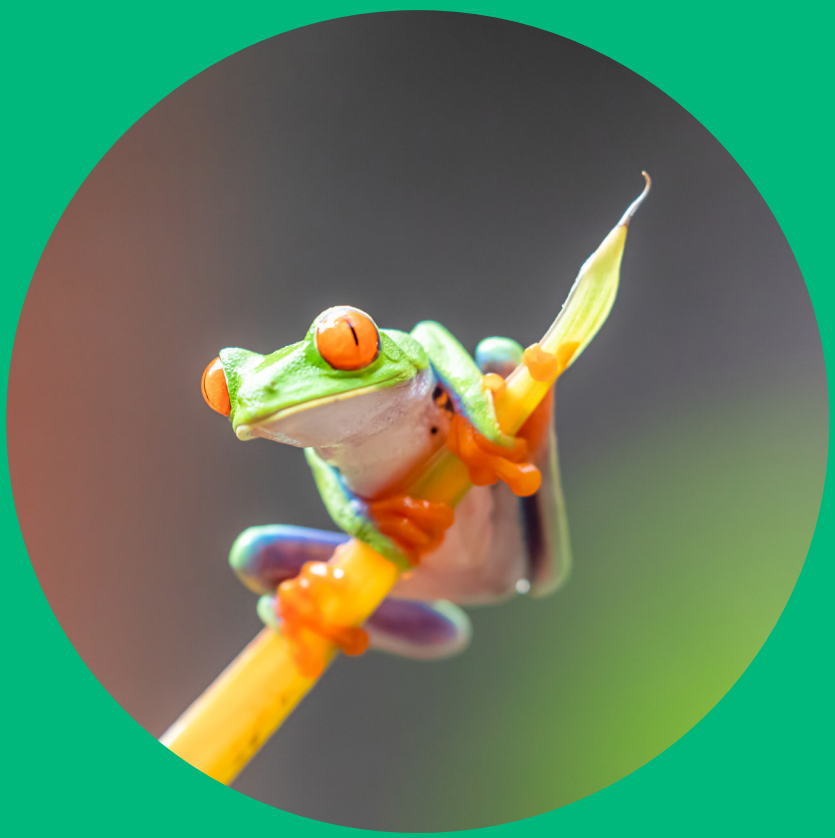
Department of Information and Communications Engineering

Methods for Convolutional Sparse Coding and Coupled Feature Learning with Applications to Image Fusion

Farshad G. Veshki



Aalto University

DOCTORAL THESES

Farshad G. Veshki

Methods for Convolutional Sparse Coding and Coupled Feature Learning with Applications to Image Fusion

Aalto University

2023

Aalto University publication series
DOCTORAL THESES 70/2023

Methods for Convolutional Sparse Coding and Coupled Feature Learning with Applications to Image Fusion

Farshad G. Veshki

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall U4, Undergraduate Centre building of the school on 21 June 2023 at 13:00.

Aalto University
School of Electrical Engineering
Department of Information and Communications Engineering

Supervising professor
Prof. Sergiy A. Vorobyov, Aalto University, Finland

Preliminary examiners
Prof. Wolfgang Heidrich, King Abdullah University of Science and Technology, Saudi Arabia
Prof. Brendt Wohlberg, Los Alamos National Laboratory, US

Opponents
Prof. Adrian Basarab, University of Lyon, France
Prof. Sergio Cruces, University of Seville, Spain

Aalto University publication series
DOCTORAL THESES 70/2023

ISBN 978-952-64-1266-5 (printed)
ISBN 978-952-64-1267-2 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-1267-2

Images: Stephanie LeBlanc via unsplash.com (cover photo)

Unigrafia Oy
Helsinki 2023

Finland

Printed matter
4041-0619

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Farshad G. Veshki

Name of the doctoral thesis
Methods for Convolutional Sparse Coding and Coupled Feature Learning with Applications to Image Fusion

Publisher School of Electrical Engineering

Unit Department of Information and Communications Engineering

Series Aalto University publication series DOCTORAL THESES 70/2023

Field of research Signal Processing Technology

Manuscript submitted 19 January 2023 **Date of the defence** 21 June 2023

Permission for public defence granted (date) 21 April 2023 **Language** English

☐ **Monograph** ☒ **Article thesis** ☐ **Essay thesis**

Abstract

The sparse approximation model, also known as the sparse coding model, represents signals as linear combinations of only a small number of elements (atoms) from a dictionary. This model is used in many applications of signal processing, machine learning, and computer vision. In many tasks, the use of dictionaries adapted to signal domains has led to significant improvements. The process of finding domain-adapted dictionaries is called dictionary learning.

Structured sparse approximation and dictionary learning has been successfully used in applications such as image fusion, where it is required to find correlated patterns in multi-measure and multimodal signals. Image fusion is the problem of combining multiple images, for example, acquired using different imaging modalities, into a single, more informative image.

A shift-invariant extension of the standard sparse approximation model that can describe the entire high-dimensional signals is referred to as convolutional sparse coding (CSC). It has been demonstrated in several studies that the CSC model is superior to its standard counterpart in representing natural signals such as audio and images.

A majority of the leading CSC and CDL algorithms are based on the alternating direction method of multipliers (ADMM) and the Fourier transform. There is only one significant difference between these methods, which is in the way they address a convolutional least-squares regression subproblem. In this thesis, we propose a novel solution for this subproblem that improves the computational efficiency of the existing algorithms. Additionally, we present an efficient ADMM-based approximate online CDL algorithm that can be used in applications that require learning large dictionaries over high-dimensional signals. Next, we propose new methods and develop computationally efficient algorithms for learning correlated features (called coupled feature learning (CFL) in this thesis) in multi-measure and multimodal signals based on sparse approximation and dictionary learning. The presented CFL algorithms potentially apply to signal and image processing tasks where a joint analysis of multiple correlated signals (e.g., multimodal images) is essential. We also propose CSC-based extensions and variations of the proposed CFL algorithm. Based on the proposed CFL methods, we develop multimodal image fusion algorithms. Specifically, the learned coupled dictionary atoms, representing correlated visual features, are used to generate unified enhanced images. We address multimodal medical image fusion, infrared and visible-light image fusion, and near-infrared and visible-light image fusion problems.

This thesis contains representative experimental results for all proposed algorithms. The effectiveness of the proposed algorithms is demonstrated based on comparisons with state-of-the-art methods.

Keywords sparse approximation, dictionary learning, convolutional sparse coding, coupled feature learning, image fusion

ISBN (printed) 978-952-64-1266-5 **ISBN (pdf)** 978-952-64-1267-2

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2023

Pages 142 **urn** http://urn.fi/URN:ISBN:978-952-64-1267-2

Preface

I would like to express my deepest gratitude to my supervisor Prof. Sergiy A. Vorobyov, whose guidance and support were invaluable throughout my doctoral studies. He provided me with the necessary resources, feedback, and encouragement to develop my ideas and pursue my research goals. His mentorship and dedication have been instrumental in shaping my academic trajectory, and I am forever grateful for his contributions to my intellectual growth.

I would also like to thank my collaborators Prof. Nora Ouzir and Prof. Esa Ollila, whose expertise and collaboration were instrumental in the success of my research. They provided me with insightful feedback, technical assistance, and valuable suggestions that significantly improved the quality and impact of my work.

I am grateful to the pre-examiners of my thesis: Prof. Wolfgang Heidrich and Prof. Brendt Wohlberg, for their valuable feedback and insights. Additionally, I would like to thank Prof. Adrian Basarab and Prof. Sergio Cruces for agreeing to serve as the opponents in my defense.

I am also grateful to my colleagues for their friendship, intellectual companionship, and moral support throughout my doctoral journey.

Helsinki, April 23, 2023,

Farshad G. Veshki

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I F. G. Veshki and S. A. Vorobyov. An Efficient Coupled Dictionary Learning Method. *IEEE Signal Processing Letters*, vol. 26(10), pp. 1441-1445, 2019.

II F. G. Veshki, N. Ouzir and S. A. Vorobyov. Image Fusion using Joint Sparse Representations and Coupled Dictionary Learning. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, pp. 8344-8348, May 2020.

III F. G. Veshki and S. A. Vorobyov. Efficient ADMM-Based Algorithms for Convolutional Sparse Coding. *IEEE Signal Processing Letters*, vol. 29, pp. 389-393, 2021.

IV F. G. Veshki, N. Ouzir, S. A. Vorobyov and E. Ollila. Multimodal Image Fusion via Coupled Feature Learning. *Signal Processing*, vol. 200, p. 108637, 2022.

V F. G. Veshki and S. A. Vorobyov. Coupled Feature Learning Via Structured Convolutional Sparse Coding for Multimodal Image Fusion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, pp. 2500-2504, May 2022.

VI F. G. Veshki and S. A. Vorobyov. Convolutional Simultaneous Sparse Approximation with Applications to RGB-NIR Image Fusion. In *Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2022.

VII F. G. Veshki and S. A. Vorobyov. Efficient Online Convolutional Dictionary

Learning Using Approximate Sparse Components. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes island, Greece, June 2023.

VIII F. G. Veshki and S. A. Vorobyov. An Efficient Approximate Method for Online Convolutional Dictionary Learning. *Submitted for publication*, 2023.

Author's Contribution

Publication I: "An Efficient Coupled Dictionary Learning Method"

The author derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Publication II: "Image Fusion using Joint Sparse Representations and Coupled Dictionary Learning"

The author derived the main algorithms and results, implemented the numerical experiments and wrote the majority of the article, incorporating comments by the co-authors.

Publication III: "Efficient ADMM-Based Algorithms for Convolutional Sparse Coding"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Publication IV: "Multimodal Image Fusion via Coupled Feature Learning"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the majority of the article, incorporating comments by the co-authors.

Publication V: "Coupled Feature Learning Via Structured Convolutional Sparse Coding for Multimodal Image Fusion"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Publication VI: "Convolutional Simultaneous Sparse Approximation with Applications to RGB-NIR Image Fusion"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Publication VII: "Efficient Online Convolutional Dictionary Learning Using Approximate Sparse Components"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Publication VIII: "An Efficient Approximate Method for Online Convolutional Dictionary Learning"

The author proposed the idea, derived the main algorithms and results, implemented the numerical experiments and wrote the article, incorporating comments by the co-authors.

Abbreviations

ADMM Alternating Direction Method of Multipliers

ASC Approximate Sparse Component

CDL Convolutional Dictionary Learning

CFL Coupled Feature Learning

CSC Convolutional Sparse Coding

CNN Convolutional Neural Networks

CT Computed Tomography

EM Expectation-Maximization

FISTA Fast Iterative Shrinkage-Thresholding Algorithm

IR Infra-Red

LS Least Squares

MMV Multi Measurement Vectors

MR Magnetic Resonance

NIR Near Infra-Red

PET Positron Emission Tomography

RGB Red-Green-Blue

SGD Stochastic Gradient Descent

SOMP Simultaneous Orthogonal Matching Pursuit

SPECT Single-Photon Emission Computed Tomography

SSA Simultaneous Sparse Approximation

SVD Singular Value Decomposition

VL Visible Light

Symbols

$\boldsymbol{D}$ Dictionary

$\boldsymbol{d}$ Convolutional filter

$\boldsymbol{I}$ Identity matrix

$\boldsymbol{s}$ Signal

$\boldsymbol{x}$ Sparse representation

ϵ Upper bound on the approximation squared error

θ Maximum number of nonzero entries in a vector

λ Sparsity regularization parameter

ρ Penalty parameter

$(\cdot)^{\circ a}$ Element-wise exponentiation to the power of a

$(\cdot)^{\star}$ Optimal value of an optimization variable

$(\cdot)^{*}$ Element-wise complex-conjugate

$(\cdot)^{+}$ Updated variable in an iterative algorithm

$(\cdot)^{H}$ Hermitian operator

$(\cdot)^{T}$ Transpose operator

$\hat{(\cdot)}$ Discrete Fourier transform of a vector

$\|\cdot\|_0$ Operator counting the number of nonzero entries in a vector

$\|\cdot\|_p$ ℓ_p-norm of a vector

$\|\cdot\|_\infty$ ℓ_∞-norm of a vector

$\|\cdot\|_{p,q}$ Mixed $\ell_{p,q}$-norm of a matrix

$\|\cdot\|_\mathrm{F}$ Frobenius norm of a matrix

$*$ Convolution operator

$\odot$ Element-wise multiplication operator

$\oslash$ Element-wise division operator

$\forall$ For all

$\boldsymbol{A}(\cdot, k)$ The k-th column of matrix $\boldsymbol{A}$

$\boldsymbol{A}(k, \cdot)$ The k-th row of matrix $\boldsymbol{A}$

$\mathcal{O}(\cdot)$ Big O notation (algorithm complexity)

$\mathrm{prox}_{\lambda\|\cdot\|_p}\left(\cdot\right)$ Proximal operator of the ℓ_p norm

$\mathcal{S}_\lambda\left(\cdot\right)$ Shrinkage operator

1. Introduction

The sparse representation model has been used in a variety of applications of signal and image processing, machine learning, and computer vision [1–4]. This model approximates a signal using a linear combination of only a small number of columns (referred to as atoms) of a matrix (called dictionary). The sparse approximation problem (also referred to as the sparse coding problem) can be formulated as

$$\underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad \|\boldsymbol{D}\boldsymbol{x}-\boldsymbol{s}\|_2^2 \leq \epsilon, \tag{1.1}$$

where $\|\cdot\|_2$ represents the Euclidean norm and $\|\cdot\|_0$ is an operator that counts the number of nonzero entries of a vector. Moreover, $\boldsymbol{D} \in \mathbb{R}^{N\times K}$, $\boldsymbol{x} \in \mathbb{R}^K$, $\boldsymbol{s} \in \mathbb{R}^N$, and ϵ represent, respectively, the dictionary, sparse representation vector, signal, and the upper bound on the approximation error. Problem (1.1) is non-convex and NP-hard. However, it can be addressed using (sub-optimal) greedy methods [5,6] or based on convex relaxation [7–9].

In many applications, the use of the sparse representation model along with a learned overcomplete dictionary has led to remarkably improved results. A learned dictionary is expected to lead to more accurate and sparser representations of its domain signals. The dictionary learning problem is commonly addressed using alternating optimization with respect to the sparse representations and the dictionary based on a training dataset [10, 11]. The dictionary optimization problem can be formulated as follows

$$\underset{\boldsymbol{D}}{\text{minimize}} \sum_{p=1}^{P} \|\boldsymbol{D}\boldsymbol{x}^p-\boldsymbol{s}^p\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{D}(\cdot,k)\|_2 = 1,\ k=1,\ldots,K, \tag{1.2}$$

where $\{\boldsymbol{s}^p\}_{p=1}^P$ is the training dataset, and the unit-norm constraint on the atoms is used to avoid scaling ambiguities. The dictionary learning problem can also be addressed using an online approach, where the dictionary is optimized incrementally after observing each training signal and finding its sparse representations. This approach is referred to as online dictionary learning [12].

Dictionary learning and sparse approximation are typically used for extraction and estimation of local patterns and features in high-dimensional signals (e.g.,

images). This usually requires a prior decomposition of the original signals into vectorized overlapping blocks (e.g., patch extraction in image processing). However, ignoring the relationships between the neighboring blocks results in multi-valued sparse representations and learning dictionaries containing similar (shifted) atoms.

Convolutional sparse coding[1] (CSC) provides a single-valued and shift-invariant model that can describe the entire high-dimensional signal. In this model, matrix-vector product $\boldsymbol{Dx}$ used in the standard sparse approximation is replaced by a sum of convolutions of dictionary filters and convolutional sparse representations (also called sparse feature maps) [13–17]. Several studies have shown that the convolutional sparse representation model significantly improves on its standard counterpart in describing natural signals such as audio and images [18–23]. However, most existing CSC and convolutional dictionary learning (CDL) algorithms have high computational costs, limiting their use to tasks including only low-dimensional signals and small datasets.

Learned dictionary atoms are commonly used as representational (e.g., visual) features to address problems that entail signal reconstruction. For example, pairs of atoms in coupled learned dictionaries are used to capture the correlated visual features in multi-measure and multimodal images. This is specifically useful for addressing different image fusion tasks. Image fusion refers to the problem of merging the information from multi-measure images or multiple images captured using different imaging sensors into a single high-quality and more informative image [24].

Over the past few years, deep learning methods have shown impressive results in various signal processing tasks, including image and speech recognition, natural language processing, and audio signal processing. These methods use large-scale neural networks to learn highly complex signal patterns, delivering state-of-the-art performance in many applications. Deep learning methods usually rely on large collections of training data. Nevertheless, although deep learning methods have become increasingly popular, sparsity-based models, such as sparse representations and dictionary learning, still have an essential role in signal processing, particularly in scenarios with limited access to the domain data or where interpretability and explainability are critical. In this thesis, we deal with such problems (image fusion) where the use of sparsity-based models is justified by the need for interpretability, limited availability of training samples (especially, in medical imaging), and also by performance superiority.

[1]The term *convolutional sparse coding* has been used to describe both convolutional sparse approximation and convolutional dictionary learning problems in some literature. In this thesis, we use this term only to refer to the convolutional sparse approximation problem.

1.1 Objectives

The main objective of this thesis is to develop effective and computationally efficient algorithms based on sparse representations and dictionary learning for extracting correlated features in multi-measure and multimodal images (high-dimensional signals with grid-like structures, in general). We also focus on developing effective image fusion methods based on the extracted correlated features in the images by using them to generate a unified reinforced representation. Furthermore, this thesis aims at developing computationally efficient CSC and CDL algorithms that can be applied to large-scale signal and image processing problems.

1.2 Contributions

- In Publication I, a simple but effective and computationally efficient method for *coupled dictionary learning* based on joint sparse approximation has been developed. In coupled dictionary learning, the relations between two correlated datasets (for example, representations of the same signals in different modalities or with different qualities) are captured using pairs of corresponding atoms in a set of dictionaries.

- In Publication II, a multifocus image fusion method based on our coupled dictionary learning algorithm has been presented. In particular, coupled dictionary learning is used to learn the mappings between focused and blurred image patches. Then, the learned focused-blurred coupled dictionaries are used to classify the relations between pairs of patches taken from the same locations in multifocus images.

- In Publication III, a computationally efficient method for the convolutional least-squares (LS) regression problem has been presented. Based on the proposed method, efficient ADMM-based CSC and CDL algorithms have been developed. In addition, we have developed an efficient algorithm for CSC in the Fourier domain with a constraint on the approximation error.

- In Publication IV, the coupled dictionary learning problem has been extended to *coupled feature learning* (CFL) in multimodal images. CFL decomposes the multimodal images into their correlated and uncorrelated components. The correlated components are estimated using a modified coupled dictionary learning method based on *simultaneous sparse approximation* (SSA). In SSA, the correlated signals are approximated using sparse representations with identical supports. This CFL model is more consistent with the characteristics of the multimodal images since the same objects can appear with varying levels of visibility in images taken using different imaging modalities. The

uncorrelated components are estimated using a constraint based on the Pearson correlation coefficient. A CFL-based multimodal image fusion method has been proposed based on the most significant representations of correlated components as well as the uncorrelated components from both input images. We have applied this method to multimodal medical and infrared-visible image fusion problems.

- In Publication V, a convolutional CFL method has been proposed where the correlated components are captured using a pair of coupled convolutional dictionaries and joint convolutional sparse representations, while the modality-specific components are estimated using a common dictionary and separate (unique) convolutional sparse representations. The resulting optimization problem has been addressed using the *alternating direction method of multipliers* (ADMM). The proposed convolutional CFL method has been applied to multimodal medical, and infrared (IR) and visible light (VL) image fusion problems.

- In Publication VI, we have presented a convolutional CFL method based on convolutional SSA with applications to the near-infrared (NIR) and visible-light image fusion problem.

- In Publication VII, an efficient ADMM-based online CDL (OCDL) algorithm based on *approximate sparse components* (ASCs) has been developed. The computational cost of the proposed method is dramatically lower than that of the other available CDL methods, making it ideal for tasks requiring CDL over large-scale data.

- In Publication VIII, we have provided a comprehensive presentation of the OCDL method in Publication VII, including detailed derivations, new algorithms, and more extensive experimental results.

- The codes for all algorithms developed in this thesis are available online at `https://users.aalto.fi/~ghorbaf1/`.

1.3 Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2 briefly reviews the existing CSC and CDL algorithms and presents the methods proposed in Publications III and VII. In Chapter 3, we discuss the CFL algorithms proposed in Publications I, V, IV and VI. Chapter 4 provides an overview of the image fusion literature and presents our CFL-based image fusion methods proposed in Publications IV, V and VI. Chapter 5 concludes this thesis by summarizing the

main results.

In each chapter, representative experimental results for the proposed methods are presented and compared to state-of-the-art algorithms. All algorithms are implemented using MATLAB. All experiments are conducted on a PC equipped with an Intel(R) Core(TM) i5-8365U 1.60GHz CPU and 16GB memory.

2. Convolutional Sparse Coding (CSC)

The CSC model describes the entire signal $\boldsymbol{s} \in \mathbb{R}^N$ using a sum of convolutions of the dictionary filters $\{\boldsymbol{d}_k \in \mathbb{R}^m\}_{k=1}^K$ and convolutional sparse representations $\{\boldsymbol{x}_k \in \mathbb{R}^N\}_{k=1}^K$, i. e.,

$$\boldsymbol{s} \simeq \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k, \tag{2.1}$$

where $*$ stands for the convolution operator. The use of the convolution operator for a shift-invariant sparse model was originally proposed by Lewicki and Sejnowski in [25] for encoding 1D time-series. Later, Mørup *et al.* [26] extended the CSC model to 2D images and music data. Numerous studies have sought to find efficient solutions to CSC and CDL problems since then [14–17, 27–35]. The CSC model has been used in a variety of signal processing and machine learning applications, including, signal restoration tasks [19, 22, 23, 36, 37], classification [38–41], image decomposition [42], fault detection [43], anomaly detection [44], source separation [19], and image reconstruction [20].

The CSC problem has been addressed based on local-block (patch-wise) sparse approximation using the existing standard sparse approximation algorithms coupled with a global signal reconstruction constraint [17, 28, 45]. Algorithms for solving variations of the CSC problem with local sparsity penalties based on mixed-norms have been proposed in [33]. Using local sparsity constraints, local priors (such as binary masks and weight maps) can be directly incorporated in the reconstruction of high-dimensional signals. Other solutions to the CSC problem in the spatial domain include the adoption of *fast iterative shrinkage-thresholding algorithm* (FISTA) [32] as well as convolutional extensions of existing greedy sparse approximation methods [46–48].

A majority of computationally efficient CSC algorithms are based on the ADMM algorithm and partly perform in the frequency (Fourier) domain [14–16, 49–51]. The main difference between these algorithms lies in the way they solve a convolutional LS regression subproblem. In this chapter, we present a novel solution to this subproblem that considerably reduces the computational costs of the most efficient existing CSC and CDL algorithms. The proposed solution to the convolutional LS regression problem is also used to develop an efficient

method that addresses the CSC problem with a constraint on the approximation error. Furthermore, in this chapter, we present an efficient OCDL method that substantially reduces the memory requirements of the existing CDL algorithms and can be used in tasks that require dictionary learning over large images.

2.1 CSC in Fourier Domain

The convolutional form of the standard sparse approximation problem (1.1) can be written as

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}} \sum_{k=1}^K \|\boldsymbol{x}_k\|_1 \quad \text{s.t.} \quad \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s} \right\|_2^2 \leq \epsilon. \tag{2.2}$$

Typically, problem (2.2) is addressed by solving its unconstrained equivalent

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2} \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s} \right\|_2^2 + \lambda \sum_{k=1}^K \|\boldsymbol{x}_k\|_1, \tag{2.3}$$

where $\lambda > 0$ is the sparsity regularization parameter. ADMM breaks the CSC problem into two main sub-problems. One of these sub-problems is a sparse approximation problem which can be straightforwardly addressed using a shrinkage operator. The challenging step is the following LS fitting problem,

$$\underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2} \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{z}_k - \boldsymbol{s} \right\|_2^2 + \frac{\rho}{2} \sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{w}_k\|_2^2, \tag{2.4}$$

where $\rho > 0$ is the ADMM penalty parameter. Based on the convolution theorem, an equivalent formulation of problem (2.4) in the Fourier domain can be written as

$$\underset{\{\hat{\boldsymbol{z}}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2} \left\| \sum_{k=1}^K \hat{\boldsymbol{d}}_k \odot \hat{\boldsymbol{z}}_k - \hat{\boldsymbol{s}} \right\|_2^2 + \frac{\rho}{2} \sum_{k=1}^K \|\hat{\boldsymbol{z}}_k - \hat{\boldsymbol{w}}_k\|_2^2, \tag{2.5}$$

where $\hat{(\cdot)}$ represents the discrete Fourier transform of a signal and $\odot$ denotes the element-wise multiplication operator. Denoting

$$\begin{aligned} \boldsymbol{\delta}_n &\triangleq [\hat{\boldsymbol{d}}_1(n), \cdots, \hat{\boldsymbol{d}}_k(n)]^T, \\ \boldsymbol{\zeta}_i &\triangleq [\hat{\boldsymbol{z}}_1(n), \cdots, \hat{\boldsymbol{z}}_k(n)]^T, \\ \boldsymbol{\omega}_i &\triangleq [\hat{\boldsymbol{w}}_1(n), \cdots, \hat{\boldsymbol{w}}_k(n)]^T, \end{aligned} \tag{2.6}$$

where $(\cdot)^T$ is the (non-conjugate) transpose operator, problem (2.5) can be seen as N independent problems

$$\underset{\boldsymbol{\zeta}_n}{\text{minimize}} \frac{1}{2} \left(\boldsymbol{\delta}_n^T \boldsymbol{\zeta}_n - \hat{s}_n \right)^2 + \frac{\rho}{2} \|\boldsymbol{\zeta}_n - \boldsymbol{\omega}_n\|_2^2. \tag{2.7}$$

Equating the derivative of the objective in (2.7) with respect to $\boldsymbol{\zeta}_n$ to zero gives

$$\begin{aligned} 0 &= \boldsymbol{\delta}_n^*(\boldsymbol{\delta}_n^T\boldsymbol{\zeta}_n - \hat{s}_n) + \rho\boldsymbol{\zeta}_n - \rho\boldsymbol{\omega}_n \\ &= (\boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T + \rho\boldsymbol{I})\boldsymbol{\zeta}_n - \hat{s}_n\boldsymbol{\delta}_n^* - \rho\boldsymbol{\omega}_n \\ &= (\boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T + \rho\boldsymbol{I})\boldsymbol{\zeta}_n - (\hat{s}_n\boldsymbol{\delta}_n^* - \boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T\boldsymbol{\omega}_n) - (\boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T + \rho\boldsymbol{I})\boldsymbol{\omega}_n, \end{aligned} \tag{2.8}$$

where $(\cdot)^*$ is the complex-conjugate of a complex number.

In Publication III, we used the third line of (2.8) and showed that (2.5) can be solved efficiently using

$$\begin{aligned} \boldsymbol{\zeta}_n^\star &= \boldsymbol{\omega}_n + (\hat{\boldsymbol{s}}_n - \boldsymbol{\delta}_n^T\boldsymbol{\omega}_n)(\boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T + \rho\boldsymbol{I})^{-1}\boldsymbol{\delta}_n^* \\ &= \boldsymbol{\omega}_n + (\hat{\boldsymbol{s}}_n - \boldsymbol{\delta}_n^T\boldsymbol{\omega}_i)(\|\boldsymbol{\delta}_n\|_2^2 + \rho)^{-1}\boldsymbol{\delta}_n^*, \end{aligned} \tag{2.9}$$

where $(\cdot)^\star$ denotes the solution to an optimization problem.

The other existing ADMM-based CSC method solves problem (2.5) using

$$\boldsymbol{\zeta}_n^\star = (\boldsymbol{\delta}_n^*\boldsymbol{\delta}_n^T + \rho\boldsymbol{I})^{-1}(\hat{s}_n\boldsymbol{\delta}_n^* + \rho\boldsymbol{\omega}_n) \tag{2.10}$$

obtained from the second line of (2.8). An efficient method for computing (2.10) based on the Sherman-Morrison formula is given in [16].

In particular, solving problem (2.5) for a batch of P training samples (signals) using the proposed method requires $((4K+1)P + 3K + 1)n$ flops, while it takes $(7KP + 3K + 1)n$ flops for solving (2.5) using the method of [16][1], indicating a considerable improvement (leading to the state-of-the-art performance) provided by our method [16].

2.2 CSC with a Constraint on the Approximation Error

It is known that for every ϵ, there exists a unique λ. Nevertheless, the appropriate value of λ is dependent on the signal and the dictionary. As a result, despite the fact that the unconstrained CSC problem is more convenient to solve, it is more favorable to address the CSC problem in the constrained form. The standard sparse approximation problem with a constraint on the approximation error has been addressed based on root-finding [52] and the augmented Lagrangian method [53].

In Publication III, we developed an efficient algorithm for solving the CSC problem with a constraint on the approximation error. Specifically, the appropriate λ values are found via root-finding by solving a single-variable optimization problem. The main steps of our algorithm are explained as follows.

The constrained CSC problem can be rewritten as

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}}\ \boldsymbol{f}\left(\{\boldsymbol{x}_k\}_{k=1}^K\right) + \sum_{k=1}^K \|\boldsymbol{x}_k\|_1, \tag{2.11}$$

[1] In this thesis, the comparisons are based on the most computationally efficient implementation of the SM method, which entails pre-computing and reusing specific quantities [16].

where $\boldsymbol{f}(\cdot)$ is the indicator function of the constraint set in (2.3), that is,

$$\boldsymbol{f}\left(\{\boldsymbol{x}_k\}_{k=1}^K\right) = \begin{cases} 0, & \text{if} \quad \boldsymbol{e}\left(\{\boldsymbol{x}_k\}_{k=1}^K\right) \leq \epsilon \\ \infty, & \text{otherwise} \end{cases}, \tag{2.12}$$

with

$$\boldsymbol{e}\left(\{\boldsymbol{x}_k\}_{k=1}^K\right) = \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s} \right\|_2^2. \tag{2.13}$$

Addressing (2.11) using ADMM leads to the following optimization problem

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}} \; \boldsymbol{f}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) + \frac{\rho}{2} \sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{w}_k\|_2^2. \tag{2.14}$$

Depending on $\{\boldsymbol{w}_k\}_{k=1}^K$, the solution to problem (2.14) is either trivial or can be found by solving an equality-constrained optimization problem. This can be written as

$$\{\boldsymbol{z}_k^\star\}_{k=1}^K = \begin{cases} \{\boldsymbol{w}_k\}_{k=1}^K, & \text{if } \boldsymbol{e}\left(\{\boldsymbol{w}_k\}_{k=1}^K\right) \leq \epsilon, \\ \underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{argmin}} \sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{w}_k\|_2^2 \quad \text{s.t. } \boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) = \epsilon, & \text{otherwise.} \end{cases} \tag{2.15}$$

Using a suitable ν, the problem in the second term of (2.15) can be reformulated as

$$\underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}} \; \boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) + \nu \sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{w}_k\|_2^2, \tag{2.16}$$

which is similar to the optimization problem in (2.4). Plugging the solution of (2.16) (which can be found using our unconstrained CSC method) into (2.13) gives

$$\boldsymbol{e}\left(\{\boldsymbol{z}_k^\star\}_{k=1}^K\right) = \frac{\nu^2}{N} \left\| \hat{\boldsymbol{r}} \oslash \left(\nu + \sum_{k=1}^K \hat{\boldsymbol{d}}_k^* \odot \hat{\boldsymbol{d}}_k\right) \right\|_2^2, \tag{2.17}$$

where $\oslash$ stands for the element-wise division operator and the division by N is required by Parseval's theorem. Thus, problem (2.14) is reduced to a single-variable optimization problem for finding the penalty parameter $\nu^\star$ that satisfies

$$\nu^\star = \left\{ \nu \mid \boldsymbol{e}\left(\{\boldsymbol{z}_k^\star\}_{k=1}^K\right) = \epsilon \right\}. \tag{2.18}$$

This can be addressed, for example, using the *bisection* method.

The complexity of our constrained CSC algorithm is the same as that of the proposed unconstrained CSC algorithm (both are of $\mathcal{O}(K)$). However, the constrained CSC algorithms results in slightly longer runtimes, which accounts for solving the single-variable optimization problem for finding $\nu^\star$ in each iteration.

2.3 Convolutional Dictionary Learning (CDL)

A common formulation of the CDL problem is written as

$$\underset{\{\{\boldsymbol{x}_k^p\}_{k=1}^K\}_{p=1}^P,\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{2}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k * \boldsymbol{x}_k^p - \boldsymbol{s}^p\right\|_2^2 + \lambda\sum_{k=1}^{K}\|\boldsymbol{x}_k^p\|_1\right) \quad \text{s.t.} \quad \|\boldsymbol{d}_k\|_2 \leq 1,\ k = 1,\ldots,K. \tag{2.19}$$

The CDL problem is usually addressed using a batch CDL approach where the sparse representations and the dictionary filters are alternatingly optimized using a training dataset [14–16, 29]. The dictionary optimization problem can be formulated as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\sum_{n=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k * \boldsymbol{x}_k^p - \boldsymbol{s}^p\right\|_2^2 + \sum_{k=1}^{K}\boldsymbol{\Omega}(\boldsymbol{d}_k), \tag{2.20}$$

where $\boldsymbol{\Omega}(\cdot)$ is the indicator function of the constraint in (2.19), that is,

$$\boldsymbol{\Omega}(\boldsymbol{d}) = \begin{cases} 0, & \text{if} \quad \|\boldsymbol{d}\|_2 \leq 1 \\ \infty, & \text{otherwise} \end{cases}.$$

2.4 CDL Based on Consensus ADMM

An efficient solution to the batch CDL problem based on the convolutional theorem and the consensus ADMM framework is provided in [50]. In this method, the dictionary optimization over the entire dataset is addressed in a distributed manner. Specifically, in one of the two main steps of ADMM algorithm, the dictionary is separately optimized with respect to each of the data samples by solving P independent optimization problems. In the second step, the global (fused) dictionary is found by projecting the average of independently optimized dictionaries onto the constraint set.

The independent optimization problems in the consensus ADMM-based CDL are convolutional LS regression problems similar to (2.4). Thus, they can be addressed more efficiently using the convolutional LS regression method proposed in Publication III.

Experimental evaluations based on image data, performed in Publication III, have shown that incorporating the proposed convolutional LS regression method in the consensus ADMM-based batch CDL algorithm leads to a significantly improved computational efficiency compared to the state-of-the-art available algorithms.

2.5 Online CDL in Fourier Domain

In batch CDL, the convolutional sparse representations for the entire dataset need to be accessed at once. This requires memory of the order of NPK, which is computationally expensive when using large training datasets. OCDL improves the computational efficiency of batch CDL by compactly storing the information in the training samples and their sparse representations using a pair of history arrays. An OCDL reformulation of problem (2.20) in the Fourier domain is written as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2}\sum_{n=1}^{N} \boldsymbol{\delta}_n^H \boldsymbol{A}_n^P \boldsymbol{\delta}_n - \sum_{p=1}^{P} \boldsymbol{\delta}_n^T \boldsymbol{b}_n^P + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k), \tag{2.21}$$

where $(\cdot)^H$ denotes the Hermitian transpose, and the history arrays $\boldsymbol{A}_n^P \in \mathbb{R}^{K\times K}$ and $\boldsymbol{b}_n^P \in \mathbb{R}^K$, $n = 1,\ldots,N$, are defined as

$$\boldsymbol{A}_n^P \triangleq \frac{1}{NP}\sum_{p=1}^{P} (\boldsymbol{\zeta}_n^p)^* (\boldsymbol{\zeta}_n^p)^T, \quad \boldsymbol{b}_n^P \triangleq \frac{1}{NP}\sum_{p=1}^{P} \hat{\boldsymbol{s}}^p(n)^* \boldsymbol{\zeta}_n^p, \tag{2.22}$$

with $\boldsymbol{\delta}_n^p$ and $\boldsymbol{\zeta}_n^p$ being the same as in (2.6). The history arrays are updated after observing each training sample and finding its sparse representations. The updates are performed using

$$\begin{aligned} \boldsymbol{A}_n^P &= \frac{1}{NP}(\boldsymbol{\zeta}_n^P)^* (\boldsymbol{\zeta}_n^P)^T + \frac{P-1}{P}\boldsymbol{A}_n^{P-1},\ n = 1,\ldots,N, \\ \boldsymbol{b}_n^P &= \frac{1}{NP}\hat{\boldsymbol{s}}^P(n)^* \boldsymbol{\zeta}_n^P + \frac{P-1}{P}\boldsymbol{b}_n^{P-1},\ n = 1,\ldots,N, \end{aligned} \tag{2.23}$$

where $\boldsymbol{A}_n^0$ and $\boldsymbol{b}_n^0$ are initialized with all-zero arrays. The dictionary is optimized by solving problem (2.21) once the updated history arrays are available. Efficient solutions to the OCDL problem in the Fourier domain have been proposed based on ADMM [27], the projected *stochastic gradient descent* (SGD) method and FISTA [27, 34].

2.5.1 Approximate Online CDL

The use of the available OCDL algorithms for learning large dictionaries over high-dimensional signals can still be prohibitively computationally costly. In Publication VII, we have proposed a novel OCDL method that dramatically reduces the computational cost of the existing algorithms.

In the proposed OCDL method the training signals are approximated in a distributed manner using P distinct dictionaries $\{\boldsymbol{c}_k^p \in \mathbb{R}^m\}_{k=1}^K$. A fusion of the separately optimized dictionaries based on the respective convolutional sparse representations is used to calculate the dictionary $\{\boldsymbol{d}_k\}_{k=1}^K$. Specifically, the quadratic term in CDL problem (2.20) is approximated using the following

upper-bound estimate

$$\sum_{p=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2=\sum_{p=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k*\boldsymbol{x}_k^p-\sum_{k=1}^{K}\boldsymbol{c}_k^p*\boldsymbol{x}_k^p+\sum_{k=1}^{K}\boldsymbol{c}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^n\right\|_2^2$$

$$\leq\sum_{p=1}^{P}\sum_{k=1}^{K}\left\|\boldsymbol{d}_k*\boldsymbol{x}_k^p-\boldsymbol{c}_k^p*\boldsymbol{x}_k^p\right\|_2^2+\sum_{p=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{c}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2, \quad (2.24)$$

where the inequality is due to the triangle inequality. Accordingly, the proposed approximate CDL problem is formulated as

$$\underset{\substack{\{\boldsymbol{d}_k\}_{k=1}^K,\\ \{\{\boldsymbol{c}_k^p\}_{k=1}^K\}_{p=1}^P}}{\text{minimize}}\frac{1}{2P}\sum_{p=1}^{P}\sum_{k=1}^{K}\left\|\boldsymbol{d}_k*\boldsymbol{x}_k^p-\boldsymbol{c}_k^p*\boldsymbol{x}_k^p\right\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}(\boldsymbol{d}_k)$$

$$+\frac{1}{2P}\sum_{p=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{c}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2+\sum_{p=1}^{P}\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{c}_k^p\right). \quad (2.25)$$

In Publication VIII, two ADMM-based online methods for addressing (2.25) have been presented. The first algorithm uses a standard approach for optimization of $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\boldsymbol{c}_k^P\}_{k=1}^K$, while the second algorithm incorporates pragmatic modifications to the first algorithm to improve the effectiveness of the proposed approximation method and lower computational costs.

Approximate OCDL Algorithm 1

Optimization problem (2.25) is jointly convex with respect to $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\{\boldsymbol{c}_k^p\}_{k=1}^K\}_{p=1}^P$. Thus, using the OCDL framework, problem (2.25) can be addressed by jointly optimizing the variables $\{\boldsymbol{c}_k^P,\boldsymbol{d}_k\}_{k=1}^K$ after observing the Pth training signal $\boldsymbol{s}^P$ and obtaining its convolutional sparse representations $\{\boldsymbol{x}_k^P\}_{k=1}^K$. Compact history arrays are used to store sufficient statistics of $\{\{\boldsymbol{c}_k^p\}_{k=1}^K\}_{p=1}^{P-1}$ and $\{\{\boldsymbol{x}_k^p\}_{k=1}^K\}_{p=1}^{P-1}$.

The following ADMM formulation is used to solve (2.25) for $\{\boldsymbol{c}_k^P,\boldsymbol{d}_k\}_{k=1}^K$

$$\underset{\substack{\{\boldsymbol{c}_k^P,\boldsymbol{d}_k\}_{k=1}^K,\\ \{\boldsymbol{f}_k^P,\boldsymbol{g}_k\}_{k=1}^K}}{\text{minimize}}\frac{1}{2P}\sum_{p=1}^{p}\sum_{k=1}^{K}\left\|\boldsymbol{g}_k*\boldsymbol{x}_k^p-\boldsymbol{f}_k^p*\boldsymbol{x}_k^p\right\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}(\boldsymbol{d}_k)$$

$$+\frac{1}{2P}\sum_{p=1}^{P}\left\|\sum_{k=1}^{K}\boldsymbol{f}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2+\sum_{p=1}^{P}\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{c}_k^p\right)$$

$$\text{s.t.}\quad \boldsymbol{g}_k=\boldsymbol{d}_k,\quad \boldsymbol{f}_k^P=\boldsymbol{c}_k^P,\ k=1,\ldots,K, \quad (2.26)$$

where $\{\boldsymbol{f}_k^P,\boldsymbol{g}_k\}_{k=1}^K$ are the (joint) ADMM auxiliary variables. The main ADMM iterations consist of the $\{\boldsymbol{f}_k^P,\boldsymbol{g}_k\}_{k=1}^K$-update step (a convolutional least-squares

fitting problem) and the $\{\boldsymbol{c}_k^P, \boldsymbol{d}_k\}_{k=1}^K$-update step (projection on the constraint set).

The $\{\boldsymbol{f}, \boldsymbol{g}\}$-update step entails solving the following optimization problems

$$\underset{\{\boldsymbol{f}_k^P\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{f}}_k^P \odot \hat{\boldsymbol{x}}_k^P - \hat{\boldsymbol{z}}_k^P\right\|_2^2 + \frac{1}{2P}\left\|\sum_{k=1}^{K}\hat{\boldsymbol{f}}_k^P \odot \hat{\boldsymbol{x}}_k^P - \hat{\boldsymbol{s}}^P\right\|_2^2 + \frac{\rho}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{f}}_k^P - \hat{\boldsymbol{q}}_k\right\|_2^2, \tag{2.27}$$

$$\underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\sum_{n=1}^{P}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_k \odot \hat{\boldsymbol{x}}_k^p - \hat{\boldsymbol{t}}_k^p\right\|_2^2 + \frac{\rho}{2}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{g}}_k - \hat{\boldsymbol{w}}_k\right\|_2^2, \tag{2.28}$$

where $\boldsymbol{z}_k^P \triangleq \boldsymbol{g}_k * \boldsymbol{x}_k^P$ and $\boldsymbol{t}_k^p \triangleq \boldsymbol{f}_k^p * \boldsymbol{x}_k^p$.

By equating the derivative of the objective in (2.27) to zero and using the Sherman-Morrison (SM) formula, the solution to the $\boldsymbol{f}$-update step can be found as

$$\left(\hat{\boldsymbol{f}}_k^P(n)\right)^\star = \left(a_n^k + \frac{(a_n^k)^2|\hat{\boldsymbol{x}}_k^P(n)|^2}{1+\sum_{k=1}^K a_n^k|\hat{\boldsymbol{x}}_k^P(n)|^2}\right)\left((\hat{\boldsymbol{x}}_k^P(n))^*\left(\hat{\boldsymbol{z}}_k^P(n) + \hat{\boldsymbol{s}}^P(n)\right) + P\rho\hat{\boldsymbol{q}}_k(n)\right), \tag{2.29}$$

where $a_n^k \triangleq (|\hat{\boldsymbol{x}}_k^P(n)|^2 + P\rho)^{-1}$.

Using precalculated values of $\sum_{k=1}^K a_n^k|\hat{\boldsymbol{x}}_k^P(n)|^2$, the $\boldsymbol{f}$-update step can be carried out with the complexity of $\mathcal{O}(KN)$ using (2.29).

The solution to (2.28) (the $\boldsymbol{g}$-update step) can be found as

$$\left(\hat{\boldsymbol{g}}_k(n)\right)^\star = \frac{\boldsymbol{\beta}_k^P(n) + \hat{\boldsymbol{w}}_k(n)}{\boldsymbol{\alpha}_k^P + \rho},\ n = 1,\ldots,N,\ k = 1,\ldots,K, \tag{2.30}$$

where history arrays $\boldsymbol{\alpha}_k^P \in \mathbb{R}^N$ and $\boldsymbol{\beta}_k^P \in \mathbb{R}^N$, $k = 1,\ldots,K$, are defined as

$$\boldsymbol{\alpha}_k^P \triangleq \frac{1}{P}\sum_{p=1}^{P}(\hat{\boldsymbol{x}}_k^p)^* \odot \hat{\boldsymbol{x}}_k^p,\ \boldsymbol{\beta}_k^P \triangleq \frac{1}{P}\sum_{p=1}^{P}(\hat{\boldsymbol{x}}_k^p)^* \odot \hat{\boldsymbol{t}}_k^p. \tag{2.31}$$

The history arrays are incrementally updated using

$$\boldsymbol{\alpha}_k^P = \frac{P-1}{P}\boldsymbol{\alpha}_k^{P-1} + \frac{1}{P}(\hat{\boldsymbol{x}}_k^P)^* \odot \hat{\boldsymbol{x}}_k^P, \tag{2.32}$$

$$\boldsymbol{\beta}_k^P = \frac{P-1}{P}\boldsymbol{\beta}_k^{P-1} + \frac{1}{P}(\hat{\boldsymbol{x}}_k^P)^* \odot \hat{\boldsymbol{t}}_k^P. \tag{2.33}$$

Approximate OCDL Algorithm 2

To improve the performance of the proposed OCDL algorithm, dictionary optimization can be performed *exactly* for the latest observed signal $\boldsymbol{s}^P$, while the proposed approximation method is used for $\{\boldsymbol{s}^p\}_{p=1}^{P-1}$. Thus, the modified

approximate CDL problem is now formulated as

$$\underset{\substack{\{\boldsymbol{d}_k\}_{k=1}^K,\\ \{\{\boldsymbol{c}_k^p\}_{k=1}^K\}_{p=1}^P}}{\text{minimize}} \frac{1}{2P}\left\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k^P - \boldsymbol{s}^P\right\|_2^2 + \frac{1}{2P}\sum_{n=1}^{P-1}\sum_{k=1}^K \left\|\boldsymbol{d}_k * \boldsymbol{x}_k^p - \boldsymbol{c}_k^p * \boldsymbol{x}_k^p\right\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}(\boldsymbol{d}_k)$$
$$+ \frac{1}{2P}\sum_{p=1}^{P-1}\left\|\sum_{k=1}^K \boldsymbol{c}_k^p * \boldsymbol{x}_k^p - \boldsymbol{s}^p\right\|_2^2 + \sum_{p=1}^N\sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{c}_k^p\right). \quad (2.34)$$

Problem (2.34) can be solved using alternating optimization with respect to $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\boldsymbol{c}_k^P\}_{k=1}^K$.

Problem (2.34) can be addressed with respect to $\{\boldsymbol{d}_k\}_{k=1}^K$ using the following ADMM formulation

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K, \{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\left\|\sum_{k=1}^K \boldsymbol{g}_k * \boldsymbol{x}_k^P - \boldsymbol{s}^P\right\|_2^2 + \frac{1}{2P}\sum_{p=1}^{P-1}\sum_{k=1}^K \left\|\boldsymbol{g}_k * \boldsymbol{x}_k^p - \boldsymbol{r}_k^p\right\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}(\boldsymbol{d}_k)$$
$$\text{s.t.} \quad \boldsymbol{g}_k = \boldsymbol{d}_k,\, k = 1,\ldots,K. \quad (2.35)$$

where $\boldsymbol{r}_k^p \triangleq \boldsymbol{c}_k^p * \boldsymbol{x}_k^p$.

The $\boldsymbol{g}$-update step requires solving the optimization problem in the form of

$$\underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\left\|\sum_{k=1}^K \hat{\boldsymbol{g}}_k \odot \hat{\boldsymbol{x}}_k^P - \hat{\boldsymbol{s}}^P\right\|_2^2 + \frac{1}{2P}\sum_{p=1}^{P-1}\sum_{k=1}^K \left\|\hat{\boldsymbol{g}}_k \odot \hat{\boldsymbol{x}}_k^p - \hat{\boldsymbol{r}}_k^p\right\|_2^2 + \frac{\rho}{2}\sum_{k=1}^K \left\|\hat{\boldsymbol{g}}_k - \hat{\boldsymbol{e}}_k\right\|_2^2. \quad (2.36)$$

Equating the derivative to zero and using the SM formula, optimization problem (2.36) can be solved as

$$\left(\hat{\boldsymbol{g}}_k^P(n)\right)^\star = \left(b_n^k + \frac{(b_n^k)^2|\hat{\boldsymbol{x}}_k^P(n)|^2}{P + \sum_{k=1}^K b_n^k|\hat{\boldsymbol{x}}_k^P(n)|^2}\right)\left(\frac{1}{P}(\hat{\boldsymbol{x}}_k^P(n))^*\hat{\boldsymbol{s}}^P(n) + \tilde{\boldsymbol{\beta}}_k^{P-1}(n) + \rho\hat{\boldsymbol{e}}_k(n)\right), \quad (2.37)$$

with $b_n^k \triangleq (\tilde{\boldsymbol{\alpha}}_k^{P-1}(n) + \rho)^{-1}$, where history arrays $\tilde{\boldsymbol{\alpha}}_k^P \in \mathbb{R}^N$ and $\tilde{\boldsymbol{\beta}}_k^P \in \mathbb{R}^N$, $k = 1,\ldots,K$, are defined as

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_k^P &\triangleq \frac{1}{P+1}\sum_{p=1}^P (\hat{\boldsymbol{x}}_k^p)^* \odot \hat{\boldsymbol{x}}_k^p, \\ \tilde{\boldsymbol{\beta}}_k^P &\triangleq \frac{1}{P+1}\sum_{p=1}^P (\hat{\boldsymbol{x}}_k^p)^* \odot \hat{\boldsymbol{r}}_k^p \end{aligned} \quad . \quad (2.38)$$

The incremental update rules for $\tilde{\boldsymbol{\alpha}}_k^P$ and $\tilde{\boldsymbol{\beta}}_k^P$ can be found as

$$\tilde{\boldsymbol{\alpha}}_k^P = \frac{P}{P+1}\tilde{\boldsymbol{\alpha}}_k^{P-1} + \frac{1}{P+1}(\hat{\boldsymbol{x}}_k^P)^* \odot \hat{\boldsymbol{x}}_k^P, \quad (2.39)$$

$$\tilde{\boldsymbol{\beta}}_k^P = \frac{P}{P+1}\tilde{\boldsymbol{\beta}}_k^{P-1} + \frac{1}{P+1}(\hat{\boldsymbol{x}}_k^P)^* \odot \hat{\boldsymbol{r}}_k^P. \quad (2.40)$$

The $\boldsymbol{g}$-update (2.37) can be performed with the complexity of $\mathcal{O}(KN)$ using precalculated values of $\sum_{k=1}^{K} b_n^k |\hat{\boldsymbol{x}}_k^P(n)|^2$.

In the modified algorithm, dictionary $\{\boldsymbol{c}_k^P\}_{k=1}^K$ is optimized only to provide a more accurate approximation of $\boldsymbol{s}^P$ (in comparison with the approximation provided using $\{\boldsymbol{d}_k\}_{k=1}^K$). It means that the second quadratic term in (2.34) is ignored in the step of $\{\boldsymbol{c}_k^P\}_{k=1}^K$ optimization. Here we rely on the fact that $\{\boldsymbol{x}_k^P\}_{k=1}^K$ are direct products of $\{\boldsymbol{d}_k\}_{k=1}^K$. As a result, considering that the approximation is based on $\{\boldsymbol{x}_k^P\}_{k=1}^K$, the resulting $\{\boldsymbol{c}_k^P\}_{k=1}^K$ cannot unfavorably deviate from $\{\boldsymbol{d}_k\}_{k=1}^K$. Problem (2.34), which needs to be solved now for $\{\boldsymbol{c}_k^P\}_{k=1}^K$ only, is then reduced to the following optimization problem

$$\underset{\{\boldsymbol{c}_k^P\}_{k=1}^K}{\text{minimize}} \frac{1}{2P} \left\| \sum_{k=1}^{K} \boldsymbol{c}_k^P * \boldsymbol{x}_k^P - \boldsymbol{s}^P \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}\left(\boldsymbol{c}_k^P\right), \tag{2.41}$$

which is a CDL problem involving a single training signal, and can be addressed using the existing CDL methods (e.g., [51]).

Computational Efficiency

The largest arrays used in the proposed approximate OCDL methods are of size KN, dramatically smaller than those used by the state-of-the-art batch CDL algorithms and OCDL algorithms, that are, KNP and K^2N, respectively. In addition, the proposed OCDL algorithms has a time complexity of $\mathcal{O}(KNP)$, which is equal to that of the most efficient batch CDL algorithm and significantly less than that of the state-of-the-art OCDL algorithm ($\mathcal{O}(K^2NP)$).

2.6 Experimental Results

In this section, the proposed CSC and CDL methods are compared to the state-of-the-art algorithms. The CSC experiments are performed using a 512×512 Lena image. The CDL experiments are conducted using the following image datasets:

1. SIPI: 40 greyscale images of size 256×256 taken from the USC-SIPI database (`http://sipi.usc.edu/database/`).

2. Flowers: 210 greyscale images of flowers of size 200×200 taken from Oxford Flower Datasets (`https://www.robots.ox.ac.uk/~vgg/data/flowers/`).

The original images are resized and converted to greyscale. The pixel values are normalized to be between 0 and 1 (the original 8-bit values are divided by 255). Since the CSR model is not able to effectively represent the low-frequency component of the signals, it is conventional that the images used for CDL are high-pass filtered [16, 31, 34]. Here, the low-frequency components of all images are removed using the *lowpass* function of the SPORCO toolbox [54] with a

regularization parameter of 5.

All CDL experiments are performed using $\lambda = 0.1\lambda_{\max}$, where $\lambda_{\max}$ is the smallest regularization parameter value that leads to all-zero sparse representations and can be obtained using ℓ_∞-norm of the gradient of the objective of CSC problem (2.3) at $\{\boldsymbol{x}_k\}_{k=1}^K = \mathbf{0}$. We use ADMM penalty parameter $\rho = 10$ and dictionary filters of size 8×8 in all experiments.

2.6.1 CSC Results

Fig. 2.1 compares the functional values over time for 50 iterations of the proposed CSC algorithms (Publication III) and the CSC method based on Sherman-Morrison formula [16] using different values of λ tested, the Lena image as the input signal, and a learned dictionary composed of 64 filters. Note that the iterations of the two unconstrained CSC methods, the proposed method discussed in Section 2.1 (red curve) and the method of [16] (blue curve), are equally effective. Thus, the use of a fixed number of iterations illustrates the difference in computational efficiency. As can be seen, the proposed algorithm is considerably more efficient for all λ values.

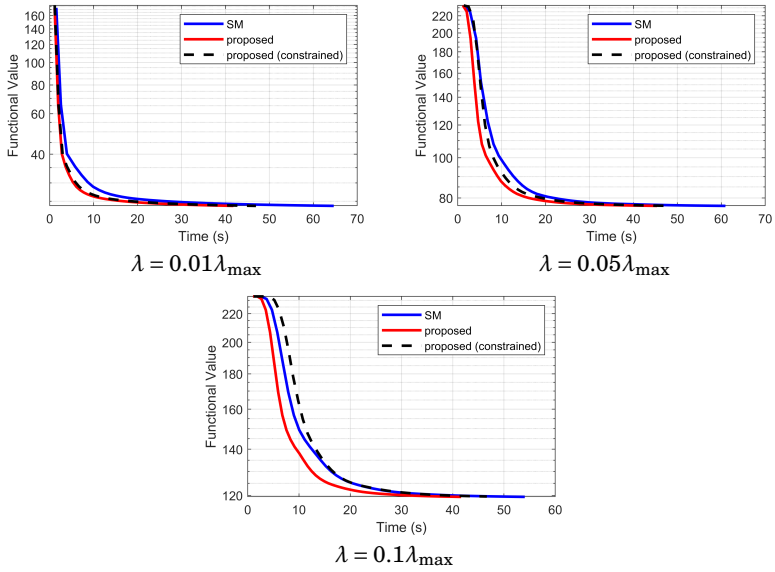


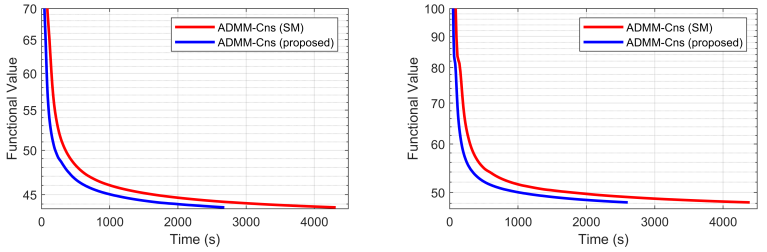
Figure 2.1. Functional values over time for the proposed CSC methods [51] and the method based on the Sherman-Morrison formula (SM) [16].

The ϵ values (the values of the quadratic functional term) obtained by executing the unconstrained CSC algorithms with different λ values are used to run the proposed constrained CSC algorithm (see dashed black curves in Fig. 2.1). It can be seen that the constrained CSC algorithm converges to the same functional values with a slightly longer runtime compared to that of the proposed

unconstrained CSC algorithm, which accounts for single variable optimization in each iteration.

2.6.2 CDL Results

Fig. 2.2 compares the functional values over time for 300 iterations of ADMM-consensus-based CDL methods based on the convolutional LS regression method proposed in Publication III (blue curve) and the Sherman-Morrison formula [31] (red curve) using different dictionary sizes K and different number of training images P (subsets of the SIPI dataset). The iterations of the two CDL methods compared are equally effective. However, as can be seen, the proposed method is substantially more computationally efficient.



$K = 64, P = 20$ $K = 128, P = 10$

Figure 2.2. Functional values over time for ADMM-consensus-based CDL algorithms based on the proposed convolutional LS regression method [51] and the Sherman-Morrison formula (SM) [31] using subsets of the SIPI dataset.

2.6.3 OCDL Results

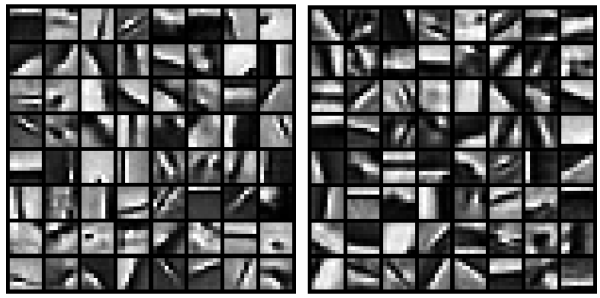
We compare the proposed OCDL method (Algorithm 2 in Publication VIII) to the OCDL method based on FISTA (with gradient calculated using the Fourier transform) proposed in [34]. For the proposed method, we use 200 ADMM iterations (maximum) with absolute and relative tolerance values of 10^{-4}. The comparisons are conducted based on the objective functional values (fval) of (2.19) for each datatset. We use 4 images taken from the SIPI dataset and 10 images taken from the Flowers dataset (different from images used for CDL) as test datasets. For the SIPI dataset, the results for both training and test datasets are reported. For the larger dataset Flowers, since it is infeasible to store all training sparse representations, the test results only are reported.

Tables 2.1 reports the objective functional values obtained using the methods tested for the SIPI ($K = 64$ and $P = 32$) and the Flowers ($K = 100$ and $P = 200$) datasets, respectively. As can be observed, the proposed method leads to a significant reduction in training time. In addition, the proposed method yields competitive results, while substantially reducing memory requirements.

Dictionaries learned using the methods compared are illustrated in Figs. 2.3 and 2.4.

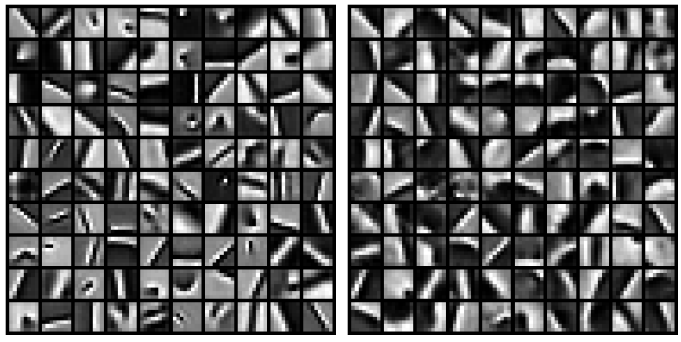
Table 2.1. The results obtained using the SIPI dataset with $K = 64$ and $P = 32$ and the Flowers dataset with $K = 100$ and $P = 200$. (K is the number of dictionary filters and P is the number of images in the training dataset.)

	SIPI			Flowers	
Methods	**train fval**	**test fval**	**train time (s)**	**test fval**	**train time (s)**
Initial dictionary	-	70.0661	-	83.5131	-
FISTA [34]	38.5268	45.1158	2211	46.8404	49244
proposed	36.6363	46.2067	796	48.9559	4614



FISTA [34] proposed

Figure 2.3. Dictionaries learned using the SIPI dataset with parameters $K = 64$ and $P = 32$.



FISTA [34] proposed

Figure 2.4. Dictionaries learned using the Flowers dataset with parameters $K = 100$ and $P = 200$.

3. Coupled Feature Learning (CFL)

Many real-world image processing and computer vision applications require joint analysis of multiple images, for example, acquired using different imaging modalities. Examples of these applications are multimodal image denoising and image fusion [24,55]. CFL aims to capture the correlated features in multimodal images by decomposing them into their correlated and uncorrelated components based on structured sparse approximation and dictionary learning. In this thesis, the extracted correlated features are used to generate unified and reinforced (fused) multimodal images. This chapter presents CFL after briefly reviewing the related concepts and relevant literature.

3.1 Related Works

3.1.1 Simultaneous Sparse Approximation (SSA)

SSA approximates a set of multi-measure signals using different linear combinations of the same subset of atoms in a dictionary, i.e., sparse representations with identical supports [56,57]. The SSA problem can be formulated as

$$\begin{aligned} &\underset{\{\boldsymbol{x}_l\}_{l=1}^{L}}{\text{minimize}} \sum_{l=1}^{L} \left(\frac{1}{2} \|\boldsymbol{D}\boldsymbol{x}_l - \boldsymbol{s}_l\|_2^2 + \lambda \|\boldsymbol{x}_l\|_0 \right) \\ &\text{s.t.} \quad \text{Supp}\left(\boldsymbol{x}_l\right) = \text{Supp}\left(\boldsymbol{x}_{l'}\right), \quad l, l' = 1, \ldots, L. \end{aligned} \tag{3.1}$$

The SSA model has been employed in various signal and image processing tasks to represent multiple dependent signals. For example, multi measurement vectors (MMV) problems [58,59], source separation [60], anomaly detection [61] and image fusion [62].

The SSA problem can be addressed using greedy methods such as *simultaneous orthogonal matching pursuit* (SOMP) [56] or based on convex relaxation using mixed-norms [57,63]. For a matrix $\boldsymbol{A} \in \mathbb{R}^{R \times C}$, the mixed $\ell_{p,q}$-norm, $p, q \geq 1$, is

defined as

$$\|\boldsymbol{A}\|_{p,q} \triangleq \left(\sum_{r=1}^{R} \|\boldsymbol{A}(r,\cdot)\|_p^q \right)^{\frac{1}{q}}.$$

For instance, the works in [64] and [57] have used the $\ell_{2,1}$- and the $\ell_{\infty,1}$-norms for addressing the SSA problem, respectively. A convex relaxation of (3.1) based on the $\ell_{2,1}$-norm can be written as

$$\underset{\{\boldsymbol{x}_l\}_{l=1}^{L}}{\text{minimize}} \frac{1}{2} \sum_{l=1}^{L} \|\boldsymbol{D}\boldsymbol{x}_l - \boldsymbol{s}_n\|_2^2 + \lambda \|\boldsymbol{X}\|_{2,1}, \tag{3.2}$$

where $\boldsymbol{X} \triangleq [\boldsymbol{x}_1 \cdots \boldsymbol{x}_L]$. Solving (3.2) entails minimizing the sum of the ℓ_2-norms of the rows of $\boldsymbol{X}$. This leads to a *row-sparse* $\boldsymbol{X}$, which is mostly zeros with only a small number of nonzero and dense rows. A convolutional extension of (3.2) has been addressed in [65]. A row-sparse structure with sparse rows can be enforced by adding an ℓ_1-norm regularization term to the objective function of (3.2) [63]. This can be written as

$$\underset{\{\boldsymbol{x}_l\}_{l=1}^{L}}{\text{minimize}} \frac{1}{2} \sum_{l=1}^{L} \|\boldsymbol{D}\boldsymbol{x}_l - \boldsymbol{s}_l\|_2^2 + \lambda_1 \|\boldsymbol{X}\|_{2,1} + \lambda_2 \|\boldsymbol{X}\|_{1,1}, \tag{3.3}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the row-sparsity and element-sparsity regularization parameters, respectively.

3.1.2 Multimodal Dictionary Learning

Here, we briefly review models based on sparse representations and dictionary learning used for representing multimodal images.

The work of [66] proposed to model L multimodal signals $\{\boldsymbol{s}_l\}_{l=1}^{L}$ (e.g., multimodal images patches) using dictionary $\boldsymbol{D} = [\boldsymbol{D}^{\mathrm{z}}\, \boldsymbol{D}^{\mathrm{e}}]$ and sparse representations $\boldsymbol{x}_l = [(\boldsymbol{x}^{\mathrm{z}})^T\, (\boldsymbol{x}_l^{\mathrm{e}})^T]^T$. In this model, the multimodal signals are assumed to contain common (identical) components represented by $\boldsymbol{z} = \boldsymbol{D}^{\mathrm{z}}\boldsymbol{x}^{\mathrm{z}}$, where $\boldsymbol{D}^{\mathrm{z}}$ and $\boldsymbol{x}^{\mathrm{z}}$ are the dictionary of common features and the common sparse representations, respectively. The modality-specific components are described using $\boldsymbol{e}_l = \boldsymbol{D}^{\mathrm{e}}\boldsymbol{x}_l^{\mathrm{z}}$, $l = 1,\ldots,L$, where $\boldsymbol{D}^{\mathrm{e}}$ is the dictionary of unique features and $\{\boldsymbol{x}_l^{\mathrm{e}}\}_{l=1}^{N}$ are the unique sparse representations (the only modality-dependent variable of the model).

In [67], the correlated components of multimodal images are captured using L coupled dictionaries $\{\boldsymbol{D}_l^{\mathrm{z}}\}_{l=1}^{L}$ and common sparse representations $\boldsymbol{x}^{\mathrm{z}}$, i.e., using $\boldsymbol{z}_l = \boldsymbol{D}_l^{\mathrm{z}}\boldsymbol{x}^{\mathrm{z}}$. Moreover, the unique components are estimated using a set of unique dictionaries and unique sparse representations ($\boldsymbol{e}_l = \boldsymbol{D}_l^{\mathrm{e}}\boldsymbol{x}_l^{\mathrm{e}}$). Thus, the multimodal signals are represented using $\boldsymbol{s}_l = \boldsymbol{D}_l^{\mathrm{z}}\boldsymbol{x}^{\mathrm{z}} + \boldsymbol{D}_l^{\mathrm{e}}\boldsymbol{x}_l^{\mathrm{e}}$, $l = 1,\ldots,L$. An extension of this model to convolutional sparse representations is provided in [55].

3.2 Coupled Dictionary Learning

Coupled dictionary learning seeks to capture the nonlinear mappings between multi-measure signals (e.g., high- and low-resolution images) using a pair of dictionaries by enforcing common sparse representations. This can be formulated as the following optimization problem

$$\begin{aligned} &\underset{\boldsymbol{D}_1, \boldsymbol{D}_2, \{\boldsymbol{x}^p\}_{p=1}^P}{\text{minimize}} \sum_{p=1}^{P} \left(\left\| \boldsymbol{D}_1 \boldsymbol{x}^p - \boldsymbol{s}_1^p \right\|_2^2 + \left\| \boldsymbol{D}_2 \boldsymbol{x}^p - \boldsymbol{s}_2^p \right\|_2^2 \right) \\ &\text{s.t.} \left\| \boldsymbol{x}^p \right\|_0 \leq \theta, \ \|\boldsymbol{D}_1(\cdot,k)\|_2 = 1, \ \|\boldsymbol{D}_2(\cdot,k)\|_2 = 1, \ \forall p,k, \end{aligned} \tag{3.4}$$

where $\{\boldsymbol{s}_1^p\}_{p=1}^P$ and $\{\boldsymbol{s}_2^p\}_{p=1}^P$ are the multi-measure signals (e.g., vectorized overlapping patches extracted from multi-measure images), $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are the coupled dictionaries, and θ is the maximum number of nonzero entries in common sparse representations $\{\boldsymbol{x}^p\}_{p=1}^P$. Coupled dictionary learning has been used in various applications, including image super-resolution [2, 68, 69], image reconstruction [70, 71], change detection [72], and image fusion [73, 74].

The coupled dictionary learning problem has been addressed based on single dictionary learning using a concatenated dictionary $\boldsymbol{D} = [\boldsymbol{D}_1^T \ \boldsymbol{D}_2^T]^T$ and a concatenated signal $\boldsymbol{s}^p = [(\boldsymbol{s}_1^p)^T \ (\boldsymbol{s}_2^p)^T]^T$, $p = 1, \ldots, P$ [2]. However, learning a joint concatenated dictionary is not equivalent to learning separate and coupled dictionaries. A coupled dictionary learning method based on bilevel optimization and the SGD method has been proposed in [68]. This method entails alternating optimization of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, where $\{\boldsymbol{x}^p\}_{p=1}^P$ is the sparse representations of $\{\boldsymbol{s}_1^p\}_{p=1}^P$ over $\boldsymbol{D}_1$. A semi-coupled dictionary learning method has been proposed in [75], where linear transformations of the same sparse representation, $\boldsymbol{x}_1^p$ and $\boldsymbol{x}_2^p = \boldsymbol{W}\boldsymbol{x}_1^p$, $p = 1, \ldots, P$, are used to describe the input signals ($\boldsymbol{W}$ is a linear mapping matrix). Considering that the coupled dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2^{'} = \boldsymbol{D}_2\boldsymbol{W}$ and common sparse representations are used to approximate the multi-measure signals, it can be seen that this method addresses the coupled dictionary problem (3.4). Nevertheless, the main disadvantage of the aforementioned methods is the high computational cost.

In Publication I, we have shown that the coupled dictionary problem can be addressed significantly more efficiently using alternating optimization with respect to the common sparse representations and the dictionaries. Specifically, in the proposed method, we have shown that the jointly optimal sparse representations can still be obtained based on the concatenated dictionaries and the concatenated signals using the existing sparse approximation algorithms. Moreover, the coupled dictionaries can be optimized disjointly (in parallel) based on a computationally efficient variation of the KSVD algorithm [11]. In particular, similar to KSVD, the atoms of the dictionaries are updated one by one to minimize the approximation error. However, instead of using a singular value decomposition (SVD), the atoms are optimized by solving an LS regression problem followed by a projection on the constraint set (unit sphere). This approach can be directly

extended to learning coupled dictionaries with atoms of varying sizes (different numbers of rows) and over multiple correlated input training data.

A comparison of the proposed coupled dictionary learning method with the existing algorithms based on their performances in an image super-resolution task presented in Publication I has demonstrated that our approach leads to promising results while significantly improving computational efficiencies.

3.3 CFL

CFL extracts the correlated features in the multimodal images and decomposes them into their correlated and uncorrelated components. This can be instrumental in applications such as multimodal image denoising, deblurring and fusion [24, 55]. The CFL's decomposition model takes into account two characteristics of the multimodal images:

1. The multimodal images depict the same object, tissue, scene, *etc.* Thus, they can contain overlapping (correlated) information.

2. As the images are captured using different imaging sensors, they can contain modality-specific (uncorrelated) information.

The CFL method proposed in Publication IV employs coupled dictionary learning to extract the correlated features as pairs of corresponding atoms in the dictionaries, while the uncorrelated components are captured using a Pearson correlation-based criterion. Since different imaging modalities can display the same underlying structures with varying levels of visibility, a modified coupled dictionary learning method is used where sparse representations with identical supports are used to describe the multimodal images (instead of using common sparse representation).

The CFL problem is formulated as

$$\begin{aligned} &\underset{\{\boldsymbol{D}_l,\{\boldsymbol{x}_l^p,\boldsymbol{e}_l^p\}_{p=1}^P\}_{l=1}^2}{\text{minimize}} \sum_{p=1}^{P}\left(\sum_{l=1}^{2}\left\|\boldsymbol{D}_l\boldsymbol{x}_l^p+\boldsymbol{e}_l^p-\boldsymbol{s}_l^p\right\|_2^2+\sum_{n=1}^{N}\boldsymbol{\phi}\left(\boldsymbol{e}_1^p(n),\boldsymbol{e}_2^p(n)\right)\right) \\ &\text{s.t.}\quad \text{Supp}\{\boldsymbol{x}_1^p\}=\text{Supp}\{\boldsymbol{x}_2^p\},\ \left\|\boldsymbol{x}_l^p\right\|_0\le\theta,\ \|\boldsymbol{D}_l(\cdot,k)\|_2=1,\ \forall p,\ k,\ l, \end{aligned} \tag{3.5}$$

where $\{\{\boldsymbol{s}_l^p\}_{p=1}^P\}_{l=1}^2$ are vectorized overlapping patches extracted from pairs of multimodal images, and $\{\{\boldsymbol{z}_l^p=\boldsymbol{D}_l\boldsymbol{x}_l^p\}_{p=1}^P\}_{l=1}^2$ and $\{\{\boldsymbol{e}_l^p\}_{p=1}^P\}_{l=1}^2$ represent their correlated and uncorrelated components, respectively. Moreover, $\boldsymbol{\phi}(\cdot,\cdot)$ is a cost function based on the squared Pearson correlation coefficient, defined as

$$\boldsymbol{\phi}\left(\boldsymbol{e}_1^p(n),\boldsymbol{e}_2^p(n)\right)=\left(\frac{(\boldsymbol{e}_1^p(n)-\mu_1^p)(\boldsymbol{e}_2^p(n)-\mu_2^p)}{\sigma_1^p\sigma_2^p}\right)^2,$$

where μ_l^p and σ_l^p are mean and standard deviation of $\boldsymbol{e}_l^p$, respectively.

Problem (3.5) is addressed using alternating optimization with respect to two blocks of variables: (i) the uncorrelated components $\{\{\boldsymbol{e}_l^p\}_{p=1}^P\}_{l=1,2}$, and (ii) the coupled dictionaries and the sparse representations $\{\boldsymbol{D}_l, \{\boldsymbol{x}_l^p\}_{p=1}^P\}_{l=1,2}$ (this is equivalent to optimization with respect to the correlated components).

The updated uncorrelated components are obtained using

$$\left(\boldsymbol{e}_1^p(n)\right)^+ = \frac{\rho \boldsymbol{t}_1^p(n) + \frac{\left(\boldsymbol{e}_2^p(n)-\mu_2^p\right)^2}{(\sigma_1^p)^2(\sigma_2^p)^2}\mu_1^p}{\rho + \frac{\left(\boldsymbol{e}_2^p(n)-\mu_2^p\right)^2}{(\sigma_1^p)^2(\sigma_2^p)^2}}, \quad \left(\boldsymbol{e}_2^p(n)\right)^+ = \frac{\rho \boldsymbol{t}_2^p(n) + \frac{\left(\boldsymbol{e}_1^p(n)-\mu_1^p\right)^2}{(\sigma_1^p)^2(\sigma_2^p)^2}\mu_2^p}{\rho + \frac{\left(\boldsymbol{e}_1^p(n)-\mu_1^p\right)^2}{(\sigma_1^p)^2(\sigma_2^p)^2}}, \tag{3.6}$$

where $\boldsymbol{t}_l^p = \boldsymbol{s}_l^p - \boldsymbol{D}_l \boldsymbol{x}_l^p$, $p = 1,\ldots,P$, $l = 1,2$, and the mean (μ_l^p) and standard deviation (σ_l^p) values are obtained based on the current values of uncorrelated components (this can be seen as using an Expectation-Maximization (EM) approach for the estimation of $\{\{\boldsymbol{e}_l^p\}_{p=1}^P\}_{l=1,2}$ that are dependent on the unobserved latent variables $\{\{\mu_l^p\}_{p=1}^P\}_{l=1,2}$ and $\{\{\sigma_l^p\}_{p=1}^P\}_{l=1,2}$).

Optimization with respect to the coupled dictionaries and the simultaneous sparse representations is addressed using a modified coupled dictionary learning method which is presented in the following section.

3.3.1 Simultaneous Coupled Dictionary Learning

Optimization with respect to $\{\boldsymbol{D}_l, \{\boldsymbol{x}_l^p\}_{p=1}^P\}_{l=1}^2$ is equivalent to solving the following optimization problem

$$\begin{aligned} &\underset{\boldsymbol{D}_1, \boldsymbol{D}_2, \{\{\boldsymbol{x}_l^p, \boldsymbol{e}_l^p\}_{l=1}^2\}_{p=1}^P}{\text{minimize}} \sum_{p=1}^{P} \left(\left\| \boldsymbol{D}_1 \boldsymbol{x}_1^p - \boldsymbol{w}_1^p \right\|_2^2 + \left\| \boldsymbol{D}_2 \boldsymbol{x}_2^p - \boldsymbol{w}_2^p \right\|_2^2 \right) \\ &\text{s.t.} \quad \text{Supp}\{\boldsymbol{x}_1^p\} = \text{Supp}\{\boldsymbol{x}_2^p\},\ \left\| \boldsymbol{x}_l^p \right\|_0 \leq \theta,\ \|\boldsymbol{D}_l(\cdot,k)\|_2 = 1,\ \forall p,\ k,\ l, \end{aligned} \tag{3.7}$$

where $\boldsymbol{w}_l^p = \boldsymbol{s}_l^p - \boldsymbol{e}_l^p$, $p = 1,\ldots,P$, $l = 1,2$. Similar to coupled dictionary learning, problem (3.7) can be addressed using alternating optimization with respect to the dictionaries and the sparse representations. Here, when sparse representations with identical supports $\{\{\boldsymbol{x}_l^p\}_{p=1}^P\}_{l=1,2}$ are available, the coupled dictionaries and the nonzero entries in the sparse representations can be updated disjointly, e.g., using the KSVD algorithm (unlike in coupled dictionary learning).

Optimization with respect to the sparse representations is addressed by modifying the SOMP algorithm [56] to incorporate the coupled dictionaries. Specifically, the atom selection rule of SOMP is modified so that the approximations are performed using the coupled dictionaries instead of sharing a single one. In each iteration, this modified SOMP selects a pair of coupled atoms $\{\boldsymbol{D}_1(\cdot,k^\star), \boldsymbol{D}_2(\cdot,k^\star)\}$ that minimizes the sum of the squared errors. This is formulated as

$$k^\star = \underset{k}{\text{argmin}} \left\| \boldsymbol{x}_1^p(k)\boldsymbol{D}_1(\cdot,k) - \boldsymbol{r}_1^p \right\|_2^2 + \left\| \boldsymbol{x}_2^p(k)\boldsymbol{D}_2(\cdot,k) - \boldsymbol{r}_2^p \right\|_2^2, \tag{3.8}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ represent the approximation residuals (i.e., $\boldsymbol{r}_1^p \triangleq \boldsymbol{s}_1^p - \boldsymbol{D}_1\boldsymbol{x}_1^p$ and $\boldsymbol{r}_2^p \triangleq \boldsymbol{s}_2^p - \boldsymbol{D}_2\boldsymbol{x}_2^p$). Problem (3.8) is typically solved via its equivalent maximiza-

tion problem, that is

$$k^\star = \underset{k}{\operatorname{argmax}} \left((\boldsymbol{r}_1^p)^T \boldsymbol{D}_1(\cdot,k)\right)^2 + \left((\boldsymbol{r}_2^p)^T \boldsymbol{D}_2(\cdot,k)\right)^2. \tag{3.9}$$

3.4 Convolutional CFL

An extension of CFL to the CSC model has been proposed in Publication V, where L multimodal images $\{\boldsymbol{s}_l\}_{l=1}^L$ are decomposed into their correlated and uncorrelated components. The correlated components are estimated using L coupled dictionaries with J convolutional filters $(\{\{\boldsymbol{d}_j^{\mathrm{z}(l)}\}_{j=1}^J\}_{l=1}^L)$ and common convolutional sparse representations $\{\boldsymbol{x}_j^{\mathrm{z}}\}_{j=1}^J$. Here, the differences in the visibility levels of coupled features in different modalities are captured in the norm of the coupled convolutional filters. The uncorrelated components are estimated using a common dictionary with K convolutional filters $(\{\boldsymbol{d}_k^{\mathrm{e}}\}_{k=1}^K)$ and L separate convolutional sparse representations $\{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^K\}_{l=1}^L$. This is formulated as the following optimization problem

$$\underset{\substack{\{\{\boldsymbol{d}_j^{\mathrm{z}(l)}\}_{j=1}^J\}_{l=1}^L, \{\boldsymbol{d}_k^{\mathrm{e}}\}_{k=1}^K, \\ \{\boldsymbol{x}_j^{\mathrm{z}}\}_{j=1}^J, \{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^K\}_{l=1}^L}}{\text{minimize}} \frac{1}{2}\sum_{l=1}^L \left\| \sum_{j=1}^J \boldsymbol{d}_j^{\mathrm{z}(l)} * \boldsymbol{x}_j^{\mathrm{z}} + \sum_{k=1}^K \boldsymbol{d}_k^{\mathrm{e}} * \boldsymbol{x}_k^{\mathrm{z}(l)} - \boldsymbol{s}_l \right\|_2^2 + \lambda_1 \sum_{j=1}^J \|\boldsymbol{x}_j^{\mathrm{z}}\|_1$$

$$+ \lambda_2 \sum_{l=1}^L \sum_{k=1}^K \|\boldsymbol{x}_k^{\mathrm{e}(l)}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{d}_j^{\mathrm{z}(l)}\| \le 1,\ \|\boldsymbol{d}_k^{\mathrm{e}}\| \le 1,\ \forall j,k,l. \tag{3.10}$$

The third term in the objective function of (3.10) enforces an element-wise sparsity in $\{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^K\}_{l=1}^L$. That means $\{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^K\}_{l=1}^L$ are sparse also along different modalities (i.e., arrays $[\boldsymbol{x}_k^{\mathrm{e}(1)}(p),\ldots,\boldsymbol{x}_k^{\mathrm{e}(L)}(p)]$, $k=1,\ldots,K$, $p=1,\ldots,P$, are sparse). When $[\boldsymbol{x}_k^{\mathrm{e}(1)}(p),\ldots,\boldsymbol{x}_k^{\mathrm{e}(L)}(p)]$ has only one nonzero entry, it means that filter $\boldsymbol{d}_k^{\mathrm{e}}$ is used to represent a feature only in one of the multimodal images (at pixel p), indicating a modality-specific feature. The overlapping nonzero entries indicate a shared feature.

In Publication V, problem (3.10) has been addressed based on consensus ADMM.

3.4.1 Convolutional SSA

In Publication VI, we proposed a convolutional SSA method that can be used to complement the convolutional CFL problem in (3.10) by replacing joint CSC (using common sparse representations). The convolutional SSA problem is

written as follows

$$\underset{\{\{\boldsymbol{x}_k^l\}_{k=1}^K\}_{l=1}^L}{\text{minimize}} \frac{1}{2}\sum_{l=1}^{L}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k * \boldsymbol{x}_k^l - \boldsymbol{s}_l\right\|_2^2 + \lambda\sum_{l=1}^{L}\sum_{k=1}^{K}\left\|\boldsymbol{x}_k^l\right\|_0 \tag{3.11}$$

$$\text{s.t.} \quad \text{Supp}\left(\boldsymbol{x}_k^l\right) = \text{Supp}\left(\boldsymbol{x}_k^{l'}\right),\ l, l' = 1,\ldots,L,\ k = 1,\ldots,K.$$

A convex relaxation of problem (3.11) based on the $\ell_{2,1}$-norm can be written as

$$\underset{\{\{\boldsymbol{x}_k^l\}_{k=1}^K\}_{l=1}^L}{\text{minimize}} \frac{1}{2}\sum_{l=1}^{L}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k * \boldsymbol{x}_k^l - \boldsymbol{s}_l\right\|_2^2 + \lambda\sum_{k=1}^{K}\|\mathcal{X}_k\|_{2,1}, \tag{3.12}$$

where $\mathcal{X}_k(n,\cdot) \triangleq [\boldsymbol{x}_k^1(n) \cdots \boldsymbol{x}_k^L(n)]$, $n = 1,\ldots,N$. Using ADMM, problem (3.12) can be broken into two simpler subproblems: a convolutional regression problem which can be addressed using [51], and the following structured sparse approximation problem

$$\underset{\{\{\boldsymbol{x}_k^l\}_{k=1}^K\}_{l=1}^L}{\text{minimize}} \frac{\rho}{2}\sum_{l=1}^{L}\sum_{k=1}^{K}\left\|\boldsymbol{x}_k^l - \boldsymbol{w}_k^l\right\|_2^2 + \lambda\sum_{k=1}^{K}\|\mathcal{X}_k\|_{2,1}. \tag{3.13}$$

Since the $\ell_{2,1}$-norm is a sum of the Euclidean norms of the rows, the solution to (3.13) can be found using

$$\left([\boldsymbol{x}_k^1(n) \cdots \boldsymbol{x}_k^L(n)]\right)^\star = \text{prox}_{\frac{\lambda}{\rho}\|\cdot\|_2}\left([\boldsymbol{w}_k^1(n) \cdots \boldsymbol{w}_k^L(n)]\right),$$

$$k = 1,\ldots,K,\ n = 1,\ldots,N, \tag{3.14}$$

where

$$\text{prox}_{\tau\|\cdot\|_2}(\boldsymbol{a}) = \left(1 - \frac{\tau}{\max(\|\boldsymbol{a}\|_2, \tau)}\right)\boldsymbol{a}. \tag{3.15}$$

The use of the $\ell_{2,1}$-norm regularization enforces a row-sparse structure which can be alternatively achieved using the $\ell_{\infty,1}$-norm regularization. The resulting optimization problem can be addressed using a similar ADMM approach where the proximal operator for the ℓ_∞ norm is used instead of $\text{prox}_{\tau\|\cdot\|_2}(\cdot)$.

A row-sparse and element-sparse structure can be enforced by adding an element-wise sparsity regularization term to the objective function of (3.11)

$$\underset{\{\{\boldsymbol{x}_k^l\}_{k=1}^K\}_{l=1}^L}{\text{minimize}} \frac{1}{2}\sum_{l=1}^{L}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k * \boldsymbol{x}_k^l - \boldsymbol{s}_l\right\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\mathcal{X}_k\|_{2,1} + \lambda_2\sum_{k=1}^{K}\|\mathcal{X}_k\|_{1,1}. \tag{3.16}$$

Problem (3.16) can be solved by replacing $\text{prox}_{\frac{\lambda}{\rho}\|\cdot\|_2}(\cdot)$ with $\text{prox}_{\frac{\lambda_1}{\rho}\|\cdot\|_2 + \frac{\lambda_1}{\rho}\|\cdot\|_1}(\cdot)$ in (3.14).

The convolutional SSA method proposed in Publication VI can be straightforwardly extended to the case of coupled dictionaries, and further to convolutional CFL using existing CDL algorithms (the coupled dictionaries can be optimized independently using their corresponding signals and their sparse representations).

3.5 Experimental Results

In this section, we provide representative experimental results for the proposed coupled dictionary learning (Publication I) and CFL (Publications IV-VI) methods.

3.5.1 Coupled Dictionary Learning Results

We evaluate the proposed coupled dictionary learning method by incorporating it in the image super-resolution algorithm of [68] and comparing the results to those obtained using the original method. The image super-resolution method of [68] uses coupled dictionaries learned over gradient features extracted from low- and high-resolution 5×5 image patches to recover the high-resolution image from the observed low-resolution input. A coupled dictionary learning method based on bilevel optimization and the SGD algorithm is proposed in [68]. We use a dataset of $10,000$ image patches to train coupled dictionaries with 512 atoms using both methods. Sparse approximation is performed using convex relaxation with ℓ_1-norm regularization parameter $\lambda = 0.05$.

Fig. 3.1 shows image super-resolution results obtained based on the two coupled dictionary methods compared. Upsized images obtained using bicubic interpolation [76] are also presented as a reference for performance gain comparison. As can be seen, the coupled dictionary learning methods yield similar results. However, the proposed method significantly reduces the computational cost. Particularly, the proposed method and the method of [68] spent 8.3 and 120.2 seconds to perform coupled dictionary learning over the training dataset, respectively.

3.5.2 CFL Results

We investigate the effectiveness of the proposed CFL (Publication IV) and convolutional CFL (Publication V) methods using multimodal *computed tomography* (CT) and *magnetic resonance* (MR) images collected from *The Whole Brain Atlas* database [77]. The proposed CFL method based on convolutional SSA (Publication VII) is applied to RGB-NIR images taken from *EPFL RGB-NIR Scene Dataset* [78] (the characteristics of the imaging modalities will be discussed in the next chapter). Dictionary atoms are of size 8×8 in all experiments.

Figs. 3.2-3.4 illustrate coupled dictionaries learned using the proposed CFL methods, where correlation between the corresponding atoms (representing the correlated multimodal features) can be clearly observed.

Figure 3.1. Image super-resolution results: (a) original 512 × 512 image; (b) down-scaled 128 × 128 image; (c) upscaled image using bicubic interpolation [76]; (d) the method of [68]; (e) the method of [68] using the coupled dictionary learning method proposed in Publication I.

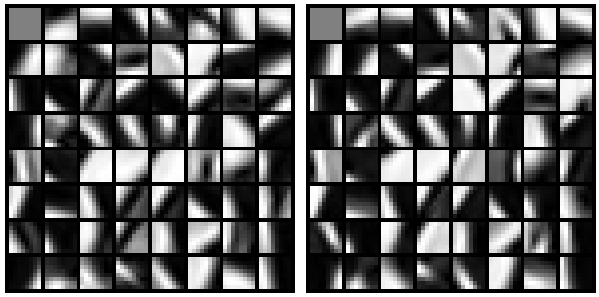


Figure 3.2. Coupled dictionaries composed of $K = 64$ atoms obtained for multimodal MR-CT images using the CFL method proposed in Publication IV.

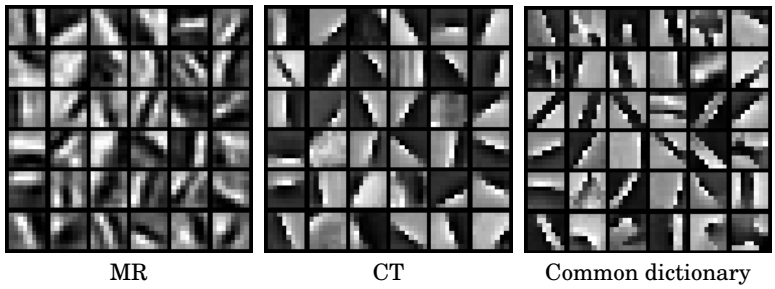


Figure 3.3. Coupled convolutional dictionaries ($K = 36$) and common dictionary of ($L = 36$ filters used to describe modality-specific features) obtained for multimodal MR-CT images using the convolutional CFL method proposed in Publication V.

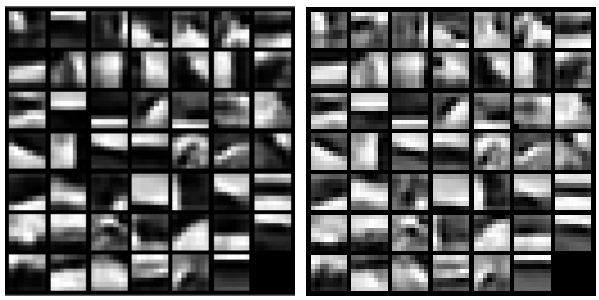


Figure 3.4. Coupled dictionaries composed of $K = 48$ convolutional filters obtained for NIR-RGB images using the method proposed in Publication VI.

Examples of multimodal images in the experiments and their decomposition components (correlated and uncorrelated) obtained using the proposed CFL methods are presented in Figs. 3.5-3.7. Since the CSC model cannot effectively describe the low-resolution components (i.e., the *base* layer) of the signals, the CSC-based CFL methods in Publications V and VI are applied to the high-resolution components only.

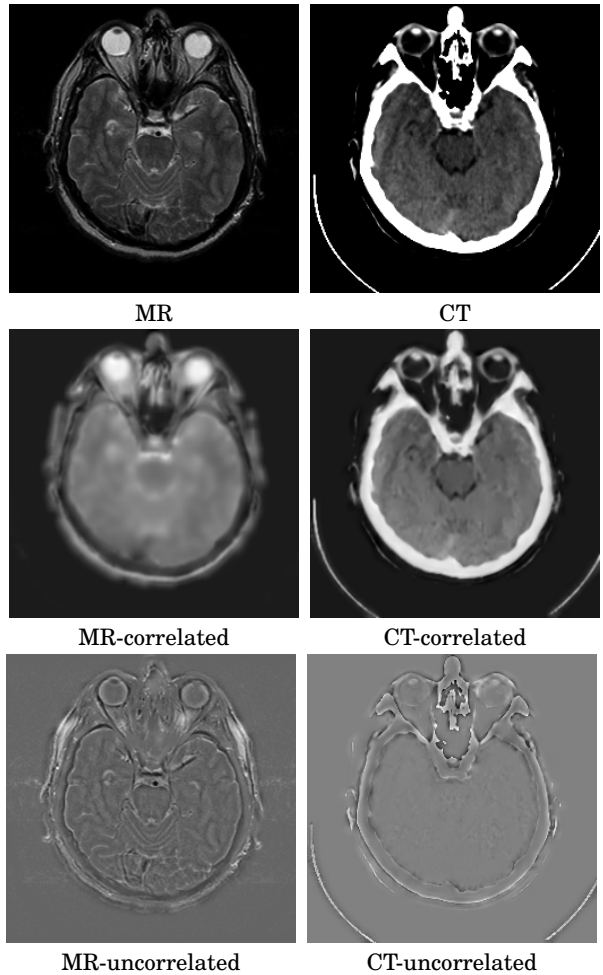


Figure 3.5. Decomposition components obtained for multimodal MR-CT images using the CFL method proposed in Publication IV.

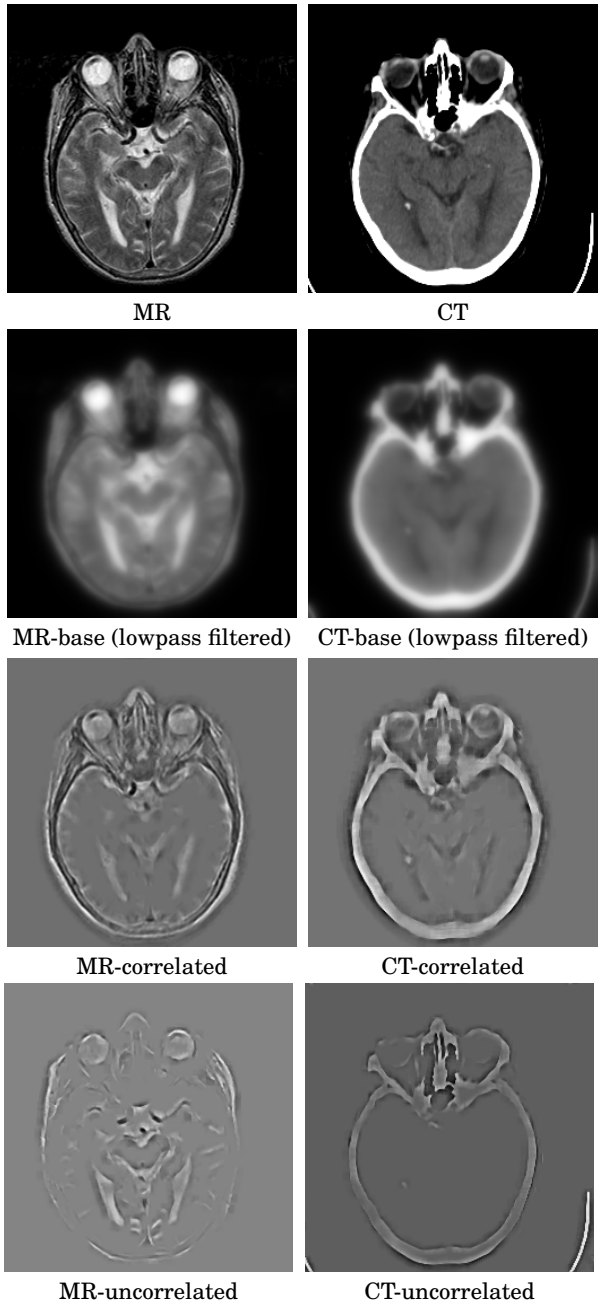


Figure 3.6. Decomposition components obtained for multimodal MR-CT images using the convolutional CFL method proposed in Publication V.

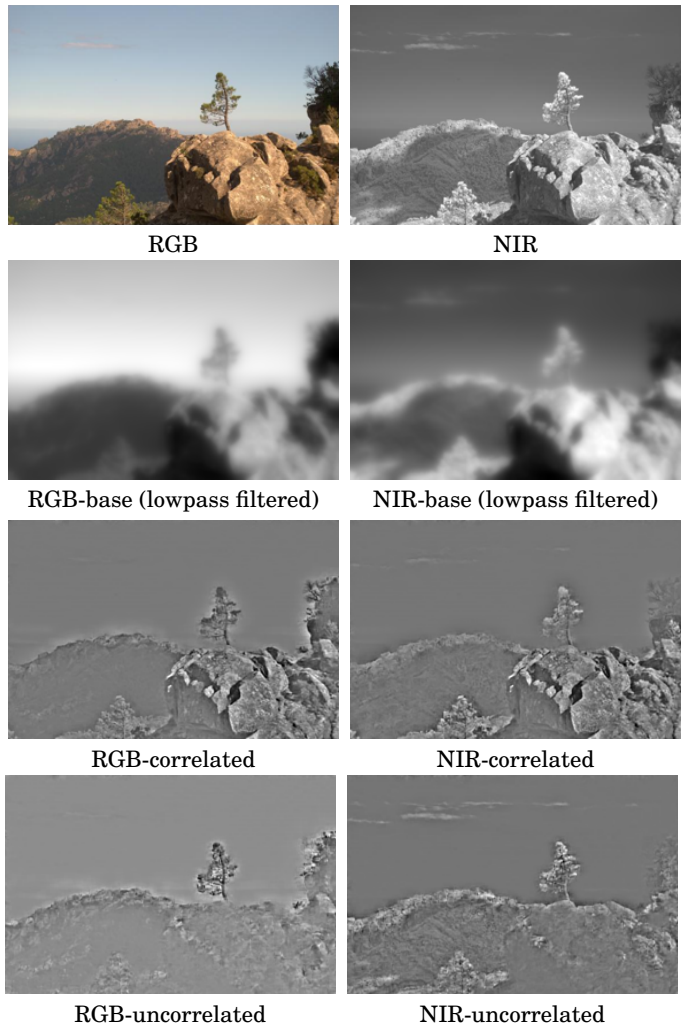


Figure 3.7. Decomposition components obtained for multimodal RGB-NIR images using the convolutional SSA-based CFL method proposed in Publication VI.

4. Multimodal Image Fusion

Multimodal image fusion seeks to integrate relevant information from images acquired with different imaging sensors into a single image without introducing noise or artifacts. Applications of multimodal image fusion include surveillance [79–81], remote sensing [82–84] and medical imaging [85–87]. In surveillance applications, a fusion of infrared and visible light images is used to enhance object detection and improve night vision [79, 88]. NIR images are used to improve the contrast-resolution of RGB images, for example, taken from vegetation scenes or in low-visibility atmospheric conditions such as fog or haze [89]. In satellite imaging, panchromatic images with high spatial resolutions and multispectral images with high spectral resolutions are combined to produce more informative and enhanced fused images [84, 90]. Multimodal medical image fusion combines the information acquired using various sensors. Anatomical imaging techniques, for example, CT and MR imaging, provide high-resolution images of internal organs and tissues. Functional imaging mechanisms such as *single-photon emission computed tomography* (SPECT) and *positron emission tomography* (PET) measure the biological activity of specific areas inside the body. This variety of information can be fused into a single image for joint analysis and easier visualization [85–87].

A typical approach for addressing the multimodal image fusion problem consists of extraction of multiscale or morphologically distinct features in the input images and then using them for generating a joint reinforced representation based on a *fusion rule*. Feature extraction is commonly performed using deterministic mathematical models such as multiscale transforms (wavelets, curvelets, shearlets, etc.) [91–94]. Other techniques employed for the same purpose include subspace learning [95], sparse representations and dictionary learning [96–100], and CNN [101, 102].

A general assumption incorporated in all aforementioned methods is that features with similar structural characteristics (e.g., resolution scale) convey correlated information. However, due to the varying (and often complementary) characteristics of multimodal images, this assumption may not be valid in many cases. For instance, in MR imaging, soft tissues (e.g., fat and liquid) are captured with a higher resolution, while the details of hard tissues (e.g., bones and

implants) are reflected more effectively in CT images. In infrared-visible images, the details in each image depict fundamentally different types of information. Thus, applying a fusion rule (e.g., based on binary selection or weighted averaging) to features representing distinct objects or characteristics (but with similar structural properties) can lead to degradation or loss of information.

The multimodal image fusion methods proposed in Publications IV-VI employ the CFL model for extracting correlated features in multimodal images instead of using conventional deterministic feature-extraction techniques. The fusion is performed using only the correlated components of the multimodal images, while the modality-specific (unique) components are preserved in the final fused images. In the following sections, we present the proposed CFL-based multimodal fusion methods.

4.1 Multimodal Image Fusion via CFL

The multimodal image fusion method in Publication IV first decomposes vectorized overlapping patches $\{\boldsymbol{s}_1^p\}_{p=1}^P$ and $\{\boldsymbol{s}_2^p\}_{p=1}^P$ (extracted from the multimodal input images) into their correlated and uncorrelated components by obtaining coupled dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, sparse representations with identical supports $\{\boldsymbol{x}_1^p\}_{p=1}^P$ and $\{\boldsymbol{x}_2^p\}_{p=1}^P$, and uncorrelated components $\{\boldsymbol{e}_1^p\}_{p=1}^P$ and $\{\boldsymbol{e}_2^p\}_{p=1}^P$ using the proposed CFL algorithm.

The sparse representations are combined using the coefficients with the largest absolute values. This ensures that correlated features (the corresponding atoms in the coupled dictionaries) with the highest visibility levels are used in the fused image. Fused correlated components $\{\boldsymbol{z}_{\mathrm{F}}^p\}_{p=1}^P$ are obtained using

$$\boldsymbol{z}_{\mathrm{F}}^p = \boldsymbol{D}_1 \boldsymbol{x}_{1(\mathrm{F})}^p + \boldsymbol{D}_2 \boldsymbol{x}_{2(\mathrm{F})}^p,\ p = 1,\ldots,P, \tag{4.1}$$

where fused sparse representations $\{\boldsymbol{x}_{1(\mathrm{F})}^p\}_{p=1}^P$ and $\{\boldsymbol{x}_{2(\mathrm{F})}^p\}_{p=1}^P$ are found as

$$\begin{aligned} \boldsymbol{x}_{1(\mathrm{F})}^p(n) &= \begin{cases} \boldsymbol{x}_1^p(n), & \text{if } |\boldsymbol{x}_1^p(n)| \geq |\boldsymbol{x}_2^p(n)| \\ 0, & \text{otherwise} \end{cases}, \quad \forall p,n, \\ \boldsymbol{x}_{2(\mathrm{F})}^p(n) &= \begin{cases} \boldsymbol{x}_2^p(n), & \text{if } |\boldsymbol{x}_2^p(n)| > |\boldsymbol{x}_1^p(n)| \\ 0, & \text{otherwise} \end{cases}, \quad \forall p,n. \end{aligned} \tag{4.2}$$

The uncorrelated components, $\{\boldsymbol{e}_1^p\}_{p=1}^P$ and $\{\boldsymbol{e}_2^p\}_{p=1}^P$, are transferred to the final fused image unaltered (to preserve the modality-specific information). The fused patches are then found using $\boldsymbol{s}_{\mathrm{F}}^p = \boldsymbol{z}_{\mathrm{F}}^p + \boldsymbol{e}_1^p + \boldsymbol{e}_2^p,\ p = 1,\ldots,P$. Finally, the fused image is reconstructed using the fused patches.

Publication IV presents extensive experimental evaluation results using multiple multimodal image datatsets, including four different combinations of medical imaging techniques, as well as infrared and visible images. The presented experimental results demonstrated that the proposed method leads to improved fusion

of local intensities and texture information compared to other state-of-the-art methods.

4.1.1 Fusion of Greyscale and Color Images

Multimodal image fusion can involve fusion of a color image with a greyscale one. For instance, functional medical images (e.g., PET) are usually presented in a color code, while anatomical medical images are available in greyscale. A standard approach for dealing with the fusion of color images is to convert them to the YCbCr (or YUV) color-space. In this new color-space, component Y (i.e., luminance) represents the grey-scale version of the image, which is used for fusion. As the color information is derived from the functional images only, the color components (Cb and Cr) are directly incorporated into the final fused image. Fig. 4.1 illustrates the block diagram of the greyscale and color image fusion method.

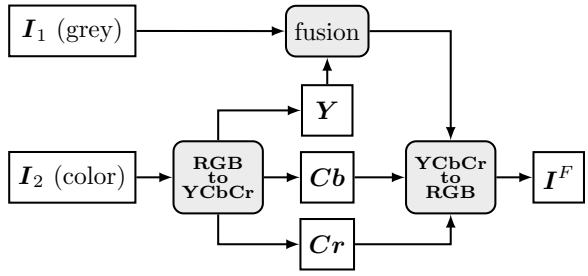


Figure 4.1. Block-diagram of the grey-scale and color image fusion method.

4.1.2 Multimodal Image Fusion via Convolutional CFL

In Publication V, a convolutional extension of the CFL-based fusion method proposed in Publication IV has been presented. In this method, input images $\{\boldsymbol{s}_l\}_{l=1}^{L}$ are first decomposed into low-resolution (base) layers $\{\boldsymbol{s}_l^{\mathrm{b}}\}_{l=1}^{L}$ and high-resolution (details) layers $\{\boldsymbol{s}_l^{\mathrm{d}}\}_{l=1}^{L}$ using low-pass filtering. The details-layers are decomposed into correlated and uncorrelated components using the proposed convolutional CFL algorithm (obtaining coupled dictionaries $\{\{\boldsymbol{d}_j^{\mathrm{z}(l)}\}_{j=1}^{J}\}_{l=1}^{L}$, common dictionary $\{\boldsymbol{d}_k^{\mathrm{e}}\}_{k=1}^{K}$, joint sparse representations $\{\boldsymbol{z}_j^{\mathrm{z}}\}_{j=1}^{J}$ and separate (unique) sparse representations $\{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^{K}\}_{l=1}^{L}$).

Dictionary of fused coupled features $\{\boldsymbol{d}_j^{\mathrm{F}}\}_{j=1}^{J}$ is formed based on the convolutional filters in $\{\{\boldsymbol{d}_j^{\mathrm{z}(l)}\}_{j=1}^{J}\}_{l=1}^{L}$ with largest variances (used as a measure of visual significance).

The fused sparse representations $\{\boldsymbol{x}_k^{\mathrm{F}}\}_{k=1}^{K}$ are found by combining redundant sparse representations $\{\{\boldsymbol{x}_k^{\mathrm{e}(l)}\}_{k=1}^{K}\}_{l=1}^{L}$ using maximum-absolute-value rule (similar to (4.2)). This ensures that the uncorrelated features and the shared features with the most significant representations are incorporated in the fused image.

The fused details-layer $\boldsymbol{s}_{\mathrm{F}}^{\mathrm{d}}$ is then reconstructed using

$$\boldsymbol{s}_{\mathrm{F}}^{\mathrm{d}} = \sum_{j=1}^{J} \boldsymbol{d}_k^{\mathrm{F}} * \boldsymbol{x}_j^{\mathrm{z}} + \sum_{k=1}^{K} \boldsymbol{d}_k^{\mathrm{e}} * \boldsymbol{x}_k^{F}.$$

The fused base-layer ($\boldsymbol{s}_{\mathrm{F}}^{\mathrm{b}}$) is obtained based on a compromise between maintaining the contrast resolutions of the fused details-layer and allowing the highest local intensities (since the pixel values in the standard images are limited to take values between 0 and 1, a plain addition of the fused components can result in saturation and loss of information). The final fused image $\boldsymbol{s}^{\mathrm{F}}$ is found using

$$\boldsymbol{s}^{\mathrm{F}} = \boldsymbol{s}_{\mathrm{F}}^{\mathrm{d}} + \boldsymbol{s}_{\mathrm{F}}^{\mathrm{b}}. \tag{4.3}$$

The experimental evaluations using medical and infrared-visible image datasets, presented in Publication V, demonstrated improved fusion performances regarding preserving the details and local intensities compared to state-of-the-art multimodal image fusion algorithms.

4.2 NIR-RGB Image Fusion based on Convolutional SSA

NIR imaging can provide higher contrast resolutions, for example, in low-visibility atmospheric conditions such as fog or haze. Therefore, NIR-RGB image fusion has been used for outdoor image enhancement [21, 89]. In the following, we present the NIR-RGB image fusion method proposed in Publication VI, which is based on convolutional SSA and CDL.

The NIR images (denoted as $\boldsymbol{s}_{\mathrm{nir}}$) are available in greyscale. Thus, they can be fused with the greyscale version of the RGB images (denoted as $\boldsymbol{s}_{\mathrm{rgb}}$). Hence, the RGB images are first converted to a color space (e.g., YCbCr), where the greyscale component ($\boldsymbol{s}_{\mathrm{grey}}$) is separated from the color components. The images are then decomposed into their base-layers $\boldsymbol{s}_{\mathrm{nir}}^{\mathrm{b}}$ and $\boldsymbol{s}_{\mathrm{grey}}^{\mathrm{b}}$, and details-layers $\boldsymbol{s}_{\mathrm{nir}}^{\mathrm{d}}$ and $\boldsymbol{s}_{\mathrm{grey}}^{\mathrm{d}}$ using lowpass filtering. The fusion is performed over the details-layers (high-resolution components).

Using the proposed convolutional SSA method and coupled dictionaries $\{\boldsymbol{d}_k^{\mathrm{nir}}\}_{k=1}^{K}$ and $\{\boldsymbol{d}_k^{\mathrm{grey}}\}_{k=1}^{K}$, the sparse representations $\{\boldsymbol{x}_k^{\mathrm{nir}}\}_{k=1}^{K}$ and $\{\boldsymbol{x}_k^{\mathrm{grey}}\}_{k=1}^{K}$ are obtained for $\boldsymbol{s}_{\mathrm{nir}}^{\mathrm{d}}$ and $\boldsymbol{s}_{\mathrm{grey}}^{\mathrm{d}}$, respectively. Coupled dictionaries for NIR and greyscale images $\{\boldsymbol{d}_k^{\mathrm{nir}}\}_{k=1}^{K}$ and $\{\boldsymbol{d}_k^{\mathrm{grey}}\}_{k=1}^{K}$ are learned beforehand using a training dataset of NIR-RGB images. The fused sparse representations $\{\boldsymbol{x}_k^{\mathrm{nir(F)}}\}_{k=1}^{K}$ and $\{\boldsymbol{x}_k^{\mathrm{grey(F)}}\}_{k=1}^{K}$ are obtained using the max-absolute-value fusion rule (allowing only the most significant coefficients at each entry). The fused greyscale details-layer $\boldsymbol{s}_{\mathrm{grey}}^{\mathrm{d(F)}}$ is then reconstructed using

$$\boldsymbol{s}_{\mathrm{grey}}^{\mathrm{d(F)}} = \sum_{k=1}^{K} \boldsymbol{x}_k^{\mathrm{nir(F)}} * \boldsymbol{d}_k^{\mathrm{nir}} + \sum_{k=1}^{K} \boldsymbol{x}_k^{\mathrm{grey(F)}} * \boldsymbol{d}_k^{\mathrm{grey}}. \tag{4.4}$$

The fused greyscale image $\boldsymbol{s}^{\mathrm{F}}_{\mathrm{grey}}$ is formed using $\boldsymbol{s}^{\mathrm{d(F)}}_{\mathrm{grey}}$ and the base-layer of the greyscale image

$$\boldsymbol{s}^{\mathrm{F}}_{\mathrm{grey}} = \boldsymbol{s}^{\mathrm{b}}_{\mathrm{grey}} + \boldsymbol{s}^{\mathrm{d(F)}}_{\mathrm{grey}}. \tag{4.5}$$

Finally, the YCbCR image with $\boldsymbol{s}^{\mathrm{F}}_{\mathrm{grey}}$ as the intensity layer and the color components of the RGB image is converted back to RGB format to generate the final fused image.

4.3 Experimental Results

In this section, we compare the CFL-based multimodal image fusion methods proposed in Publications IV-VI with state-of-the-art algorithms. The experiments are conducted using the following datasets:

1. Multimodal medical image fusion: multimodal CT-MR and MR-PET images taken from *The Whole Brain Atlas* database [77].

2. IR-VL image fusion: a pair of IR-VL images taken from `https://github.com/hli1221/imagefusion_resnet50/tree/master/IV_images`.

3. RGB-NIR image fusion: outdoor images taken from *EPFL RGB-NIR Scene Dataset* [78].

The comparisons are conducted using two medical image fusion methods: a method based on the convolutional neural networks and Laplacian pyramids (CNN) [103] and a method based on Laplacian redecomposition (LRD) [93]. We also use two IR-VL image fusion methods: a method that incorporates a hierarchical Bayesian model (Bayes) [88] and a method based on deep learning (Resnet) [101]. We compare the proposed RGB-NIR image fusion method to a method based on the top-hat transform (Top-Hat) [89]. For all methods, we use the default parameters tuned by the authors of the corresponding papers.

4.3.1 Multimodal Medical Image Fusion Results

Figs. 4.2 and 4.3 show results for CT-MR and MR-PET image fusion using the compared medical image fusion methods. As can be seen, the CNN-based fusion method results in a significant loss of local intensities in both experiments. The LRD method leads to a loss of high-resolution details and results in color distortions (see Fig. 4.3). Both proposed fusion methods provide improved results in terms of preserving the local intensities and high-resolution details.

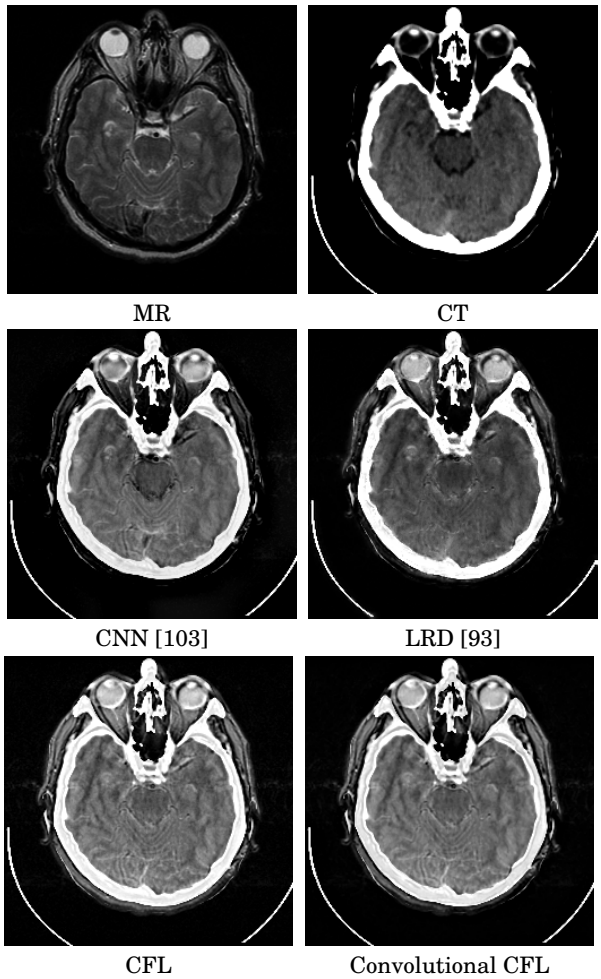


Figure 4.2. MR-CT image fusion results obtained using different methods.

4.3.2 Visible-Light and Infrared Image Fusion Results

A pair of multimodal VL-IR images and their fusion results obtained using different methods are shown in Fig. 4.4. As can be seen, similar to the case of multimodal medical image fusion, the proposed fusion methods lead to enhanced contrast resolution and overall visibility.

4.3.3 RGB-NIR Image Fusion Results

Figs. 4.5 and 4.6 show examples of NIR and RGB images and their fusion results obtained using the method proposed in Publication IV and the method of [89]. Improvements obtained by using the proposed method can be clearly observed.

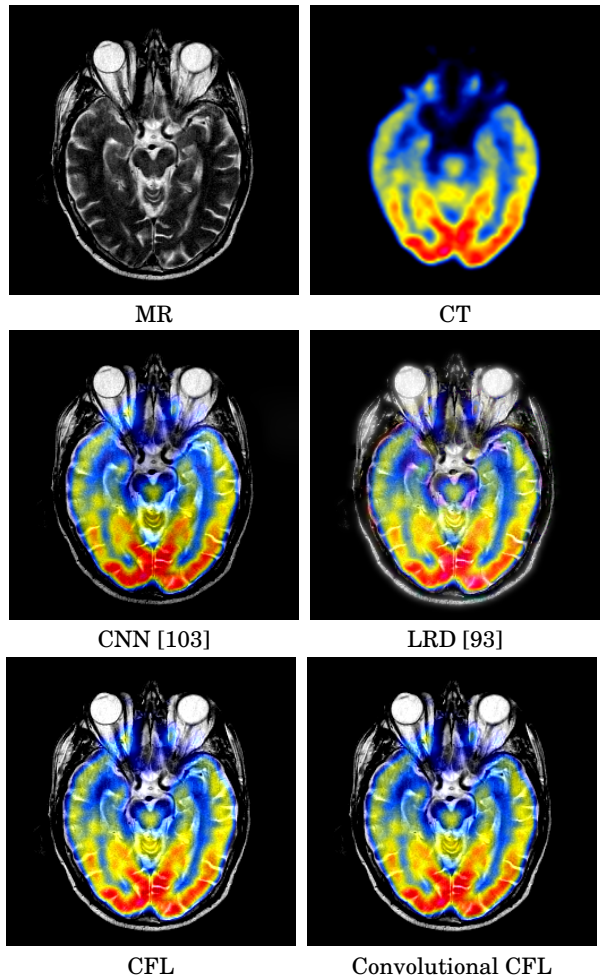


Figure 4.3. MR-PET image fusion results obtained using different methods.

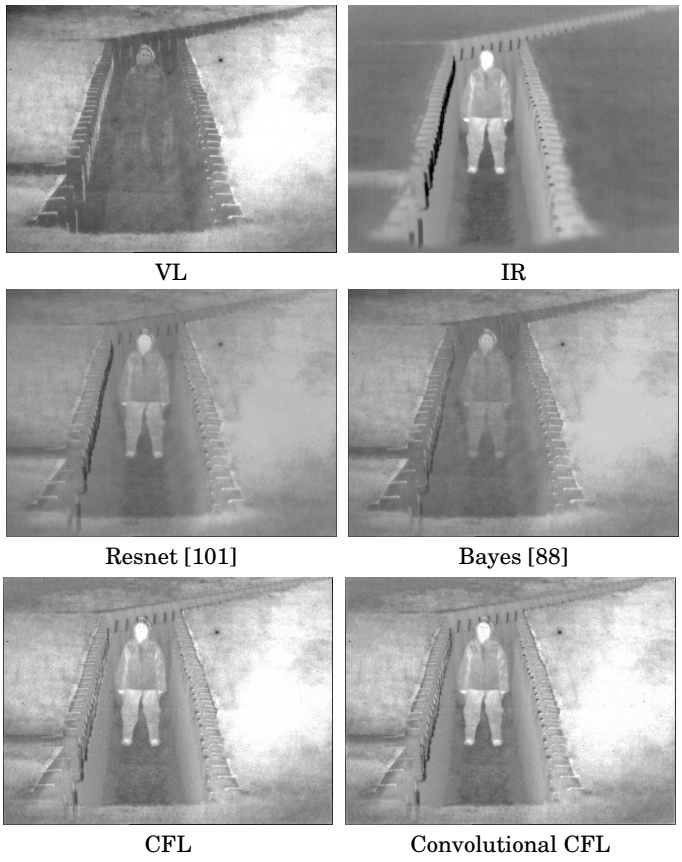
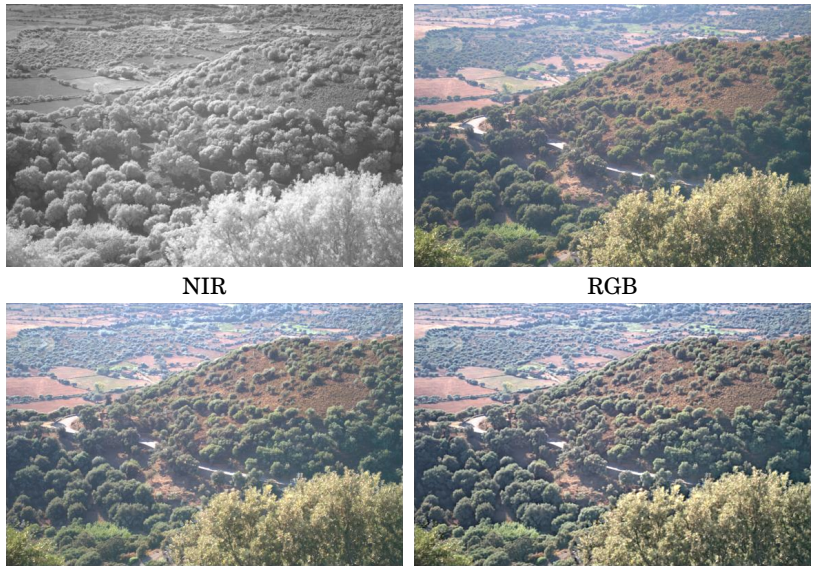


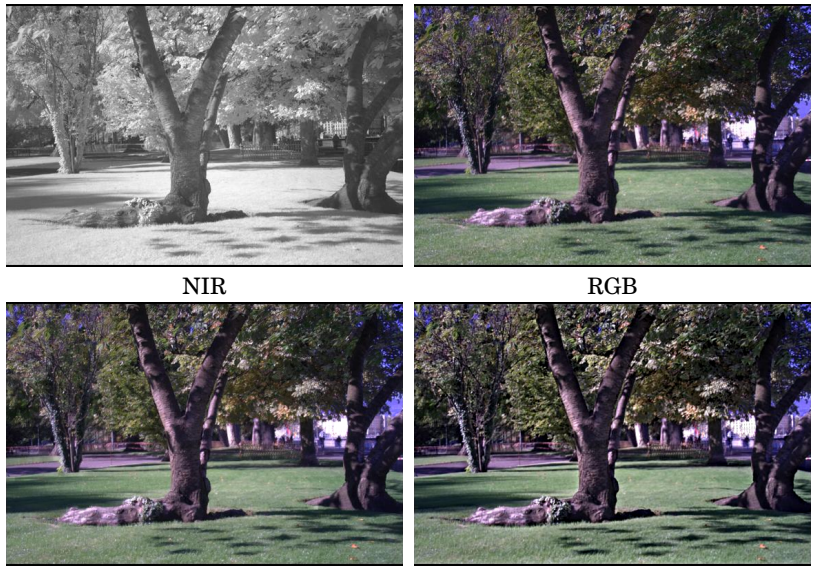
Figure 4.4. VL-IR image fusion results obtained using different methods.



NIR RGB

Top-Hat transform [89] Convolutional SSA

Figure 4.5. NIR-RGB image fusion results.



NIR RGB

Top-Hat transform [89] Convolutional SSA

Figure 4.6. NIR-RGB image fusion results.

5. Conclusions

In this thesis, we have developed computationally efficient CSC and CDL algorithms that can be used in large-scale signal and image processing applications. Specifically, we have proposed a novel convolutional LS regression method that improves the efficiency of existing ADMM-based CSC and CDL algorithms. Additionally, we have proposed an efficient approximate ADMM-based OCDL method suitable for applications that require learning large dictionaries over high-dimensional signals.

Furthermore, we have presented new methods and developed computationally efficient algorithms for learning correlated features in multi-measure and multimodal signals based on sparse approximation and dictionary learning frameworks. We have also developed extensions and variations of the proposed CFL method based on the CSC model. The presented CFL methods can be potentially used in various signal and image processing applications that require a joint analysis of multiple correlated data. Specifically, we have proposed multimodal image fusion methods based on the proposed CFL algorithms. We used the learned coupled features to generate unified and reinforced (fused) images. We have addressed multimodal medical, IR-IV, and NIR-RGB image fusion problems.

Image fusion is a task where dictionary learning and sparse representations remain superior to deep learning-based methods. This is largely due to the absence of naturally fused images available for end-to-end supervised learning. Instead, fused images are synthesized by combining the information in multiple input images. Moreover, access to training data can be limited in many cases, for example, in medical image fusion.

In contrast, dictionary learning allows visual features to be learned as atoms of the dictionaries using very small datasets or even a single image. Based on the sparse representation and dictionary learning model, these learned visual features can be used as building blocks for constructing the fused image. Additionally, the magnitude of the sparse coefficients can be interpreted as a measure of the significance or visibility level of the visual features. By relying on these interpretations of the sparse model, the image fusion task can be addressed more effectively and efficiently with a smaller training dataset, fewer

parameters, shorter training time, etc.

Representative experimental results obtained using the proposed algorithms have been provided at the end of each chapter. The effectiveness of the proposed methods has been evaluated based on comparisons with state-of-the-art algorithms.

5.1 Potential Future Works

The CFL methods proposed in this thesis are specifically applicable to signals with grid-like structures, such as images and time series. It could be interesting to consider extending the CFL model to graph signals, for example, based on the existing graph dictionary learning algorithms [104, 105].

We proposed general image fusion methods with an emphasis on algorithmic simplicity to demonstrate the effectiveness of the proposed CFL algorithm. More comprehensive CFL-based image fusion methods incorporating the imaging modalities' characteristics can enhance image fusion performance.

References

[1] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[2] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[3] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[5] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[6] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, (Pacific Grove, CA, USA), pp. 40–44, Nov. 1993.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.

[8] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 629–654, 2008.

[9] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[10] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process.*, vol. 5, (Phoenix, AZ, USA), pp. 2443–2446, Mar. 1999.

[11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, 2006.

[12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Mach. Learn.*, (Montreal, Quebec, Canada), pp. 689–696, June 2009.

[13] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2528–2535, June 2010.

[14] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Portland, OR, USA), pp. 391–398, June 2013.

[15] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, (Boston, MA, USA), pp. 5135–5143, June 2015.

[16] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, 2016.

[17] V. Papyan, Y. Romano, M. Elad, and J. Sulam, "Convolutional dictionary learning via local processing," in *Proc. IEEE Int. Conf. Comput. Vis.*, (Venice, Italy), pp. 5306–5314, Oct. 2017.

[18] A. Cogliati, Z. Duan, and B. Wohlberg, "Piano transcription with convolutional sparse lateral inhibition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 392–396, 2017.

[19] P. Jao, L. Su, Y. Yang, and B. Wohlberg, "Monaural music source separation using convolutional sparse coding," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 11, pp. 2158–2170, 2016.

[20] P. Bao, W. Xia, K. Yang, W. Chen, M. Chen, Y. Xi, S. Niu, J. Zhou, H. Zhang, H. Sun, Z. Wang, and Y. Zhang, "Convolutional sparse coding for compressed sensing CT reconstruction," *IEEE Trans. Med. Imaging*, vol. 38, no. 11, pp. 2607–2619, 2019.

[21] X. Hu, F. Heide, Q. Dai, and G. Wetzstein, "Convolutional sparse coding for RGB+NIR imaging," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1611–1625, 2018.

[22] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, (Santiago, Chile), pp. 1823–1831, Dec. 2015.

[23] M. Li, Q. Xie, Q. Zhao, W. Wei, S. Gu, J. Tao, and D. Meng, "Video rain streak removal by multiscale convolutional sparse coding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, (Salt Lake City, UT, USA), pp. 6644–6653, June 2018.

[24] F. G. Veshki, N. Ouzir, S. A. Vorobyov, and E. Ollila, "Multimodal image fusion via coupled feature learning," *Signal Process.*, vol. 200, p. 108637, 2022.

[25] M. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in *Advances in Neural Information Processing Systems*, vol. 11, pp. 730–736, Dec. 1998.

[26] M. Mørup, M. N. Schmidt, and L. K. Hansen, "Shift invariant sparse coding of image and music data," *DTU Informatics, Tech. Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep. IMM2008-04659*, 2008.

[27] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Scalable online convolutional sparse coding," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4850–4859, 2018.

[28] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5687–5701, 2017.

[29] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis.*, (Venice, Italy), pp. 4290–4298, Oct. 2017.

[30] G. Peng, "Adaptive ADMM for dictionary learning in convolutional sparse representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3408–3422, 2019.

[31] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[32] R. Chalasani, J. C. Principe, and N. Ramakrishnan, "A fast proximal method for convolutional sparse coding," in *Proc. Int. Jt. Conf. Neural Netw.*, (Dallas, TX, USA), pp. 1–5, Aug. 2013.

[33] I. Rey-Otero, J. Sulam, and M. Elad, "Variations on the convolutional sparse coding model," *IEEE Trans. Signal Process.*, vol. 68, pp. 519–528, 2020.

[34] J. Liu, C. Garcia-Cardona, B. Wohlberg, and W. Yin, "First-and second-order methods for online convolutional dictionary learning," *SIAM J. Imaging Sci.*, vol. 11, no. 2, pp. 1589–1628, 2018.

[35] H. Sreter and R. Giryes, "Learned convolutional sparse coding," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process.*, (Calgary, AB, Canada), pp. 2191–2195, Apr. 2018.

[36] L. Yang, C. Li, J. Han, C. Chen, Q. Ye, B. Zhang, X. Cao, and W. Liu, "Image reconstruction via manifold constrained convolutional sparse coding for image sets," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 7, pp. 1072–1081, 2017.

[37] J. Kang, D. Hong, J. Liu, G. Baier, N. Yokoya, and B. Demir, "Learning convolutional sparse coding on complex domain for interferometric phase restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 826–840, 2021.

[38] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariance sparse coding for audio classification," *arXiv:1206.5241*, 2012.

[39] W. Luo, J. Li, J. Yang, W. Xu, and J. Zhang, "Convolutional sparse autoencoders for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3289–3294, 2018.

[40] H. Chang, J. Han, C. Zhong, A. M. Snijders, and J. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1182–1194, 2018.

[41] C. Cheng, H. Li, J. Peng, W. Cui, and L. Zhang, "Deep high-order tensor convolutional sparse coding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[42] H. Zhang and V. M. Patel, "Convolutional sparse and low-rank coding-based image decomposition," *IEEE Trans. Imag. Process.*, vol. 27, no. 5, pp. 2121–2133, 2018.

[43] J. Ding, "Fault detection of a wheelset bearing in a high-speed train using the shock-response convolutional sparse-coding technique," *Measurement*, vol. 117, pp. 108–124, 2018.

[44] D. Carrera, G. Boracchi, A. Foi, and B. Wohlberg, "Detecting anomalous structures by convolutional sparse models," in *Proc. Int. Jt. Conf. Neural Netw.*, (Killarney, Ireland), pp. 1–8, July 2015.

[45] E. Zisselman, J. Sulam, and M. Elad, "A local block coordinate descent algorithm for the CSC model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Long Beach, CA, USA), pp. 8200–8209, June 2019.

[46] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," in *Proc. European Signal Processing Conference*, (Lausanne, Switzerland), pp. 1–5, Aug. 2008.

[47] M. Pachitariu, A. M. Packer, N. Pettit, H. Dalgleish, M. Hausser, and M. Sahani, "Extracting regions of interest from biological images with convolutional sparse block coding," in *Advances in Neural Information Processing Systems*, vol. 26, (Lake Tahoe, Nevada, USA), Dec. 2013.

[48] C. Rusu, B. Dumitrescu, and S. A. Tsaftaris, "Explicit shift-invariant dictionary learning," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 6–9, 2014.

[49] B. Kong and C. C. Fowlkes, "Fast convolutional sparse coding (FCSC)," *Department of Computer Science, University of California, Irvine, Tech. Rep*, vol. 3, 2014.

[50] M. Šorel and F. Šroubek, "Fast convolutional sparse coding using matrix inversion lemma," *Digit. Signal Process.*, vol. 55, pp. 44–51, 2016.

[51] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389–393, 2021.

[52] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2009.

[53] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2011.

[54] B. Wohlberg, "SParse Optimization Research COde (SPORCO)." Software library available from `http://purl.org/brendt/software/sporco`, 2017.

[55] F. Gao, X. Deng, M. Xu, J. Xu, and P. L. Dragotti, "Multi-modal convolutional dictionary learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1325–1339, 2022.

[56] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Process.*, vol. 86, p. 572–588, 2006.

[57] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.

[58] F. Boßmann, S. Krause-Solberg, J. Maly, and N. Sissouno, "Structural sparsity in multiple measurements," *IEEE Trans. Signal Process.*, vol. 70, pp. 280–291, 2021.

[59] B. Zheng, C. Zeng, S. Li, and G. Liao, "The MMV tail null space property and DOA estimations by tail-$\ell_{2,1}$ minimization," *Signal Process.*, vol. 194, p. 108450, 2022.

[60] S. H. Fouladi, S. Chiu, B. D. Rao, and I. Balasingham, "Recovery of independent sparse sources from linear mixtures using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6332–6346, 2018.

[61] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, 2015.

[62] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10–19, 2012.

[63] W. Chen, D. Wipf, Y. Wang, Y. Liu, and I. J. Wassell, "Simultaneous Bayesian sparse approximation with structured sparse models," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6145–6159, 2016.

[64] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[65] B. Wohlberg, "Convolutional sparse representation of color images," in *2016 IEEE Southwest Symposium on Image Analysis and Interpretation*, (Santa Fe, NM, USA), pp. 57–60, Mar. 2016.

[66] Q. Zhang, Y. Fu, H. Li, and J. Zou, "Dictionary learning method for joint sparse representation-based image fusion," *Opt. Eng.*, vol. 52, no. 5, p. 057006, 2013.

[67] P. Song, X. Deng, J. F. C. Mota, N. Deligiannis, P. L. Dragotti, and M. R. D. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 57–72, 2019.

[68] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.

[69] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral super resolution of hyperspectral images via coupled dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2777–2797, 2018.

[70] P. Song, L. Weizman, J. F. C. Mota, Y. C. Eldar, and M. R. D. Rodrigues, "Coupled dictionary learning for multi-contrast MRI reconstruction," *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 621–633, 2020.

[71] F. Deeba, S. Kun, F. Ali Dharejo, and Y. Zhou, "Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm," *IET Image Process.*, vol. 14, no. 11, pp. 2365–2375, 2020.

[72] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, 2016.

[73] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1284–1294, 2014.

[74] F. G. Veshki, N. Ouzir, and S. A. Vorobyov, "Image fusion using joint sparse representations and coupled dictionary learning," in *IEEE Int. Conf. Acous., Speech, Signal Process.*, (Barcelona, Spain), pp. 8344–8348, May 2020.

[75] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Providence, RI, USA), pp. 2216–2223, June 2012.

[76] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 6, pp. 1153–1160, 1981.

[77] Harvard Medical School, "The Whole Brain Atlas ." `http://www.med.harvard.edu/AANLIB/`. [Online; accessed Dec-2022].

[78] EPFL, "RGB-NIR scene dataset ." `https://www.epfl.ch/labs/ivrl/research/downloads/rgb-nir-scene-dataset/`. [Online; accessed Dec-2022].

[79] X. Zhang, P. Ye, H. Leung, K. Gong, and G. Xiao, "Object fusion tracking based on visible and infrared images: A comprehensive review," *Inf. Fusion*, vol. 63, pp. 166–187, 2020.

[80] L. Ren, Z. Pan, J. Cao, H. Zhang, and H. Wang, "Infrared and visible image fusion based on edge-preserving guided filter and infrared feature decomposition," *Signal Process.*, vol. 186, p. 108108, 2021.

[81] F. I. Arnous, R. M. Narayanan, and B. C. Li, "Application of multidomain data fusion, machine learning and feature learning paradigms towards enhanced image-based SAR class vehicle recognition," in *Radar Sensor Technology XXV*, vol. 11742, pp. 35–46, Mar. 2021.

[82] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, 2021.

[83] Y. Peng, W. Li, X. Luo, J. Du, Y. Gan, and X. Gao, "Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal–spectral fusion of remote sensing images," *Inf. Fusion*, vol. 65, pp. 21–36, 2021.

[84] L. J. Deng, M. Feng, and X. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, 2019.

[85] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, 2017.

[86] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A review of multimodal medical image fusion techniques," *Comput. Math. Methods. Med.*, vol. 2020, 2020.

[87] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal medical image fusion review: Theoretical background and recent advances," *Signal Process.*, vol. 183, 2021.

[88] Z. Zhao, S. Xu, C. Zhang, J. Liu, and J. Zhang, "Bayesian fusion for infrared and visible images," *Signal Process.*, vol. 177, pp. 1–12, 2020.

[89] M. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, "Visible-NIR image fusion based on top-hat transform," *IEEE Trans. Image Process.*, vol. 30, pp. 4962–4972, 2021.

[90] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[91] G. Li, Y. Lin, and X. Qu, "An infrared and visible image fusion method based on multi-scale transformation and norm optimization," *Inf. Fusion*, vol. 71, pp. 109–129, 2021.

[92] W. Tan, P. T. H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Comput. Appl.*, pp. 1–21, 2020.

[93] X. Li, X. Guo, P. Han, X. Wang, H. Li, and T. Luo, "Laplacian re-decomposition for multimodal medical image fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6880–6890, 2020.

[94] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, S. Lee, and K. He, "Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain," *Infrared Phys. Technol.*, vol. 88, pp. 1–12, 2018.

[95] Y. Yang, S. Cao, S. Huang, and W. Wan, "Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–16, 2021.

[96] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, 2019.

[97] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, 2018.

[98] C. Xing, M. Wang, C. Dong, C. Duan, and Z. Wang, "Using taylor expansion and convolutional sparse representation for image fusion," *Neurocomputing*, vol. 402, pp. 437–455, 2020.

[99] F. Liu, L. Chen, L. Lu, A. Ahmad, G. Jeon, and X. Yang, "Medical image fusion method by using laplacian pyramid and convolutional sparse representation," *Concurrency Computat. Pract. Exper.*, vol. 32, no. 17, p. e5632, 2020.

[100] C. Xing, Y. Cong, Z. Wang, and M. Wang, "Fusion of hyperspectral and multispectral images by convolutional sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[101] H. Li, X. Wu, and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," *Infrared Phys. Technol.*, vol. 102, 2019.

[102] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform," *Expert Syst. Appl.*, vol. 171, p. 114574, 2021.

[103] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th Int. Conf. Inf. Fusion*, (Xi'an, China), pp. 1–7, July 2017.

[104] C. Vincent-Cuaz, T. Vayer, R. Flamary, M. Corneli, and N. Courty, "Online graph dictionary learning," in *Proc. 38th Int. Conf. Mach. Learn*, vol. 139, pp. 10564–10574, July 2021.

[105] D. Thanou, D. I. Shuman, and P. Frossard, "Parametric dictionary learning for graph signals," in *2013 IEEE Global Conference on Signal and Information Processing*, (Austin, TX, USA), pp. 487–490, IEEE, Dec. 2013.

Errata

Publication I

In the first paragraph of Section III, it is stated “this helps to bypass the non-convexity of dictionary update problem”, which is not accurate; the dictionary update problem is convex. Simultaneously updating the dictionary atoms and the nonzero entries in the sparse representations is non-convex.

Publication I

F. G. Veshki and S. A. Vorobyov. An Efficient Coupled Dictionary Learning Method. *IEEE Signal Processing Letters*, vol. 26(10), pp. 1441-1445, 2019.

An Efficient Coupled Dictionary Learning Method

Farshad G. Veshki and Sergiy A. Vorobyov, *Fellow, IEEE*

***Abstract*—In this letter, we present a generic and computationally efficient method for coupled dictionary learning (CDL). The proposed method enforces relations between the corresponding atoms of dictionaries learned to represent two related (but not necessarily of the same dimensionality) feature spaces, aiming that each pair of related signals from the two feature spaces has the same sparse representation with respect to their corresponding learned dictionaries. Coupled learned dictionaries have various applications in many sparse representation-based recognition and reconstruction problems, where the two related feature spaces are representing the same signal of different modalities or different qualities. The presented experimental comparisons show that the results obtained using our proposed CDL method are competitive to those of the state-of-the-art CDL methods in performance, while the proposed method has a significantly lower computational cost. Furthermore, the proposed method can be straightforwardly used for learning coupled dictionaries from more than two related feature spaces.**

***Index Terms*—Coupled dictionary learning, feature space learning, sparse representation.**

I. Introduction

SPARSITY and overcompleteness has been successfully used for diverse applications in signal processing over the last decade [1]–[4]. The fact exploited is that the signal $\boldsymbol{x}$ can be described as a linear combination of few atoms over an overcomplete dictionary $\boldsymbol{D}$, and the problem of seeking such sparse representation can be formulated as $\min_{\boldsymbol{\alpha}}\|\boldsymbol{\alpha}\|_0$ s.t. $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is the sparse vector of coefficients for atoms in the dictionary $\boldsymbol{D}$ and $\|\cdot\|_0$ denotes the operator that counts the number of non-zero entries in a vector.

Many applications have benefited remarkably from using the above approach with learned overcomplete dictionary [5]–[8]. Representative examples of dictionary learning algorithms include the K-SVD method [9], the method of optimal directions (MOD) [10], and the online dictionary learning (OLD) method [11]. "Good" dictionaries are expected to be highly adaptive to the observed signals and to lead to accurate sparse representations.

While the *single dictionary* model has been extensively studied, there exists also a *coupled dictionary* viewpoint to sparsity and overcompleteness, where a coupled dictionary is needed to represent the double feature space, e.g., low-resolution (LR) and high-resolution (HR) images in image processing [2]. The combination of learned coupled dictionaries and sparse approximation is shown to be superior for representing double feature spaces [12]–[19]. Signal reconstruction problems [12]–[14], recognition tasks [16], [17], and multi-focus image fusion [18] are examples of applications of coupled dictionaries.

A majority of existing CDL algorithms aim to learn two related feature spaces through burdensome complex procedures, while the computationally demanding nature of dictionary learning algorithms becomes more restrictive when we need to learn two dictionaries simultaneously. In this letter, we propose a fast and easy to implement CDL scheme based on joint sparse coding and computationally cheap atom update rules, which dramatically reduces the computational cost compared to the existing CDL methods without sacrificing the performance even slightly.

II. Problem Statement

The CDL aims to find a pair of dictionaries $\{\boldsymbol{D}_1 \in \mathbb{R}^{m_1\times n}, \boldsymbol{D}_2 \in \mathbb{R}^{m_2\times n}\}$ best representing two subsets of p training signals $\boldsymbol{X}_1 = [[\boldsymbol{x}_1]_1, \ldots, [\boldsymbol{x}_1]_p]$ and $\boldsymbol{X}_2 = [[\boldsymbol{x}_2]_1, \ldots, [\boldsymbol{x}_2]_p]$ in such a way that if a linear combination of atoms of $\boldsymbol{D}_1$ models a signal in $\boldsymbol{X}_1$, the same linear combination of atoms of $\boldsymbol{D}_2$ also models the corresponding signal in $\boldsymbol{X}_2$. Notice that the dimensionalities of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are not necessarily the same. Then the CDL problem can be formulated as the following optimization problem [12]

$$\begin{aligned}&\min_{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{\Gamma}} \omega\,\|\boldsymbol{X}_1 - \boldsymbol{D}_1\boldsymbol{\Gamma}\|_{\mathrm{F}}^2 + (1-\omega)\,\|\boldsymbol{X}_2 - \boldsymbol{D}_2\boldsymbol{\Gamma}\|_{\mathrm{F}}^2\\ &\text{s.t. } \|\boldsymbol{\gamma}_i^{\mathrm{c}}\|_0 \leqslant T_0,\ \|[\boldsymbol{d}_1]_t\|_2 = 1, \|[\boldsymbol{d}_2]_t\|_2 = 1, \forall t, i \end{aligned} \tag{1}$$

where $[\boldsymbol{d}_1]_t$ and $[\boldsymbol{d}_2]_t$ are the t-th dictionary atoms (columns) of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, respectively, T_0 is the constraint value on sparsity, $\boldsymbol{\gamma}_i^{\mathrm{c}}$ denotes the i-th column of $\boldsymbol{\Gamma}$ ("c" is for "column"), ω ($0 \leq \omega \leq 1$) controls the two approximation errors associated to $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, $\|\cdot\|_2$ is the Euclidean norm of a vector, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix.

A commonly used approximation for (1) is based on reformulating (1) as a joint dictionary learning problem (2)

$$\min_{\boldsymbol{D},\boldsymbol{\Gamma}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{\Gamma}\|_{\mathrm{F}}^2 \text{ s.t. } \|\boldsymbol{\gamma}_i^{\mathrm{c}}\|_0 \leqslant T_0,\ \forall i \tag{2}$$

where $\boldsymbol{X} \triangleq \begin{bmatrix}\sqrt{\omega}\boldsymbol{X}_1\\ \sqrt{1-\omega}\boldsymbol{X}_2\end{bmatrix}$ and $\boldsymbol{D} \triangleq \begin{bmatrix}\sqrt{\omega}\boldsymbol{D}_1\\ \sqrt{1-\omega}\boldsymbol{D}_2\end{bmatrix}$. Problem (2) can be addressed using any single dictionary learning method. However, problems (1) and (2) are not equivalent with respect to $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. Thus, the jointly learned dictionaries are not guaranteed to be individually adaptive to $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively.

Manuscript received July 1, 2019; revised August 3, 2019; accepted August 3, 2019. Date of publication August 8, 2019; date of current version August 23, 2019. This work was supported in part by the Academy of Finland under Grants 299243 and 319822. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marco F. Duarte. *(Corresponding author: Sergiy A. Vorobyov.)*

The authors are with the Department of Signal Processing and Acoustics, Aalto University, FI-00076 Aalto, Finland (e-mail: farshad.ghorbaniveshki@aalto.fi; svor@ieee.org).

This letter has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/LSP.2019.2934045

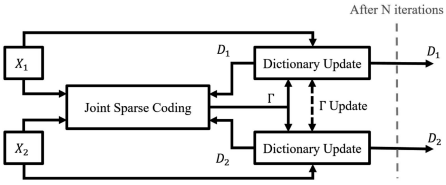


Fig. 1. Block-diagram of the proposed CDL method.

A bilevel optimization scheme which directly addresses (1) has been also proposed in [12]. It alternatively optimizes $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, where $\boldsymbol{\Gamma}$ is the sparse represenation of $\boldsymbol{X}_1$ over $\boldsymbol{D}_1$. It is the state-of-the-art CDL method for our developments. Moreover, instead of enforcing an identical sparse representation, the method in [15] learns $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ which yeild $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2 = \boldsymbol{W}\boldsymbol{\Gamma}_1$ representing $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively. Here $\boldsymbol{W}$ is a linear mapping from $\boldsymbol{\Gamma}_1$ to $\boldsymbol{\Gamma}_2$. This method also solves the CDL problem (1) in the sense that the product of the dictionaries $\boldsymbol{D}_1$ and $\hat{\boldsymbol{D}}_2 = \boldsymbol{D}_2\boldsymbol{W}$ and the common sparse representation $\boldsymbol{\Gamma}_1$ will reconstruct $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively. The main problem with the existing CDL methods is their high computational complexities.[1]

III. Proposed Method

In this letter, we propose to exploit the fact that CDL problem (1) is equivalent to the joint dictionary learning problem (2) with respect to $\boldsymbol{\Gamma}$. Performing the sparse coding jointly, apart from simplying the algorithm, improves the effectivness of the CDL update cycles.[2] We also show that the atoms of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be learned disjointly from each other, only with respect to their corresponding rows in $\boldsymbol{\Gamma}$. This helps to bypass the non-convexity of dictionary update problem, that arises from the sparsity constraint on $\boldsymbol{\Gamma}$, and to learn the optimal atoms using computationally cheap update rules only with respect to the nonzero entries of $\boldsymbol{\Gamma}$.

Thus, we split the optimization variables in problem (1) into two subsets: $\{\boldsymbol{\Gamma}\}$ and $\{\boldsymbol{D}_1, \boldsymbol{D}_2\}$. Then (1) can be addressed in alternating manner by iterating between two phases, where in the first phase $\boldsymbol{\Gamma}$ is optimized under the constraint $\|\boldsymbol{\gamma}_i^{\mathrm{c}}\|_0 \leqslant T_0$ – a *joint sparse coding* problem, and in the second phase $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are optimized under the constraints $\|[\boldsymbol{d}_1]_t\|_2 = 1$ and $\|[\boldsymbol{d}_2]_t\|_2 = 1$, respectively, – a *coupled dictionary update* problem. The general procedure for our CDL method is then summarized in the block-diagram presented in Fig. 1. The dashed arrow in the block diagram indicates that in order to preserve the same sparse representation for both $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, the updates of $\boldsymbol{\Gamma}$ need to be performed in a coupled manner also during the dictionary update phase. The dictionaries can be initialized by any fixed basis overcomplete dictionary, e.g., discrete cosine transform (DCT) dictionary or a Gaussian random matrix of appropriate size with (l_2-norm) normalized columns. Moreover, we show that our method can be easily extended to learning coupled dictionaries from multiple feature spaces.

[1] See Section IV for detailed studies of complexity.

[2] We explain this in details later in SubSection IV-B.

A. Joint Sparse Coding

Problems (1) and (2) are equivalent with respect to only $\boldsymbol{\Gamma}$. Thus, the optimal $\boldsymbol{\Gamma}$ for CDL scenario is equal to the sparse approximation of joint training data with respect to the joint dictionary, i.e., $\boldsymbol{X}$ and $\boldsymbol{D}$ (defined after (2)), respectively. Such a sparse coding problem can be solved using many available sparse coding algorithms, for example, orthogonal matching pursuit (OMP) [20], focal underdetermined system solver (FOCUSS) [21], and least angle regression (LARS) [22], [23].

B. Coupled Dictionary Update

For the common sparse representation $\boldsymbol{\Gamma}$, problem (1) needs to be solved then over the coupled dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. The corresponding optimization problem is given as

$$\min_{\boldsymbol{D}_1,\boldsymbol{D}_2} \omega \left\| \boldsymbol{X}_1 - \sum_t [\boldsymbol{d}_1]_t \boldsymbol{\gamma}_t^{\mathrm{r}} \right\|_{\mathrm{F}}^2 + (1-\omega) \left\| \boldsymbol{X}_2 - \sum_t [\boldsymbol{d}_2]_t \boldsymbol{\gamma}_t^{\mathrm{r}} \right\|_{\mathrm{F}}^2 \tag{3}$$

subject to the constraints in (1). Here, $\boldsymbol{\gamma}_i^{\mathrm{r}}$ is the i-th row of $\boldsymbol{\Gamma}$ ("r" is for "row"), and we rewrite the products $\boldsymbol{D}_1\boldsymbol{\Gamma}$ and $\boldsymbol{D}_2\boldsymbol{\Gamma}$ as the sums of vector outer products $[\boldsymbol{d}_1]_t\boldsymbol{\gamma}_t^{\mathrm{r}}$ and $[\boldsymbol{d}_2]_t\boldsymbol{\gamma}_t^{\mathrm{r}}$, respectively. Then each pair of corresponding atoms can be updated disjointly from the others and independent from ω. Thus, to update the t-th pair of atoms, we fix the remaining atoms, and rewrite the optimization problem (3) as

$$[\boldsymbol{d}_i]_t = \underset{[\boldsymbol{d}_i]_t}{\operatorname{argmin}} \left\| [\boldsymbol{E}_i]_t - [\boldsymbol{d}_i]_t \, [\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}} \right\|_{\mathrm{F}}^2, \; i = 1, 2 \tag{4}$$

where $[\boldsymbol{E}_i]_t$ represents the approximation residuals excluding the contribution of t-th atom, and it is defined as

$$[\boldsymbol{E}_i]_t \triangleq \left[\boldsymbol{X}_i - \sum_{s \neq t} [\boldsymbol{d}_i]_s \boldsymbol{\gamma}_t^{\mathrm{r}} \right]_{\boldsymbol{f_t}}, \boldsymbol{f_t} = \{i | [\boldsymbol{\gamma}_t^{\mathrm{r}}]_i \neq 0\}.$$

Here $\boldsymbol{f_t}$ is the set of indices of nonzero entries of $\boldsymbol{\gamma}_t^{\mathrm{r}}$. In a single dictionary update problem, K-SVD [9] addresses (4) by finding the rank-1 approximation of $[\boldsymbol{E}_i]_t$ using singular value decomposition (SVD) and simultaneously updates $[\boldsymbol{d}_i]_t$ and $[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}$. However, this approach is not applicable to the coupled dictionary update case, since $[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}$ needs to be preserved identical for both dictionaries. Instead, we address (4) by solving two seperate least squares (LS) problems. The solutions can be easily found as $[\boldsymbol{d}_i]_t = [\boldsymbol{E}_i]_t [\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}^T / \|[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}\|_2^2$, $i = 1, 2$. Since we need to normalize the l_2-norm of each atom to one anyway, the normalization term $\|[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}\|_2^2$ can be dropped. Then the atom update rule is

$$[\boldsymbol{d}_i]_t = [\boldsymbol{E}_i]_t [\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}^T, \; i = 1, 2. \tag{5}$$

After updating $[\boldsymbol{d}_i]_t, i = 1, 2$, we need to update $[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}$. Since $[\boldsymbol{d}_i]_t$ is a unit vector, the solution of (3), this time over $[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}}$, can be efficiently found as

$$[\boldsymbol{\gamma}_t^{\mathrm{r}}]_{\boldsymbol{f_t}} = \boldsymbol{d}_t^T \boldsymbol{E}_t \tag{6}$$

where $\boldsymbol{d}_t \triangleq \begin{bmatrix} \sqrt{\omega}[\boldsymbol{d}_1]_t \\ \sqrt{1-\omega}[\boldsymbol{d}_2]_t \end{bmatrix}$ and $\boldsymbol{E}_t \triangleq \begin{bmatrix} \sqrt{\omega}[\boldsymbol{E}_1]_t \\ \sqrt{1-\omega}[\boldsymbol{E}_2]_t \end{bmatrix}$.

When $\boldsymbol{f_t}$ is empty, i.e., when $[\boldsymbol{d}_1]_t$ and $[\boldsymbol{d}_2]_t$ are not used in the approximation of any training sample, they can be substituted

Algorithm 1: Coupled Dictionary Learning.

Input: $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and $\boldsymbol{D}_0 =$ DCT dictionary.
Initialization: Set $\boldsymbol{D}_1 := \boldsymbol{D}_0$, $\boldsymbol{D}_2 := \boldsymbol{D}_0$.
Number of update cycles $:= N$.
for N cycles **do**
  Joint sparse coding:
    Find $\boldsymbol{\Gamma}$ over $\boldsymbol{X}$, $\boldsymbol{D}$ (2);
  for $t = 1 \cdots$ number of atoms **do**
    Find $\boldsymbol{f}_t = \{i | [\boldsymbol{\gamma}_t^{\mathrm{r}}]_i \neq 0\}$;
    if $\boldsymbol{f}_t \neq \emptyset$
      Update $[\boldsymbol{d}_1]_t$ and $[\boldsymbol{d}_2]_t$ using (5);
      Normalize the atoms:
      $[\boldsymbol{d}_1]_t = \frac{[\boldsymbol{d}_1]_t}{\|[\boldsymbol{d}_1]_t\|_2}$ and $[\boldsymbol{d}_2]_t = \frac{[\boldsymbol{d}_2]_t}{\|[\boldsymbol{d}_2]_t\|_2}$;
      Update $\boldsymbol{\gamma}_t^{\mathrm{r}}$ using (6);
    else
      Update $[\boldsymbol{d}_1]_t$ and $[\boldsymbol{d}_2]_t$ using (7);
    end if
  end for
end for
Output: The pairwise correlated dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$.

with the pair of inputs $[\boldsymbol{x}_1]_j$ and $[\boldsymbol{x}_2]_j$ that jointly have the largest approximation residuals. This can be formulated as

$$[\boldsymbol{d}_i]_t = \frac{[\boldsymbol{x}_i]_j}{\|[\boldsymbol{x}_i]_j\|_2},\ j = \underset{j}{\operatorname{argmax}} \|\boldsymbol{x}_j - \boldsymbol{D}\boldsymbol{\gamma}_j^{\mathrm{c}}\|_2^2,\ i = 1, 2. \quad (7)$$

C. Summary of the Algorithm

The overall algorithm for CDL can be then summarized as in Algorithm 1.

D. Complexity Analysis

Let s_i^{c}, s_i^{r} be the number of nonzero entries in i-th column and row of $\boldsymbol{\Gamma}$, respectively, and S be the total number of nonzeros elements. For sparse coding, we recommend LARS/Homotopy l_1-minimization algorithm [23]. The computational complexity of k-step Homotopy algorithms for a general $m \times n$ dictionary is bounded by $sm^2 + smn$ flops [27], where s is the number of nonzero coefficients. Then the complexity of sparse coding for our method is bounded by $\sum_{i=1}^{p} s_i^{\mathrm{c}}(m_1 + m_2)^2 + s_i^{\mathrm{c}}(m_1 + m_2)n = S((m_1 + m_2)^2 + (m_1 + m_2)n)$ flops per each learning cycle. Also, the complexity of atom update phase (equations (5) and (6)) is $\sum_{i=1}^{n} 2s_i^{\mathrm{r}}(m_1 + m_2) = 2S(m_1 + m_2)$ flops per learning cycle.

E. CDL for More Than Two Feature Spaces

The proposed method can be directly applied to the case where M coupled dictionaries need to be learned from M related feature spaces. This is the case, e.g., in distributed compressed sensing where three dictionaries need to be learned (for common and two innovation components) [28], [29]. For such case, the objective function in (1) can be rewritten as

$$\min_{\boldsymbol{D}_1,\ldots,\boldsymbol{D}_M,\boldsymbol{\Gamma}} \sum_{j=1}^{M} \lambda_j \|\boldsymbol{X}_j - \boldsymbol{D}_j\boldsymbol{\Gamma}\|_{\mathrm{F}}^2$$

where λ_j ($\sum_{j=1}^{M} \lambda_j = 1, \lambda_j \geq 0$) controls the tradeoff among approximation errors. Then, for the joint sparse coding phase, $\boldsymbol{X}$ and $\boldsymbol{D}$ are formed as

$$\boldsymbol{X} = \left[\sqrt{\lambda_1}\boldsymbol{X}_1^T, \ldots, \sqrt{\lambda_M}\boldsymbol{X}_M^T\right]^T,$$

$$\boldsymbol{D} = \left[\sqrt{\lambda_1}\boldsymbol{D}_1^T, \ldots, \sqrt{\lambda_M}\boldsymbol{D}_M^T\right]^T.$$

The atom update rule (5) is the same for multiple dictionary learning case. The only difference is that $i = 1, \ldots, M$ this time. Then, the rows of $\boldsymbol{\Gamma}$ can be updated using (6) with

$$\boldsymbol{d}_t = \left[\sqrt{\lambda_1}[\boldsymbol{d}_1]_t^T, \ldots, \sqrt{\lambda_M}[\boldsymbol{d}_M]_t^T\right]^T,$$

$$\boldsymbol{E}_t = \left[\sqrt{\lambda_1}[\boldsymbol{E}_1]_t^T, \ldots, \sqrt{\lambda_M}[\boldsymbol{E}_M]_t^T\right]^T.$$

IV. Experimental Results

In this section, we compare our CDL method with the method of [12] in terms of the performance in an image super-resolution (SR) problem, convergence speed, and algorithm complexity. We employ the CDL based signle-image SR algorithm used in [12]. The experiments are performed on a PC running an Intel(R) Xeon(R) 3.40 GHz CPU.

A. CDL for Image Super Resolution

The image SR algorithm of [12] employs two dictionaries learned over two datasets of corresponding LR and HR images to recover the SR image patches from their LR versions. Instead of the original LR signals, features containing their median frequency band, which are known to contain the most relevant information [2], are extracted using four 1-D filters. The dimensionality of the obtained feature vectors are then four times higher than the original signals. We train three pairs of dictionaries (where $n = 512$ always) using our method, the method of [12], and K-SVD [9] (for joint dictionary learning as in (2)). The parameter ω is always equal to 0.5. Both CDL algorithms are initialized using Gaussian random dictionaries. The training data[3] includes 10,000 vectorized HR intensity patches ($\boldsymbol{X_2} \in \mathbb{R}^{25\times n}$), and their corresponding LR feature vectors ($\boldsymbol{X_1} \in \mathbb{R}^{100\times n}$). Using each pair of learned dictionaries, while the rest of parameters are kept unchanged, we apply the SR algorithm of [12] to a four times downsized version of Lena image, then compare the results in Fig. 2. The results are also compared with the upsized image using bicubic interpolation [24].

The results in Fig. 2 show that the SR images obtained using the coupled learned dictionaries are significantly better than that of bicubic interpolation, which is excessively blurred, and the SR image obtained using jointly learned dictionaries which has non-smooth edges and contains artefacts. There is almost no difference between the results obtained using the dictionaries learned by our method and that of [12]. We apply the SR algorithm to two more images, Child and Peppers and summarize the root mean squared error (RMSE) values[4] in Table I, which show that the coupled dictionaries learned by [12] and the proposed

[3]The training data is taken from the demo software of [2] made available online by the authors of [25].

[4]RMSE values are calculated as $\sqrt{\frac{1}{p}\sum_{i=1}^{p}(\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\gamma}_i^{\mathrm{c}})^2}$.

Fig. 2. (a) Input (128×128), (b) original image (512×512), (c) upsized image using bicubic interpolation [24]; the reconstructed images using (d) jointly learned dictionaries, (e) Yang *et al.* CDL method [12], and (f) the proposed CDL method.

TABLE I
RMSE RESULTS FOR SR RECONSTRUCTION OF THREE TEST IMAGES USING BICUBIC INTERPOLATION, AND SR METHOD OF [12] FOR THREE PAIRS OF LEARNED DICTIONARIES, LEARNED BY: K-SVD, CDL METHOD OF [12], AND THE PROPOSED METHOD

	K-SVD [9](Joint)	Bicubic interpolation [24]	Yang et. al. [12]	Proposed
Lena	7.6314	8.546	7.3968	7.3485
Child	5.8584	6.5328	5.6994	5.7277
Peppers	8.1263	8.7691	7.9389	7.9124

method have similar performance, and both of them perform better than jointly learned dictionaries.

B. Convergence Speed

To compare the convergence speed of the two CDL algorithms, we visualize the density of $\boldsymbol{\Gamma}$ (see Fig. 3(a)) and RMSE values associated with $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ (see Fig. 3(b,c)), during 15 CDL cycles. In the first few cycles, the density of $\boldsymbol{\Gamma}$ is higher for our method, because we perform the sparse coding jointly, while in [12] it is performed only for $\boldsymbol{X_1}$. The latter also results in lower RMSE for [12] in approximation of $\boldsymbol{X_1}$, however, it leads to higher RMSE for that of $\boldsymbol{X_2}$. In each dictionary update phase, the lower bounds of error for the approximations $\boldsymbol{D_1\Gamma} \simeq \boldsymbol{X_1}$ and $\boldsymbol{D_2\Gamma} \simeq \boldsymbol{X_2}$ are directly dependent on the overlaps between the row-space of $\boldsymbol{\Gamma}$ and those of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, respectively. The

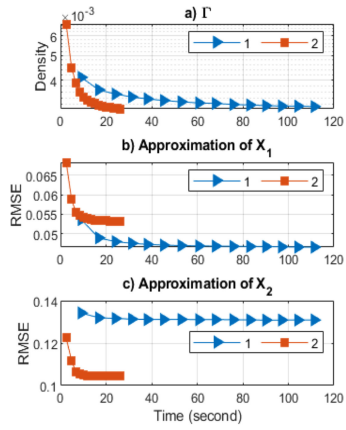


Fig. 3. Comparing the proposed CDL method – 2 and the method of [12] – 1, in terms of: a) density of $\boldsymbol{\Gamma}$ (the ratio between the number of nonzeros and total number of entries), b) RMSE of approximaion of $\boldsymbol{X}_1$, and c) RMSE of approximaion of $\boldsymbol{X}_2$, over 15 CDL cycles.

row-space of a jointly optimal $\boldsymbol{\Gamma}$ approximately spans that of $\boldsymbol{X}$ ($\boldsymbol{D\Gamma} \simeq \boldsymbol{X}$) which means that it also approximately spans the row-spaces of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ (since $\boldsymbol{X}$ has the rows of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$). However, the latter is not guaranteed if $\boldsymbol{\Gamma}$ is obtained only for $\boldsymbol{X}_1$ as in [12].

The density of $\boldsymbol{\Gamma}$ in both experiments converges to around 0.003 (in average 1.53 nonzero entries per column). In Fig. 3, the superiority of the proposed method in terms of convergence speed is clearly visible. The convergence takes nearly 25 and 90 seconds for our method and that of [12], respectively.

C. Complexity Comparison

The computational complexity of the CDL algorithm is not given in [12]. Thus, we calculate it and compare to that of our proposed method. We use LARS/Homotopy algorithm [23] also for [12]. The complexity of sparse coding and updating $\boldsymbol{D}_1$ for [12] is at least $SN(2m_1^2 + 2m_1 n + 5m_1 + 3m_2)$ flops. Learning $\boldsymbol{D}_2$ using Lagrange dual method proposed in [26] (as recommended in [12]) also needs $p(N + 2M)(n^2 + nm_2)$ flops where M is the number of iterations of the conjugate gradient descent method that [26] uses for finding the optimal dual variables. Considering that $m_2, m_1 \ll n \ll p < S$, it is easy to see that the total computational cost of our CDL method is significantly lower than that of the method of [12].

V. CONCLUSION

A novel computationally efficient CDL algorithm that enforces common sparse representations for double feature spaces and can be straightforwardly extended to learn coupled dictionaries from more than two feature spaces has been proposed. The performance and convergence speed of the proposed method have been compared to the state-of-the-art CDL method. The comparison results show that the proposed method reduces dramatically the computational costs, which is crucially important for computationally costly tasks such as CDL.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Mia, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[2] J. Yang, J. Wright, T. S. Huang, and Y. Mia, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, May. 2010.

[3] L. Zhang, W. D. Zhou, P. C. Cheng, J. Liu, Z. Yan, and T. Wang, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.

[4] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Honolulu, HI, USA, 2007, pp. 126–133.

[5] I. ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 3501–3508.

[6] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, 2011, pp. 457–464.

[7] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Nov. 2012.

[8] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse Model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661–677, Feb. 2013.

[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[10] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Phoenix, AZ, USA, 1999, pp. 2443–2446.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. ACM Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 689–696.

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

[13] J. Ahmed, R. A. Memon, M. Waqas, M. I. Mangrio, and S. Ali, "Selective sparse coding based coupled dictionary learning algorithm for single image super-resolution," in *Proc. Int. Conf. Comput., Math. Eng. Technol.*, Sukkur, Pakistan, 2018, pp. 1–5.

[14] J. Sadasivan, S. Mukherjee, and C. S. Seelamantula, "Joint dictionary training for bandwidth extension of speech signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, China, 2016, pp. 5925–5929.

[15] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Rhode Island, USA, 2012, pp. 2216–2223.

[16] D. Mandal and S. Biswas, "Generalized coupled dictionary learning approach with applications to cross-Modal matching," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3826–3837, Jun. 2016.

[17] F. Huang and Y. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proc. IEEE Int. Conf. Comp. Vis.*, Sydney, NSW, Australia, 2013, pp. 2496–2503.

[18] R. Gao, S. A. Vorobyov, and H. Zhao, "Multi-focus image fusion via coupled dictionary training," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, China, 2016, pp. 1666–1670.

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

[20] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[22] B. Efron, T. Hastie, J. Trevor, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[23] I. Dori and D. L. Donoho, "Solution of l_1 minimization problems by LARS/Homotopy methods," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Toulouse, France, Jul. 2006, pp. 636–639.

[24] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[25] J. Yang, [Online]. Available: http://www.ifp.illinois.edu/∼jyang29/ScSR.html. Accessed on: Mar. 5, 2019.

[26] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inform. Process. Syst.*, Dec. 2007, vol. 19, no. 2, pp. 801–808.

[27] D. L. Donoho and Y. Tsaig, "Fast solution of l_1-Norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory.*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[28] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Conf. Rec. 39th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2005, pp. 1537–1541.

[29] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, Florence, Italy, Oct. 2012, vol. 7572, pp. 186–199.

Publication II

F. G. Veshki, N. Ouzir and S. A. Vorobyov. Image Fusion using Joint Sparse Representations and Coupled Dictionary Learning. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, pp. 8344-8348, May 2020.

Image Fusion using Joint Sparse Representations and Coupled Dictionary Learning

Farshad G. Veshki, Nora Ouzir and Sergiy A. Vorobyov, *Fellow, IEEE*

***Abstract*—The image fusion problem consists in combining complementary parts of multiple images captured, for example, with different focal settings into one image of higher quality. This requires the identification of the sharpest areas in sets of input images. Recently, it was shown that coupled dictionary learning can successfully capture the relationships between high- and low-resolution patches in the context of single image super-resolution. In this work, to identify the sharp image patches, we propose an improved discriminative coupled dictionary learning approach using joint sparse representations in blurred and focused dictionaries. In addition, a pixel-wise processing of the boundaries (*i.e.,* patches containing blurred and focused pixels) is proposed. The experimental results using two natural image datasets, as well as a sequence of *in vivo* microscopy images, show the competitiveness of the proposed method compared to state-of-the-art algorithms in terms of accuracy and computational time.**

***Index Terms*—Image fusion, coupled dictionary learning, joint sparse representations.**

I. Introduction

Image fusion is a post-processing technique that combines the relevant information from multiple images captured with different tools or parameters in a single image. Image fusion techniques seek to preserve image quality without resorting to often costly specialized optic sensors [1]–[3]. Due to the potential for a considerable cost reduction, image fusion has attracted increased attention in various fields, including remote sensing or medical imaging [4]–[7].

State-of-the-art image fusion methods can be grouped into two main categories: spatial and transform domain methods. The first approach relies on measures such as spatial frequency [8] and phase congruency [9] that allow the significance level of pixels (or image patches) to be evaluated. A fused image is then obtained by assigning the elements with the highest significance levels to their corresponding locations in the final image. In the second approach, the input images are transformed, and fusion is performed over the transform coefficients before generating the all-in-focus image by using inverse transform. Typical transform domain approaches use frequency-based transforms, such as wavelets [10].

One emerging image fusion approach utilizes sparse representations in dictionaries learned from the data itself [11]–[14]. These methods exploit the fact that patches of natural images can be compactly represented by a linear combination of only few *atoms* of an over-complete dictionary. Sparsity is then used either as a quality measure [11], or in order to learn the desired features from the labelled training data [12]. However, these methods commonly employ a single (usually focused) dictionary. The main drawback of the single dictionary approach is that it fails to take advantage of the features in the degraded and noisy images. Specifically, the inability of properly describing degraded patches can noticeably reduce the accuracy of the sparse representation. To bypass this issue, a simultaneous learning of blurred and focused dictionaries was proposed in [14]. However, this method does not exploit the core advantage of coupled dictionary learning (CDL), which is based on a common sparse representation that expresses the correlation between blurred and focused features. Specifically, separate sparse codes are used for each dictionary, which are then averaged in order to perform fusion [14].

The CDL technique has led to state-of-the-art performance in various image processing applications [15]–[18]. CDL is designed to learn a pair of dictionaries for capturing the relationships between two correlated input data. In particular, the dictionaries are *coupled*, in the sense that they use common sparsity coefficients to reconstruct the data. For example, coupled dictionaries have been successfully used for describing the connection between high- and low-resolution features in image super-resolution [17]. Similarly, one can use CDL to capture blurred and focused image features [14]. In the context of multi-focus image fusion, CDL can be formulated as the following minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{D}_F, \boldsymbol{D}_B, \boldsymbol{A}} \|\boldsymbol{X}_{\mathrm{Ft}} - \boldsymbol{D}_F \boldsymbol{A}\|_{\mathcal{F}}^2 + \|\boldsymbol{X}_{\mathrm{Bt}} - \boldsymbol{D}_B \boldsymbol{A}\|_{\mathcal{F}}^2 \\ &\text{s.t. } \|\boldsymbol{\alpha}_i\|_0 \leqslant K,\ \|[\boldsymbol{d}_{\mathrm{Ft}}]_j\|_2 = 1, \|[\boldsymbol{d}_{\mathrm{Bt}}]_j\|_2 = 1, \forall j, i \end{aligned} \tag{1}$$

where $\boldsymbol{D}_{\mathrm{F}} \in \mathbb{R}^{n \times q}$ and $\boldsymbol{D}_{\mathrm{B}} \in \mathbb{R}^{n \times q}$ represent the focused and blurred dictionaries, respectively, with $[\boldsymbol{d}_{\mathrm{F}}]_j$, $[\boldsymbol{d}_{\mathrm{B}}]_j$ referring to their j-th columns (*i.e.,* atoms). The focused and blurred training data are denoted as $\boldsymbol{X}_{\mathrm{Ft}}$ and $\boldsymbol{X}_{\mathrm{Bt}}$, respectively. The i-th column of the common sparse representation matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{\alpha}_i$. Finally, $\|\cdot\|_0$ is the operator counting the number of non-zero coefficients in a vector, K is the maximum number of such coefficients, and $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm. In image fusion, CDL can be interpreted as an approximation of the underlying blurring function in the form of a linear transformation between any two tight column-wise corresponding subspaces of $\boldsymbol{D}_{\mathrm{F}}$ and $\boldsymbol{D}_{\mathrm{B}}$. Once the coupled dictionaries are learned, the joint sparse representation of two input image patches can be used for reconstruction or for building a fusion rule, as will be explained later in this paper.

In this work, we introduce a novel CDL-based image fusion method. The proposed approach consists of two main stages. First, CDL is used to capture the relationships between blurred

The authors are with Aalto University, Dept. Signal Processing and Acoustics, FI-00076, AALTO, Finland. E-mail: farshad.ghorbaniveshki@aalto.fi, nora.ouzir@aalto.fi, svor@ieee.org

and focused features by learning coupled dictionaries from labelled training data. In order to improve the discriminability of the dictionaries, a CDL method using structured incoherency is proposed. In a second stage, the reconstruction errors (obtained after a sparse coding phase) of a pair of input patches are used to identify the focused patch. In contrast to the method of [14], we do not separate the sparse representations in each dictionary, but rather promote the correlation between blurred and focused images by enforcing joint sparsity. Furthermore, we employ the reconstruction errors instead of the sparsity-level as a discriminative rule. Finally, an all-in-focus image is obtained by averaging the selected focused patches. The effect of blocking artefacts is mitigated by applying a sliding window approach. Since the patches located between blurred and focused areas contain varying focus levels, we propose a pixel-wise strategy for handling boundary regions. Experimental results using two natural image datasets and a sequence of *in vivo* microscopy images are presented in Section III. The results show the proposed method to be more effective than several existing fusion approaches.

II. Image Fusion using CDL

A. Problem formulation

We consider the fusion problem where, for simplicity but without loss of generality, two images $\boldsymbol{I}_1 \in \mathbb{R}^{N\times M}$ and $\boldsymbol{I}_2 \in \mathbb{R}^{N\times M}$ with varying focus levels are fused into an all-in-focus image $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$. We propose to use a patch-wise approach where n_p patches of size n are extracted from $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$, then concatenated into two matrices $\boldsymbol{X}_1 \in \mathbb{R}^{n\times n_p}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{n\times n_p}$, respectively. The corresponding single input patches are denoted by $\boldsymbol{x}_1 \in \mathbb{R}^n$ or $\boldsymbol{x}_2 \in \mathbb{R}^n$. After selecting the focused patches $\boldsymbol{x}_F$ from the pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and concatenating them into a matrix $\boldsymbol{X}_F \in \mathbb{R}^{n\times n_p}$, the final all-in-focus image is obtained by weighted averaging as follows

$$\boldsymbol{Y} = \mathcal{P}^*(\boldsymbol{X}_F), \tag{2}$$

where $\mathcal{P}(\cdot) : \mathbb{R}^{N\times N} \mapsto \mathbb{R}^{n\times n_p}$ is a linear operator that extracts n_p overlapping patches of size n from an image, and $\mathcal{P}^*(\cdot)$ is its adjoint operation, which places each patch into its location in the image and performs averaging depending on the amount of overlap between patches, *i.e.,* $\mathcal{P}^*[\mathcal{P}(\boldsymbol{I})] = \boldsymbol{I}$. The following subsections first provide some details on the CDL approach used for learning the dictionaries $\boldsymbol{D}_\mathrm{F}$ and $\boldsymbol{D}_\mathrm{B}$. Then the sparse representation-based selection rule allowing the patches $\boldsymbol{x}_F$ to be selected is presented.

B. CDL with structured incoherency

Prior to fusion, two coupled dictionaries $\boldsymbol{D}_\mathrm{F}$ (focused) and $\boldsymbol{D}_\mathrm{B}$ (blurred) are learned using labelled training data. The dictionaries are obtained by solving (1), as explained in Section I. In this work, we choose to solve (1) using the algorithm proposed in [22]. More specifically, the method of [22] is based on an iterative minimization approach, that alternates between minimizations with respect to the dictionaries $\boldsymbol{D}_\mathrm{F}$ and $\boldsymbol{D}_\mathrm{B}$, and the sparse codes in $\boldsymbol{A}$. The sparse coding step is solved using the orthogonal matching pursuit (OMP) [19], while the dictionary update is obtained by solving the following minimization problem:

$$[\boldsymbol{d}_c]_j = \underset{[\boldsymbol{d}_c]_j}{\operatorname{argmin}} \left\| [\boldsymbol{E}_c]_j - [\boldsymbol{d}_c]_j \left[\boldsymbol{\alpha}_j^T\right]_{\boldsymbol{f_j}} \right\|_{\mathcal{F}}^2, \; c \in \{\mathrm{F}, \mathrm{B}\}, \tag{3}$$

such that

$$[\boldsymbol{E}_c]_j \triangleq \left[\boldsymbol{X}_c - \sum_{s\neq j} [\boldsymbol{d}_c]_s \boldsymbol{\alpha}_j^T \right]_{\boldsymbol{f_j}} \text{ and } \boldsymbol{f_j} = \{i | [\boldsymbol{\alpha}_j^T]_i \neq 0\},$$

where $\boldsymbol{\alpha}_j^T$ is the jth row of $\boldsymbol{A}$, $\boldsymbol{f_j}$ is an indicator function that selects the non-zero entries in $\boldsymbol{\alpha}_j^T$, and the subscript c stands for the labels F (for focused) or B (for blurred training data). Since the dictionaries are used to classify the patches as either belonging to class F or B (see Subsection II-C), it is desirable that the dictionaries provide the best discriminative power. In this work, a discriminative constraint based on structured incoherency is added to (3). Note that this approach has been successfully used in [21] for the discriminative dictionary learning. The key idea is to add a constraint that enforces each dictionary to be weak at representing other classes. The incoherency term takes the form $\mathcal{C}(\boldsymbol{D}_\mathrm{F}, \boldsymbol{D}_\mathrm{B}) = \|\boldsymbol{D}_\mathrm{F}^T \boldsymbol{D}_\mathrm{B}\|_{\mathcal{F}}^2$. The dictionary update problem becomes

$$\begin{gathered}[\boldsymbol{d}_c]_j = \underset{[\boldsymbol{d}_c]_j}{\operatorname{argmin}} \left\| [\boldsymbol{E}_c]_j - [\boldsymbol{d}_c]_j \left[\boldsymbol{\alpha}_j^T\right]_{\boldsymbol{f_j}} \right\|_{\mathcal{F}}^2 + \lambda \|\boldsymbol{D}_h^T [\boldsymbol{d}_c]_j\|_2^2 \\ c \in \{\mathrm{F}, \mathrm{B}\}, h = \{\mathrm{F}, \mathrm{B}\} - c,\end{gathered} \tag{4}$$

where $\lambda > 0$ controls the trade-off between the reconstructive and discriminative properties of the dictionaries $\boldsymbol{D}_\mathrm{F}$ and $\boldsymbol{D}_\mathrm{B}$. Note that in (4), the incoherency term is formulated atom-wise. The corresponding atom update rule is then formulated as follows:

$$[\boldsymbol{d}_c]_j = \left(\lambda \boldsymbol{D}_h \boldsymbol{D}_h^T + \|\boldsymbol{\alpha}_j^T\|_2^2 \boldsymbol{I_d}\right)^{-1} [\boldsymbol{E}_c]_j [\boldsymbol{\alpha}_j^T]_{\boldsymbol{f_j}}^T, \; c \in \{\mathrm{F}, \mathrm{B}\} \tag{5}$$

where $\boldsymbol{I_d}$ denotes the identity matrix. For a more detailed description of the remaining steps of the CDL algorithm, the reader is referred to [22].

C. Fusion using sparse representation

A classical way of classifying input signals using sparse representation is by evaluating their reconstruction errors [20]. The key idea is that each element should be assigned to the class providing the smallest reconstruction error, *i.e.,* the best sparse representation. However, since multi-focus images are highly correlated, standard dictionary learning methods lead to a considerable overlap between the dictionaries $\boldsymbol{D}_\mathrm{F}$ and $\boldsymbol{D}_\mathrm{B}$, making classification impractical. The CDL framework described in Subsection II-B allows to overcome this issue by insuring that the learnt dictionaries are sufficiently independent and discriminative. In particular, classification can be performed using a concatenation of the coupled dictionaries $\left[\boldsymbol{D}_\mathrm{F}^T \; \boldsymbol{D}_\mathrm{B}^T\right]^T$ and $\left[\boldsymbol{D}_\mathrm{B}^T \; \boldsymbol{D}_\mathrm{F}^T\right]^T$. These matrices can be used as means of describing the function between two patches $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, *i.e.,* blurring and deblurring, respectively. Note that since identifying focused patches is a binary classification problem, it is sufficient to use one dictionary $\boldsymbol{D} \triangleq \left[\boldsymbol{D}_\mathrm{F}^T \; \boldsymbol{D}_\mathrm{B}^T\right]^T$.

1) Selection rule: In this work, the focused patches $\boldsymbol{x}_F$ are selected using the reconstruction error-based approach. More specifically, let e_1 and e_2 be the reconstruction errors associated with the blurring and deblurring functions, respectively. This can be formulated as

$$\begin{cases} e_1 = \| \left[\boldsymbol{x}_1^T \ \boldsymbol{x}_2^T\right]^T - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 \\ e_2 = \| \left[\boldsymbol{x}_2^T \ \boldsymbol{x}_1^T\right]^T - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2, \end{cases} \tag{6}$$

where $\boldsymbol{\alpha}$ contains the associated sparse codes. The selection rule is then based on the fact that a relatively smaller value of e_1 indicates that $\boldsymbol{x}_1$ has undergone blurring (resulting in $\boldsymbol{x}_2$). For the purpose of processing the boundary region (explained below), a pixel-wise sparse representation error is also defined as follows:

$$\begin{cases} \boldsymbol{e'_1} = (\boldsymbol{x}_1 - \boldsymbol{D}_{\mathrm{F}}\boldsymbol{\alpha})^2 + (\boldsymbol{x}_2 - \boldsymbol{D}_{\mathrm{B}}\boldsymbol{\alpha})^2 \\ \boldsymbol{e'_2} = (\boldsymbol{x}_1 - \boldsymbol{D}_{\mathrm{B}}\boldsymbol{\alpha})^2 + (\boldsymbol{x}_2 - \boldsymbol{D}_{\mathrm{F}}\boldsymbol{\alpha})^2, \end{cases} \tag{7}$$

The selection rule for the pair of patches $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ can then be expressed as

$$\begin{cases} \boldsymbol{x}_F = \boldsymbol{x}_1, & \text{if } e_1 < e_2 \\ \boldsymbol{x}_F = \boldsymbol{x}_2, & \text{otherwise} \end{cases} \tag{8}$$

Note that in (8) the equality condition is not taken into account since e_1 and e_2 cannot be equal.

2) Fusion and processing of boundary regions: Once the selection rule is applied to all the patches in the images, a patch-wise decision mask $\boldsymbol{M}_p$ is straightforwardly obtained (such that the patches in $\boldsymbol{M}_p$ contain values $c \in \{1, 2\}$). The pixel-wise decision mask is reconstructed using

$$\boldsymbol{M} = \mathcal{P}^*(\boldsymbol{M}_p), \tag{9}$$

where $\boldsymbol{M}$ now contains values $c' \in [1, 2]$. One could directly use $\boldsymbol{M}$ to form the final image $\boldsymbol{Y}$ by assigning pixels according to their labels at each location. However, this approach can lead to errors around the boundaries between blurred and focused regions, caused by the weighted averaging of patches containing both blurred and focused pixels. In order to bypass this issue, a pixel-wise labelling is proposed for these regions. More specifically, the mask $\boldsymbol{M}$ is used to first detect the boundary region, which contains all the pixels with labels in the interval $]1, 2[$. In the second step, the pixel-wise sparse coding errors (7) are used to assign new labels to the pixels within the boundary region.[1] Finally, the all-in-focus image $\boldsymbol{Y}$ is obtained by assigning pixels according to their labels in M to each location (k, l), as follows:

$$\boldsymbol{Y}_{kl} = (2 - \boldsymbol{M}_{kl})\boldsymbol{I}_{1,kl} + (\boldsymbol{M}_{kl} - 1)\boldsymbol{I}_{2,kl} \tag{10}$$

The different steps of the proposed image fusion algorithm are summarized in Algorithm 1.

Algorithm 1 Image Fusion using CDL

Input: Input images $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$, and learnt coupled dictionary $\boldsymbol{D} = [\boldsymbol{D}_{\mathrm{F}}^T, \boldsymbol{D}_{\mathrm{B}}^T]^T$.

1: Patch extraction: $\boldsymbol{X}_1 = \mathcal{P}(\boldsymbol{I}_1)$ and $\boldsymbol{X}_2 = \mathcal{P}(\boldsymbol{I}_2)$;
2: Subtract the mean of each patch: $\boldsymbol{x}_1 = \boldsymbol{x}_1 - \text{mean}(\boldsymbol{x}_1)$ and $\boldsymbol{x}_2 = \boldsymbol{x}_2 - \text{mean}(\boldsymbol{x}_2)$;
3: **for** each patch:
4: Find $\boldsymbol{\alpha}$ using OMP;
5: Compute the errors e_1 and e_2 in (6);
6: Find $\mathbf{x}_F$ by applying the selection rule (8);
7: **end for**
8: Reconstruct the decision mask using (9);
9: Process decision boundaries using the errors in (7);
10: Form the final all-in-focus image $\boldsymbol{Y}$ using the decision mask.

Output: The all-in-focus image $\boldsymbol{Y}$.

III. Experiments

In this section, the proposed method is compared with four recent fusion methods including state-of-the-art. Specifically, these methods include one transform domain approach using Laplacian pyramids (LP) [28], a spatial domain method using phase congruency (PC) [9], a sparse representation-based approach using convolutional dictionaries (CSR) [29], and a method for microscopic medical image fusion using mean-shift segmentation [32]. For the data without available ground-truth, the evaluation is based on the normalized mutual information (NMI) [26], the objective image fusion performance measure ($Q_{AB/F}$) [25] and Tone mapped index (TMQI) [27]. When ground-truth is available, we also use the mean-squared-error (MSE). First, experiments are conducted on two widely used multi-focus datasets referred to as *Lytro* [23] and *Grayscale* [24]. Note that the Grayscale dataset provides ground-truth all-in-focus images, which can be used to compute the MSE. In a second experiment, a sequence of real medical images is used to validate the proposed method. Specifically, a sequence of 15 partially blurred and noisy microscopy images is used [31]. All experiments are performed on a PC running an Intel(R) Xeon(R) 3.40GHz CPU.

The coupled dictionaries $\boldsymbol{D}_{\mathrm{F}}$ and $\boldsymbol{D}_{\mathrm{B}}$ are learnt using 40000 pairs of blurred and focused patches extracted from 4 images

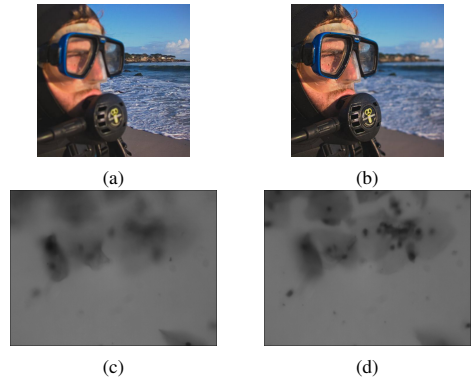


Fig. 1: Examples of multi-focus images from the Lytro dataset (a,b) and the microscopic medical image sequence (c,d).

[1]Figure 2-c shows an example of a mask obtained using the proposed approach.

of the Lytro dataset (the rest of the images are used for testing). The dictionary learning parameters are set empirically with a sparsity parameter $K = 10$, a redundancy of 4 and a patch size of 16×16. The resulting dictionaries are of size 256×1024. Finally, the maximum number of iterations of the CDL step is set to 20.

A. Fusion results using Lytro and Grayscale images

The fusion results are quantitatively compared in terms of average NMI, $Q_{AB/F}$, and MSE for the two considered datasets. The results reported in Table. I show that the best overall performance is provided by the proposed and PC methods for both datasets. More specifically, the PC and proposed approaches lead to similar NMI and $Q_{AB/F}$ values, while the proposed method results in significantly lower MSE. One can also see that the proposed method provides competitive execution times.

Fig. 2 shows the resulting all-in-focus images obtained for the input images in Fig. 1-(a,b) using the proposed and PC methods. Note that the mask is first computed for the grayscale version of the images before applying it to the RGB layers. A visual inspection of the resulting all-in-focus images shows that the proposed method allows the edges to be preserved. In particular, the boundaries between blurred and focused regions are sharper, as can be clearly observed in the magnified regions in Fig. 2-d and Fig. 2-e.

	Dataset	$Q_{AB/F}$	NMI	$TMQI$	MSE	Execution time (s)
LP	Gray	0.7434	1.0406	0.9347	6.0308	0.0056
	Lytro	0.7524	1.0306	0.6628	–	0.0107
PC	Gray	0.7535	1.2216	0.9321	6.8984	0.5104
	Lytro	0.7397	1.2089	0.6648	–	1.0557
CSR	Gray	0.7198	1.0296	0.9292	4.6927	54.8437
	Lytro	0.7304	1.0340	0.6619	–	99.1661
Us	Gray	0.7512	1.1772	0.9331	3.3453	1.4180
	Lytro	0.7561	1.1913	0.6628	–	2.5627

TABLE I: The average results of NMI, $Q_{AB/F}$, TMQI, MSE and execution time for the Lytro and Grayscale datesets.

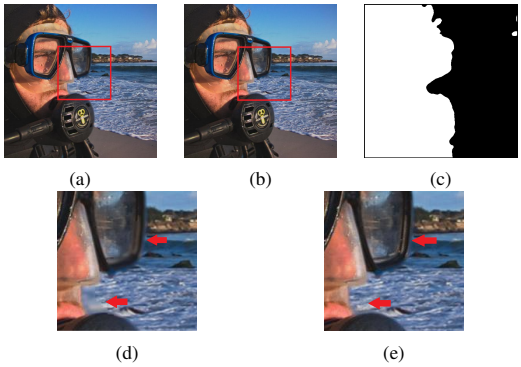


Fig. 2: Fusion results by (a) PC and (b) the proposed method, the corresponding magnified regions (d,e), and (c) the mask obtained using the proposed method.

B. Fusion of in vivo Pap Smear images

Papanicolaou test (Pap smear) images are used for the automated diagnosis of cervical cancer. However, these images are characterized by the presence of different focus levels due to the limited depth of field of the microscope. In order to bypass this limitation, multiple images with different focal settings are acquired and fused into one all-in-focus image [30]. In this work, we use a sequence of 15 multi-focus Pap smear images (of size 480×640 pixels). The sequence is processed using a single-elimination approach. This means that the fusion is conducted sequentially and pair-wise, *i.e.,* the fusion result from each pair of images is in turn fused with the next input image. The quantitative evaluation of the results obtained using the proposed method, PC, and the method of [32] are summarized in Table. II. The resulting fused images are displayed in Fig. 3.

The results reported in Table. II show that the proposed method provides higher $Q_{AB/F}$ and NMI values, corresponding to more edge information and a higher fidelity of pixel intensities. Fig. 3 confirms these findings, as one can see that the proposed method provides sharper edges (green arrows) and preserves details (red arrows) that are missing in the image obtained by the method of [32].

	$Q_{AB/F}$	NMI	$TMQI$
The method of [32]	0.6063	5.9135	0.7949
PC	0.6212	6.9133	0.7739
The proposed method	0.6257	7.1941	0.7755

TABLE II: Fusion performance for the Pap smear images.

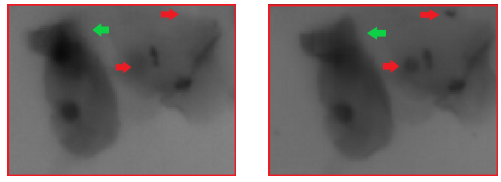


Fig. 3: Magnified region in the fused image obtained by (a) the method of [32], and (b) the proposed method.

IV. Conclusion

This paper has introduced an image fusion algorithm using coupled dictionary learning and joint sparse representation. First, a coupled dictionary learning approach with an additional incoherency constraint has been used to learn blurred and focused dictionaries. Secondly, the patch-wise sparse representation errors have been used to construct a fusion rule for input patches with unknown focus levels. In addition, a pixel-wise processing of the boundary regions has been proposed. Experiments have been conducted using two natural image datasets and a sequence of *in vivo* microscopic images (Pap smear). A comparison with state-of-the-art approaches has shown the competitiveness of the proposed method in terms of various image fusion metrics.

REFERENCES

[1] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Opt. Eng.*, vol. 32, pp. 2824–2836, Mar. 1993.

[2] M. Born and E. Wolf, *Principles of Optics*. *Cambridge Univ. Press.*, 1999.

[3] Q. Zhang, and B. L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, pp. 1334–1346, Jul. 2009.

[4] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Apr. 2007.

[5] G. Pajares and J. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[6] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proc. IEEE Int. Conf. Image Process.*, Santa Barbara, CA, 1997, pp. 288–291.

[7] V. D. Calhoun and T. Adali, "Feature-based fusion of medical imaging data," in *IEEE Trans. Inf. Technol. Biomedicine*, vol. 13, no. 5, pp. 711–720, Sep. 2009.

[8] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhuang, and Y. Zhang, "Multi-Focus Image Fusion Based on Spatial Frequency in Discrete Cosine Transform Domain," in *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 220–224, Sep. 2014.

[9] K. Zhan, Q. Li, J. Teng, M. Wang, and J. Shi, "Multifocus image fusion using phase congruency," in *Journal of Electronic Imaging*, vol. 24, no. 3, pp. 0330141–03301412, May. 2015.

[10] H. Li, L. Li, and J. Zhang, "A novel DWT based multi-focus image fusion method," in *Procedia Eng.*, vol. 24, pp. 177–181, 2011.

[11] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," in *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.

[12] M. Nejati, S. Samavi, and S. Hirani, "Multi-focus image fusion using dictionary-based sparse representation," in *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.

[13] Q. Zhang, and M. D. Levine, "Robust multi-focus image fusion using multi-task sparse representation and spatial context," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2045–2058, Mar. 2016.

[14] R. Gao, S. A. Vorobyov, and H. Zhao, "Multi-focus image fusion via coupled dictionary training," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Shanghai, China, 2016, pp. 1666–1670.

[15] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, May. 2010.

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

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

[18] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Rhode Island, USA, 2012, pp. 2216–2223.

[19] J. A. Tropp, and A.C. Gilbert,"Signal recovery From random measurements via orthogonal matching pursuit," in *Proc. IEEE Trans. Inf. Proc.*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[20] K. Skretting, and J. H. Husoy," Texture classification using sparse frame-based representations," in *EURASIP Journal on Advances in Signal Processing,*, vol. 2006, pp. 102–112, 2006.

[21] I. Ramirez, P. Sprechmann, and G. Sapiro,"Classification and clustering via dictionary learning with structured incoherence and shared features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, pp. 3501–3508, Jun. 2010.

[22] F. G. Veshki, and S. A. Vorobyov, "An Efficient Coupled Dictionary Learning Method," in *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1441–1445, Aug. 2019.

[23] M. Nejati, S. Samavi, and S. Hirani, "Lytro Multi Focus Dataset," [Online]. Available: http://mansournejati.ece.iut.ac.ir/content/lytro-multi-focus-dataset. [Accessed Oct. 16, 2019].

[24] J. Saeedi, K. Faez, "Multi Focus Image Dataset," [Online]. Available: https://www.researchgate.net/profile/Jamal_Saeedi/publication/273000238_multifocus_image_dataset/data/54f489b80cf2ba6150635697/multi-focus-image-dataset.rar. [Accessed Oct. 16, 2019].

[25] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[26] M. Hossny, S. Nahavandi and D. Creighton, "Comments on Information measure for performance of image fusion," *Electron. Lett.*, pp. 1066–1067, 2008.

[27] H. Yeganeh, and Z. Wang, "Objective Quality Assessment of Tone Mapped Images," *IEEE Transactios on Image Processing*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

[28] W. Wang and F. Chang, "A Multi-focus Image Fusion Method Based on Laplacian Pyramid," *Journal of Computers*, vol. 6, no. 12, pp. 2559–2566, 2011.

[29] Y. Liu, X. Chen, R. K. Ward and Z. J. Wang, "Image Fusion With Convolutional Sparse Representation," in *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[30] J. Tian and L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure," in *Signal Processing*, vol. 92, no. 9, pp. 2137–2146, 2012.

[31] S. Tello-Mijares, "Multi focus image fusion," [Online]. Available: https://drive.google.com/drive/folders/1bcrJM8Kw0mzKF8VKpHg1WVlOxWGulwdj. [Accessed Feb. 2, 2020].

[32] S. Tello-Mijares and J. Bescos, "Region-based multifocus image fusion for the precise acquisition of Pap smear images," in *Journal of Biomedical Optics*, vol. 23, no. 5, pp. ()056005)1–9, May. 2018.

Publication III

F. G. Veshki and S. A. Vorobyov. Efficient ADMM-Based Algorithms for Convolutional Sparse Coding. *IEEE Signal Processing Letters*, vol. 29, pp. 389-393, 2021.

Efficient ADMM-Based Algorithms for Convolutional Sparse Coding

Farshad G. Veshki and Sergiy A. Vorobyov, *Fellow, IEEE*

***Abstract*—Convolutional sparse coding improves on the standard sparse approximation by incorporating a global shift-invariant model. The most efficient convolutional sparse coding methods are based on the alternating direction method of multipliers and the convolution theorem. The only major difference between these methods is how they approach a convolutional least-squares fitting subproblem. In this letter, we present a novel solution for this subproblem, which improves the computational efficiency of the existing algorithms. The same approach is also used to develop an efficient dictionary learning method. In addition, we propose a novel algorithm for convolutional sparse coding with a constraint on the approximation error. Source codes for the proposed algorithms are available online.**

***Index Terms*—Convolutional sparse coding, constrained sparse approximation, dictionary learning, alternating direction method of multipliers.**

I. Introduction

SPARSE representations are widely used in various applications of signal and image processing [1]–[6]. The sparse synthesis model admits that natural signals can be approximated using a linear combination of only a small number of atoms (columns) of a dictionary (matrix). A common formulation of the sparse coding problem is given as

$$\underset{\boldsymbol{x}}{\text{minimize}}\ \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{D}\boldsymbol{x}-\boldsymbol{s}\|_2^2 \le \epsilon, \tag{1}$$

where $\boldsymbol{D}$ is the dictionary, $\boldsymbol{x}\in\mathbb{R}^m$ is the sparse representation vector, $\boldsymbol{s}\in\mathbb{R}^n$ is the signal, and ϵ represents the upper bound on the approximation error. Moreover, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_1-norm and the Euclidean norm, respectively. The problem of finding sparsity promoting dictionaries is called dictionary learning [7], [8].

The applications of sparse representations and dictionary learning usually involve either or both extraction and estimation of local features. Typically, this is handled by a prior decomposition of the original signal into vectorized overlapping blocks (*e.g.*, patches in image processing). However, this strategy results in multi-valued representations. Moreover, since the relationships among neighboring blocks are ignored, dictionaries learned using this approach contain shifted versions of the same features.

Manuscript received December 3, 2021; accepted December 9, 2021. Date of publication December 14, 2021; date of current version January 28, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shiwen He. *(Corresponding author: Sergiy A. Vorobyov.)*

The authors are with the Department of Signal Processing and Acoustics, Aalto University, Espoo 02150, Finland (e-mail: farshad.ghorbaniveshki@aalto.fi; svor@ieee.org).

Digital Object Identifier 10.1109/LSP.2021.3135196

Convolutional sparse coding (CSC) incorporates a single-valued and shift-invariant model that represents the entire signal. In this model, the product $\boldsymbol{D}\boldsymbol{x}$ in the standard sparse coding problem is replaced by a sum of convolutions. The convolutional form of the standard sparse coding problem (1) can be written as follows

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}}\ \sum_{k=1}^K \|\boldsymbol{x}_k\|_1 \quad \text{s.t.} \quad \Big\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s}\Big\|_2^2 \le \epsilon, \tag{2}$$

where $*$ denotes the convolution operator (usually, with "$same$" padding) and $\boldsymbol{x}_k\in\mathbb{R}^n$ and $\boldsymbol{d}_k\in\mathbb{R}^m,\ k=1,\dots,K$, are the sparse coefficient maps and the dictionary filters, respectively. Several applications have shown that CSC outperforms its standard version in modeling natural signals, such as audio and images [9]–[20].

A majority of available CSC algorithms, including [21]–[29], are based on the alternating direction method of multipliers (ADMM) [30]. ADMM breaks the CSC problem into two main sub-problems, one of which is a sparse approximation problem which can be efficiently addressed using a shrinkage operator, and the other entails a convolutional least-squares regression. An efficient solution to the second sub-problem based on the convolution theorem and the Sherman-Morrison formula is given in [23]. CSC problem (2) is typically addressed by solving its unconstrained equivalent

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}}\ \frac{1}{2}\Big\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s}\Big\|_2^2 + \lambda\sum_{k=1}^K \|\boldsymbol{x}_k\|_1, \tag{3}$$

where $\lambda>0$ is a regularization parameter. It is known that there is a unique λ for each ϵ. However, the appropriate value of λ also depends on $\boldsymbol{s}$ and $\{\boldsymbol{d}_k\}_{k=1}^K$. Thus, despite being more convenient to solve, the unconstrained reformulation (3) introduces undesirable data dependency to the CSC algorithm.

A common approach for convolutional dictionary learning (CDL) entails optimizing the filters and the sparse coefficient maps using a batch of P training signals [22]–[25]. This problem can be formulated as

$$\underset{\{\boldsymbol{x}_k^p\}_{k=1}^K,\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ \sum_{p=1}^P\left(\frac{1}{2}\Big\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k^p - \boldsymbol{s}^p\Big\|_2^2 + \lambda\sum_{k=1}^K \|\boldsymbol{x}_k^p\|_1\right)$$
$$\text{s.t.}\quad \{\boldsymbol{d}_k\}_{k=1}^K \in \boldsymbol{\mathcal{D}}, \tag{4}$$

where $\boldsymbol{\mathcal{D}}=\{\boldsymbol{d}_k \mid \|\boldsymbol{d}_k\|_2\le 1,\ k=1,\dots,K\}$. The CDL problem is usually addressed by alternating optimization with respect to $\{\boldsymbol{x}_k^p\}_{k=1}^K$ and $\{\boldsymbol{d}_k\}_{k=1}^K$ [21]–[23]. Several works have shown that (4) can be solved for $\{\boldsymbol{d}_k\}_{k=1}^K$ effectively and efficiently using ADMM in frequency domain [31].

The contributions of this letter are summarized as follows: (i) we present an efficient approach for solving the convolutional

least-squares fitting which leads to a constant improvement on the complexity of the existing CSC algorithms; (ii) we use the proposed solution to improve the efficiency of existing CDL methods; (iii) we propose a novel algorithm for solving the CSC problem with a constraint on the approximation error based on our solution to the unconstrained CSC problem. MATLAB implementations of the proposed algorithms are available at GitHub repository [32].

II. Proposed Algorithms

A. Unconstrained CSC

The ADMM formulation of the unconstrained CSC problem (3) can be written in the form

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K,\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2}\Big\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{z}_k - \boldsymbol{s}\Big\|_2^2 + \lambda \sum_{k=1}^K \|\boldsymbol{x}_k\|_1$$

$$\text{s.t.} \quad \boldsymbol{z}_k = \boldsymbol{x}_k, \quad k = 1, \ldots, K.$$

The ADMM iterations are

$$\{\boldsymbol{z}_k^{t+1}\}_{k=1}^K = \underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{argmin}} \frac{1}{2}\Big\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{z}_k - \boldsymbol{s}\Big\|_2^2 + \frac{\rho}{2}\sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{x}_k^t + \boldsymbol{u}_k^t\|_2^2$$

$$\{\boldsymbol{x}_k^{t+1}\}_{k=1}^K = \underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{argmin}} \lambda \sum_{k=1}^K \|\boldsymbol{x}_k\|_1 + \frac{\rho}{2}\sum_{k=1}^K \|\boldsymbol{z}_k^{t+1} - \boldsymbol{x}_k + \boldsymbol{u}_k^t\|_2^2$$

$$\boldsymbol{u}_k^{t+1} = \boldsymbol{u}_k^t + \boldsymbol{z}_k^{t+1} - \boldsymbol{x}_k^{t+1}, \qquad k = 1, \ldots, K,$$

where $\rho > 0$ is the penalty parameter and $\{\boldsymbol{u}_k\}_{k=1}^K$ are the scaled Lagrangian multipliers. The second subproblem ($\boldsymbol{x}$-update step) can be addressed in an element-wise manner using the shrinkage operator

$$\boldsymbol{x}_k^{t+1} = \mathcal{S}_{\lambda/\rho}\left(\boldsymbol{z}_k{}^{t+1} + \boldsymbol{u}_k^t\right), \qquad k = 1, \ldots, K.$$

For completeness, it is reminded the shrinkage operator is defined as $\mathcal{S}_\kappa(a) = \text{sign}(a)\max(0, |a| - \kappa)$.

The challenging step is solving the first subproblem ($\boldsymbol{z}$-update), which entails solving the optimization problem

$$\underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2}\left\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{z}_k - \boldsymbol{s}\right\|_2^2 + \frac{\rho}{2}\sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{w}_k\|_2^2. \quad (5)$$

Using the convolution theorem, problem (5) in Fourier domain can be written as

$$\underset{\{\hat{\boldsymbol{z}}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2}\left\|\sum_{k=1}^K \hat{\boldsymbol{d}}_k \odot \hat{\boldsymbol{z}}_k - \hat{\boldsymbol{s}}\right\|_2^2 + \frac{\rho}{2}\sum_{k=1}^K \|\hat{\boldsymbol{z}}_k - \hat{\boldsymbol{w}}_k\|_2^2, \quad (6)$$

where $\hat{(\cdot)}$ denotes the discrete Fourier transform of a signal and $\odot$ stands for the element-wise multiplication operator. Note that the filters $\{\boldsymbol{d}_k\}_{k=1}^K$ are zero-padded to the size of $\{\boldsymbol{z}_k\}_{k=1}^K$ before performing the discrete Fourier transform.

Denoting $\boldsymbol{\delta}_i = [\hat{\boldsymbol{d}}_1(i), \ldots, \hat{\boldsymbol{d}}_k(i)]^T$, $\boldsymbol{\zeta}_i = [\hat{\boldsymbol{z}}_1(i), \ldots, \hat{\boldsymbol{z}}_k(i)]^T$, $\boldsymbol{\omega}_i = [\hat{\boldsymbol{w}}_1(i), \ldots, \hat{\boldsymbol{w}}_k(i)]^T$, $i = 1, \ldots, n$, where $(\cdot)^T$ is the (non-conjugate) transpose operator, we can see that problem (6) can be addressed as n independent problems

$$\underset{\boldsymbol{\zeta}_i}{\text{minimize}} \frac{1}{2}\left(\boldsymbol{\delta}_i^T\boldsymbol{\zeta}_i - \hat{s}_i\right)^2 + \frac{\rho}{2}\|\boldsymbol{\zeta}_i - \boldsymbol{\omega}_i\|_2^2.$$

Equating the derivative with respect to $\boldsymbol{\zeta}_i$ to zero, we have

$$\begin{aligned} 0 &= \bar{\boldsymbol{\delta}}_i\left(\boldsymbol{\delta}_i^T\boldsymbol{\zeta}_i - \hat{s}_i\right) + \rho\boldsymbol{\zeta}_i - \rho\boldsymbol{\omega}_i \\ &= (\bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T + \rho\boldsymbol{I})\boldsymbol{\zeta}_i - \hat{s}_i\bar{\boldsymbol{\delta}}_i - \rho\boldsymbol{\omega}_i \\ &= (\bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T + \rho\boldsymbol{I})\boldsymbol{\zeta}_i - (\hat{s}_i\bar{\boldsymbol{\delta}}_i - \bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T\boldsymbol{\omega}_i) - (\bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T + \rho\boldsymbol{I})\boldsymbol{\omega}_i, \end{aligned} \quad (7)$$

which gives

$$\begin{aligned} \boldsymbol{\zeta}_i^\star &= \boldsymbol{\omega}_i + (\hat{\boldsymbol{s}}_i - \boldsymbol{\delta}_i^T\boldsymbol{\omega}_i)(\bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T + \rho\boldsymbol{I})^{-1}\bar{\boldsymbol{\delta}}_i \\ &= \boldsymbol{\omega}_i + (\hat{\boldsymbol{s}}_i - \boldsymbol{\delta}_i^T\boldsymbol{\omega}_i)(\|\boldsymbol{\delta}_i\|_2^2 + \rho)^{-1}\bar{\boldsymbol{\delta}}_i, \end{aligned} \quad (8)$$

where $(\cdot)^\star$ denotes the solution to an optimization problem and $\bar{(\cdot)}$ is the complex-conjugate of a complex number.

Denoting

$$\hat{\boldsymbol{c}}_k^\rho = \bar{\hat{\boldsymbol{d}}}_k \oslash \left(\rho + \sum_{k=1}^K \bar{\hat{\boldsymbol{d}}}_k \odot \hat{\boldsymbol{d}}_k\right), \quad \hat{\boldsymbol{r}} = \hat{\boldsymbol{s}} - \sum_{k=1}^K \hat{\boldsymbol{d}}_k \odot \hat{\boldsymbol{w}}_k \quad (9)$$

(here $\oslash$ stands for the element-wise division operator), the solution to the $\boldsymbol{z}$-update step based on (8) can be found as

$$\hat{\boldsymbol{z}}_k^\star = \hat{\boldsymbol{w}}_k + \hat{\boldsymbol{c}}_k^\rho \odot \hat{\boldsymbol{r}}. \quad (10)$$

Computational Complexity: The available ADMM-based CSC algorithms usually address the $\boldsymbol{z}$-update step as

$$\boldsymbol{\zeta}_i^\star = (\bar{\boldsymbol{\delta}}_i\boldsymbol{\delta}_i^T + \rho\boldsymbol{I})^{-1}(\hat{s}_i\bar{\boldsymbol{\delta}}_i + \rho\boldsymbol{\omega}_i), \quad (11)$$

which can be inferred from the second line of (7). Computing (11) using matrix inversion results in a complexity of $\boldsymbol{\mathcal{O}}(K^3)$ [21]. However, the work of [23] demonstrated that this can be reduced to $\boldsymbol{\mathcal{O}}(K)$ using the Sherman-Morrison formula. The complexity of the proposed method is also of $\boldsymbol{\mathcal{O}}(K)$. However, using further simplifications, the proposed approach eliminates the need for explicit matrix inversion and requires fewer computations. In particular, performing the $\boldsymbol{z}$-update step on a batch of P images using the proposed method requires $((4K+1)P + 3K + 1)n$ flops, while it takes $(7KP + 3K + 1)n$ flops using the method of [23], indicating a considerable improvement provided by our method.

B. CSC With a Constraint on the Approximation Error

The ADMM reformulation of problem (2) is given as

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K,\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}} \boldsymbol{f}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) + \sum_{k=1}^K \|\boldsymbol{x}_k\|_1 \text{ s.t. } \boldsymbol{z}_k = \boldsymbol{x}_k, \ \forall k,$$

where $\boldsymbol{f}(\cdot)$ is an indicator function of the constraint set in (3), that is,

$$\boldsymbol{f}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) = \begin{cases} 0, & \text{if } \ \boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) \le \epsilon \\ \infty, & \text{otherwise} \end{cases},$$

with

$$\boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) = \left\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{z}_k - \boldsymbol{s}\right\|_2^2. \quad (12)$$

The ADMM iterations are

$$\{\boldsymbol{z}_k^{t+1}\}_{k=1}^K = \underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{argmin}} \boldsymbol{f}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right) + \frac{\rho}{2}\sum_{k=1}^K \|\boldsymbol{z}_k - \boldsymbol{x}_k^t + \boldsymbol{u}_k^t\|_2^2$$

$$\{\boldsymbol{x}_k^{t+1}\}_{k=1}^K = \underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{argmin}} \sum_{k=1}^K \|\boldsymbol{x}_k\|_1 + \frac{\rho}{2}\sum_{k=1}^K \|\boldsymbol{z}_k^{t+1} - \boldsymbol{x}_k + \boldsymbol{u}_k^t\|_2^2$$

$$\boldsymbol{u}_k^{t+1} = \boldsymbol{u}_k^t + \boldsymbol{z}_k^{t+1} - \boldsymbol{x}_k^{t+1}, \qquad k = 1, \ldots, K.$$

The $\boldsymbol{z}$-update step requires solving the following optimization problem

$$\underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}}\ \boldsymbol{f}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right)+\frac{\rho}{2}\sum_{k=1}^K\|\boldsymbol{z}_k-\boldsymbol{w}_k\|_2^2. \tag{13}$$

Depending on $\{\boldsymbol{w}_k\}_{k=1}^K$, problem (13) either has a trivial solution or it is equivalent to an equality-constrained optimization problem. This can be expressed as

$$\{\boldsymbol{z}_k^\star\}_{k=1}^K= \begin{cases}\{\boldsymbol{w}_k\}_{k=1}^K, & \text{if } \boldsymbol{e}\left(\{\boldsymbol{w}_k\}_{k=1}^K\right)\le\epsilon\\ \underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{argmin}}\sum_{k=1}^K\|\boldsymbol{z}_k-\boldsymbol{w}_k\|_2^2 & \text{s.t. } \boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right)=\epsilon, \quad \text{otherwise}\end{cases}. \tag{14}$$

Using a suitable regularization parameter ν, the problem in the second term of (14) can be reformulated as

$$\underset{\{\boldsymbol{z}_k\}_{k=1}^K}{\text{minimize}}\ \boldsymbol{e}\left(\{\boldsymbol{z}_k\}_{k=1}^K\right)+\nu\sum_{k=1}^K\|\boldsymbol{z}_k-\boldsymbol{w}_k\|_2^2, \tag{15}$$

which has the same form as problem (5). Finding the solution of (15) using (10) and plugging it into (12) leads to

$$\boldsymbol{e}\left(\{\boldsymbol{z}_k^\star\}_{k=1}^K\right)=\frac{\nu^2}{n}\left\|\hat{\boldsymbol{r}}\oslash\left(\nu+\sum_{k=1}^K\bar{\hat{\boldsymbol{d}}}_k\odot\hat{\boldsymbol{d}}_k\right)\right\|_2^2,$$

where the division by n is required by Parseval's theorem. Thus, problem (13) is simplified to a single-variable optimization problem for finding the optimal parameter $\nu^\star$, which satisfies $\nu^\star=\left\{\nu\mid\boldsymbol{e}\left(\{\boldsymbol{z}_k^\star\}_{k=1}^K\right)=\epsilon\right\}$. Considering that $\boldsymbol{e}(\{\boldsymbol{z}_k^\star\}_{k=1}^K)$ is strictly monotonically increasing in $\nu>0$, this problem can be efficiently addressed, for example, using the *secant* method. Once $\nu^\star$ is known, the $\boldsymbol{z}$-update can be performed as $\hat{\boldsymbol{z}}_k^\star=\hat{\boldsymbol{w}}_k+\hat{\boldsymbol{c}}_k^{\nu^\star}\odot\hat{\boldsymbol{r}}, k=1,\ldots,K$, where $\hat{\boldsymbol{c}}_k^{\nu^\star}$ and $\hat{\boldsymbol{r}}$ are calculated using (9).

C. Dictionary Update

Addressing CDL problem (4) over $\{\boldsymbol{d}_k\}_{k=1}^K$ is equivalent to solving the optimization problem

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ \frac{1}{2}\sum_{p=1}^P\left\|\sum_{k=1}^K\boldsymbol{d}_k*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2+\boldsymbol{\Omega}\Big(\{\boldsymbol{d}_k\}_{k=1}^K\Big), \tag{16}$$

where $\boldsymbol{\Omega}\big(\boldsymbol{d}_k\big)$ is an indicator function associated with the constraint set in (4). Problem (16) can be efficiently addressed using the consensus ADMM method [31]. The consensus ADMM formulation of (16) is given as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ \frac{1}{2}\sum_{p=1}^P\left\|\sum_{k=1}^K\boldsymbol{g}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2+\boldsymbol{\Omega}\left(\{\boldsymbol{d}_k\}_{k=1}^K\right)$$

$$\text{s.t.}\quad \boldsymbol{g}_k^p=\boldsymbol{d}_k,\qquad k=1,\ldots,K,\quad p=1,\ldots,P$$

with the ADMM iterations

$$\{\boldsymbol{g}_k^{p,t+1}\}_{k=1}^K$$

$$=\underset{\{\boldsymbol{g}_k^p\}_{k=1}^K}{\text{argmin}}\left(\frac{1}{2}\left\|\sum_{k=1}^K\boldsymbol{g}_k^p*\boldsymbol{x}_k^p-\boldsymbol{s}^p\right\|_2^2+\frac{\sigma}{2}\sum_{k=1}^K\|\boldsymbol{g}_k^p-\boldsymbol{d}_k^t+\boldsymbol{v}_k^{p,t}\|_2^2\right)$$

$$\{\boldsymbol{d}_k^{t+1}\}_{k=1}^K$$

$$=\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{argmin}}\left(\boldsymbol{\Omega}\left(\{\boldsymbol{d}_k\}_{k=1}^K\right)+\frac{\sigma}{2}\sum_{k=1}^K\|\boldsymbol{d}_k-\frac{1}{P}\sum_{p=1}^P(\boldsymbol{g}_k^{p,t+1}+\boldsymbol{v}_k^{p,t})\|_2^2\right)$$

$$\boldsymbol{v}_k^{p,t+1}=\boldsymbol{v}_k^{p,t}+\boldsymbol{g}_k^{p,t+1}-\boldsymbol{d}_k^{t+1},\ k=1,\ldots,K,\ p=1,\ldots,P.$$

The first subproblem ($\boldsymbol{g}$-update) is similar to problem (5). Thus, it can be efficiently addressed using the proposed approach in Section II-A. The use of the Fourier domain-based approach requires $\{\boldsymbol{g}_k^p\}_{k=1}^K$ to be the same size as $\{\boldsymbol{x}_k^p\}_{k=1}^K$. Hence, the filters $\{\boldsymbol{d}_k\}_{k=1}^K$ are zero-padded to the size of $\{\boldsymbol{x}_k^p\}_{k=1}^K$ to be conformable with $\{\boldsymbol{g}_k^p\}_{k=1}^K$. Subproblem $\boldsymbol{d}$-update can be solved by projecting $\frac{1}{P}\sum_{p=1}^P(\boldsymbol{g}_k^{p,t+1}+\boldsymbol{v}_k^{p,t})$ onto the constraint set. This can be done simply by mapping the entries outside the constraint support to zero before normalizing the ℓ_2-norm. This approach can be also used for learning multiscale dictionaries, *i.e.*, filters with different sizes.

D. CDL Algorithm

CDL problem given by (4) is addressed using alternation approach (by alternating between CSC (see Section II-A) and dictionary update (see Section II-C) subproblems) to find a local optimum. We use a single iteration for each subproblem (the updated ADMM variables are used to initiate the succeeding iterations). This approach has been shown to be effective while simplifying the algorithm [23], [31]. We also use the variable coupling approach suggested in [33] which is shown to provide a better numerical stability [23], [31]. Specifically, the sparse (shrunk) variable $\{\boldsymbol{x}_k^p\}_{k=1}^K$ and the constrained filters $\{\boldsymbol{d}_k\}_{k=1}^K$ are passed to the next subproblem. The dictionary can be initialized using norm-normalized Gaussian random filters. All other ADMM variables are initialized using zero arrays of appropriate sizes.

The performances of the proposed algorithms can be substantially improved using ADMM extensions such as *over-relaxation* [30, Section 3.4.3] and *varying penalty parameter* [30, Section 3.4.1]. The work of [23] provides detailed explanations of how these extensions can be incorporated into ADMM-based CSC and CDL algorithms.

III. Experimental Results

In this section, we first compare the proposed unconstrained CSC algorithm with the state-of-the-art method, which uses the Sherman-Morrison formula in convolutional fitting step (the SM method) [23]. Then, we compare our unconstrained and constrained CSC methods in terms of convergence speed. Finally, we compare the proposed CDL algorithm with three available methods. All methods are based on the same alternating approach explained in Section II-D and use ADMM in both phases (CSC and dictionary update). All compared methods use the SM method in CSC phase. The compared dictionary learning methods are based on the conjugate gradient method (CG) [23], the iterative Sherman-Morrison method (ISM) [23] and a method based on the consensus ADMM framework and the Sherman-Morrison formula (SM-cns) [31].

A 512×512 greyscale Lena image is used in the CSC experiments. The CDL experiments are performed using a dataset of 20 images taken from the USC-SIPI database [34]. All images in the dataset are converted to greyscale and resized to 256×256

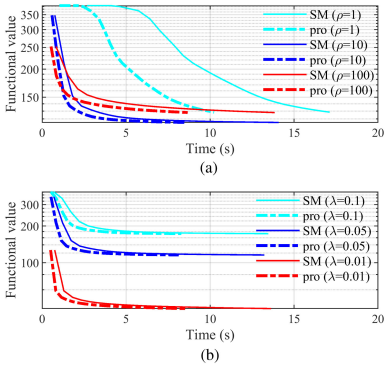


Fig. 1. Functional values over time for the proposed unconstrained CSC method (pro) and the SM method using (a) different values of ρ for $\lambda = 0.05$, and (b) different values of λ for $\rho = 10$. A dictionary of 16 filters of size 8×8 is used in both cases.

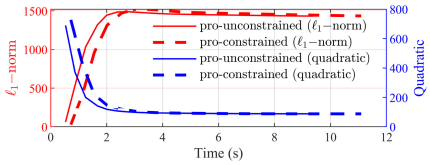


Fig. 2. The quadratic and ℓ_1-norm functional values for the proposed unconstrained and constrained CSC methods using $\lambda = 0.05$ ($\epsilon = 88.1886$), $\rho = 10$. A dictionary of 16 filters of size 8×8 is used.

pixels. All methods are implemented using MATLAB. All experiments are conducted on a PC equipped with an Intel(R) Core(TM) i5-8365 U 1.60 GHz CPU.

A. CSC Results

Fig. 1 shows the functional values over time for 25 iterations of the proposed unconstrained CSC method and the SM method using different values of ρ and λ tested. We use a fixed number of iterations to display the deference in efficiencies (the iterations of the two methods are equally effective). As it can be seen, the proposed method is significantly more efficient for all λ and ρ values. The algorithm complexities have been compared in Section II-A.

The proposed constrained and unconstrained CSC methods are compared in Fig. 2. Specifically, we executed the unconstrained CSC method using $\lambda = 0.05$, then we used the observed quadratic functional value ($\epsilon = 88.1886$) to run our constrained CSC method, while keeping the rest of the parameters unchanged. As it can be seen, the quadratic and the ℓ_1-norm functionals converge to the same values in both CSC methods. The constrained method results in a longer runtime, which accounts for optimization with respect to ν in each iteration.

B. CDL Results

In Fig. 3, the functional values over time for 50 iterations of all CDL methods using different dataset sizes P are compared. The complexity of the ISM method is of $\mathcal{O}(KP^2)$, which makes it inefficient when P is large. CG improves scalability, but slows

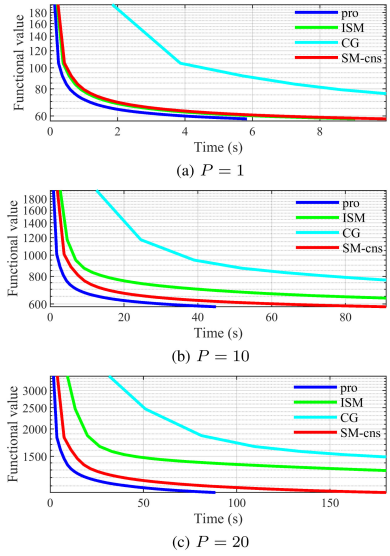


Fig. 3. Functional values over time using different values of P, $\rho = 10$, $\lambda = 0.05$, $K = 16$ filters of size 8×8.

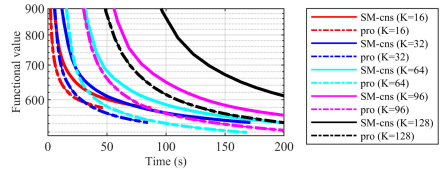


Fig. 4. Functional values over time using different K values, $P = 10$, $\rho = 10$, $\lambda = 0.05$ and filters of size 8×8.

down the convergence. The complexities of the proposed method and SM-cns are both of $\mathcal{O}(KP)$, and their iterations are equally effective. However, as it can be seen, the proposed method is substantially faster. This is achieved by using the method explained in Section II-A instead of the Sherman-Morrison formula, in both the $\boldsymbol{z}$-update step (CSC phase) and the $\boldsymbol{g}$-update step (dictionary update phase).

In Fig. 4, the convergence speeds of the proposed CDL method and SM-cns using different dictionary sizes K are compared. The improved computational efficiency of the proposed method can be clearly observed.

IV. Conclusion

An efficient solution for the convolutional least-squares fitting problem has been presented. The proposed method has been used to substantially improve the efficiency of the state-of-the-art convolutional sparse coding and dictionary learning algorithms. In addition, a novel method for convolutional sparse approximation with a constraint on the approximation error has been proposed.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[2] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[6] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195–208, 2015.

[7] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Phoenix, AZ, USA, 1999, pp. 2443–2446.

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[9] A. Cogliati, Z. Duan, and B. Wohlberg, "Piano transcription with convolutional sparse lateral inhibition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 392–396, Apr. 2017.

[10] P. Jao, L. Su, Y. Yang, and B. Wohlberg, "Monaural music source separation using convolutional sparse coding," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 11, pp. 2158–2170, Nov. 2016.

[11] P. Bao *et al.*, "Convolutional sparse coding for compressed sensing CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2607–2619, Nov. 2019.

[12] Y. Liu, X. Chen, R. Ward, and Z. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[13] X. Hu, F. Heide, Q. Dai, and G. Wetzstein, "Convolutional sparse coding for RGB NIR imaging," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1611–1625, Apr. 2018.

[14] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vision*, Santiago, Chile, 2015, pp. 1823–1831.

[15] M. Li *et al.*, "Video rain streak removal by multiscale convolutional sparse coding," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 6644–6653.

[16] B. Wohlberg and P. Wozniak, "PSF estimation in crowded astronomical imagery as a convolutional dictionary learning problem," *IEEE Signal Process. Lett.*, vol. 28, pp. 374–378, 2021.

[17] M. Li, X. Cao, Q. Zhao, L. Zhang, and D. Meng, "Online rain/snow removal from surveillance videos," *IEEE Trans. Image Process.*, vol. 30, pp. 2029–2044, 2021.

[18] X. Feng, C. Fang, X. Lou, and K. Hu, "Research on infrared and visible image fusion based on tetrolet transform and convolution sparse representation," *IEEE Access*, vol. 9, pp. 23498–23510, 2021.

[19] J. Wei, L. Mi, Y. Hu, J. Ling, Y. Li, and Z. Chen, "Effects of lossy compression on remote sensing image classification based on convolutional sparse coding," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, pp. 1–5, 2021.

[20] Y. Zhu, Y. Lu, Q. Gao, and D. Sun, "Infrared and visible image fusion based on convolutional sparse representation and guided filtering," *J. Electron. Imag.*, vol. 30, no. 4, pp. 1–17, 2021.

[21] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 391–398.

[22] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Boston, MA, USA, 2015, pp. 5135–5143.

[23] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.

[24] G. Peng, "Adaptive ADMM for dictionary learning in convolutional sparse representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3408–3422, Jul. 2019.

[25] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in *Proc. IEEE Int. Conf. Comput. Vision*, Venice, Italy, 2017, pp. 4290–4298.

[26] V. Papyan, Y. Romano, M. Elad, and J. Sulam, "Convolutional dictionary learning via local processing," in *Proc. IEEE Int. Conf. Comput. Vision*, Venice, Italy, 2017, pp. 5306–5314.

[27] I. Rey-Otero, J. Sulam, and M. Elad, "Variations on the convolutional sparse coding model," *IEEE Trans. Signal Process.*, vol. 68, pp. 519–528, 2020.

[28] V. Papyan, J. Sulam, and M. Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5687–5701, Nov. 2017.

[29] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Scalable online convolutional sparse coding," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4850–4859, Oct. 2018.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[31] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 366–381, Sep. 2018.

[32] [Online]. Available: https://github.com/FarshadGVeshki/Convolutional-Sparse-Coding.git

[33] C. Garcia-Cardona and B. Wohlberg, "Subproblem coupling in convolutional dictionary learning," in *Proc. IEEE Int. Conf. Image Process.*, Beijing, China, 2017, pp. 1697–1701.

[34] [Online]. Available: http://sipi.usc.edu/database/

Publication IV

F. G. Veshki, N. Ouzir, S. A. Vorobyov and E. Ollila. Multimodal Image Fusion via Coupled Feature Learning. *Signal Processing*, vol. 200, p. 108637, 2022.

Signal Processing 200 (2022) 108637

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Multimodal image fusion via coupled feature learning

Farshad G. Veshki[a], Nora Ouzir[b], Sergiy A. Vorobyov[a], Esa Ollila[a,*]

[a] Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland
[b] University of Paris-Saclay Inria, CentraleSupelec, CVN, France

ARTICLE INFO

Article history:
Received 14 June 2021
Revised 20 May 2022
Accepted 28 May 2022
Available online 14 June 2022

Keywords:
Multimodal image fusion
Coupled dictionary learning
Joint sparse representation
Multimodal medical imaging
Infrared images

ABSTRACT

This paper presents a multimodal image fusion method using a novel decomposition model based on coupled dictionary learning. The proposed method is general and can be used for a variety of imaging modalities. In particular, the images to be fused are decomposed into correlated and uncorrelated components using sparse representations with identical supports and a Pearson correlation constraint, respectively. The resulting optimization problem is solved by an alternating minimization algorithm. Contrary to other learning-based fusion methods, the proposed approach does not require any training data, and the correlated features are extracted online from the data itself. By preserving the uncorrelated components in the fused images, the proposed fusion method significantly improves on current fusion approaches in terms of maintaining the texture details and modality-specific information. The maximum-absolute-value rule is used for the fusion of correlated components only. This leads to an enhanced contrast-resolution without causing intensity attenuation or loss of important information. Experimental results show that the proposed method achieves superior performance in terms of both visual and objective evaluations compared to state-of-the-art image fusion methods.

1. Introduction

Multimodal image fusion aims at combining relevant information from images acquired with different sensors into a single image. The fused images are expected to preserve all important information in the source images without introducing distortions or artifacts. Multimodal image fusion has been used in a variety of applications, including surveillance [1–3], remote sensing [4–6] and medical imaging [7–9]. In various surveillance applications, fusion of infrared and visible images is used to aggregate the visual details of the optical images with thermal information captured in the infrared images, for example, to improve night vision [1,10]. In satellite imaging, the high spatial resolution of panchromatic images and the high spectral resolution of multispectral images are merged to generate more informative and high-quality fused images [6]. Multimodal medical image fusion combines the information captured using various medical imaging techniques. Anatomical imaging techniques, such as computed tomography (CT) and magnetic resonance (MR) imaging, provide high-resolution images of internal organs. Functional imaging techniques measure the biological activity of specific regions inside the organs. Single-photon emission computed tomography (SPECT) and positron emission tomography (PET) are typical examples of this type of techniques. Combining this variety of (often complementary) information into a single image, in addition to easing the visualization of multiple images, can potentially enable a joint analysis providing relevant and new information about the patient [7–9].

Multimodal image fusion methods mostly rely on the extraction of different types of features in the images before selecting or combining the most relevant ones. One way of achieving this is by transforming the images into a domain where the relevant features would arise naturally. A common approach employs multi-scale transformation (MST) techniques to extract features from different levels of resolution and select them using an appropriate fusion rule in a subsequent step [11–18]. The final fused image is obtained by applying the inverse MST to the combined multi-scale features. For example, a recent MST-based fusion method using a non-subsampled shearlet transform (NSST) has been proposed in [13]. The fusion rule is based on a pulse coupled neural network (PCNN), weighted local energy, and a modified Laplacian. In a different study, the MST is based on local Laplacian filtering (LLF), and the fusion rule relies on the information of interest (IOI) criterion [15].

Another approach seeks to learn the relevant features from the images. The methods that follow this approach often use sparse

* Corresponding author.
E-mail addresses: farshad.ghorbaniveshki@aalto.fi (F.G. Veshki), nora.ouzir@centralesupelec.fr (N. Ouzir), sergiy.vorobyov@aalto.fi (S.A. Vorobyov), esa.ollila@aalto.fi (E. Ollila).

https://doi.org/10.1016/j.sigpro.2022.108637

representations (SR) and dictionary learning [19–23]. Subspace learning techniques have also been used for this purpose [24,25]. In this category of methods, different strategies have been proposed to deal with the presence of drastically different components in the images. For example, the images are separated into low-resolution (base) and high-resolution (detail) components prior to the dictionary learning phase in [19,23]. In [20,21], morphological characteristics, such as cartoon and texture components, are used. Deep neural networks have also been utilized for multimodal image fusion. However, the applications of these techniques are mostly limited to generating decision/weight maps for the fusion of raw image pixels [26] or multi-scale features extracted using standard MST methods [27,28].

In general, both MST- and learning-based fusion methods assume that features with similar structural properties convey redundant information. These features are therefore deemed appropriate for fusion. However, this assumption does not hold in many cases due to the varying characteristics of imaging modalities. For example, the highest resolution level of MR and CT images usually depicts very different types of tissues, *i.e.,* soft tissues, such as fat and liquid, in MR and hard structures, such as bones and implants, in CT [8]. In the context of surveillance, the details in infrared and visible images display fundamentally different information. Applying a fusion rule (e.g., binary selection) to features representing distinct objects or characteristics can cause a loss of useful information. Using averaging techniques can mitigate this loss, but it leads to an attenuation of the original intensities (particularly when there is a weak signal in one of the input images). In general, it is not always meaningful to apply a fusion rule when there is no guarantee that features with the same levels of resolution or morphological characteristics encompass the same type of information. Conversely, different imaging techniques can provide varying resolutions for the same underlying structures.

This paper presents an image fusion method using a novel decomposition model base on an SR multi-component approach. Unlike other multi-component fusion methods (*e.g.*, [11,13,20,28]), we do not make assumptions on the characteristics of the correlated features nor rely on deterministic decomposition models. Instead, a general data-driven model enables us to preserve the essential features in both input modalities and reduce the loss of information. Specifically, coupled dictionary learning (CDL) [29] is used to learn features from input images and simultaneously decompose them into correlated and uncorrelated components. The core idea of the proposed model consists of two aspects:

1. The uncorrelated components contain modality-specific information that should appear in the final fused image. Therefore, these components should be preserved entirely rather than subjected to some fusion rule.
2. Since the input images still represent the same scene, organ, *etc*., they can contain a significant amount of similar or overlapping information. This redundant information is taken into account in the correlated components and considered relevant for fusion.

The coupled dictionaries play a key role here; each pair of corresponding atoms (columns) in the dictionaries, represent a correlated feature. This allows us to choose the best candidate for fusion based on the most significant dictionary atom without any loss of information. A summary of the proposed methodology and contributions is provided in the following.

- We employ a general learning-based decomposition model suitable for fusing images from various imaging modalities.
- A CDL method based on simultaneous sparse approximation is proposed for estimating the correlated features. In order to incorporate variability in the appearance of correlated features, we relax the assumption of equal SRs by enforcing common supports only.
- The uncorrelated components are estimated using a Pearson correlation-based constraint (enforcing low correlation).
- An alternating optimization method is designed for simultaneous dictionary learning and image decomposition.
- The final fusion step combines direct summation with the max-absolute-value rule.
- A MATLAB implementation of the proposed fusion method is available online at [30].

A thorough experimental comparison with current multimodal image fusion methods is conducted using 80 pairs of real multimodal images. The data comprises four different combinations of medical imaging techniques, including MR-CT, MR-PET and MR-SPECT, as well as infrared and visible images. The experimental results show that the proposed method results in a better fusion of local intensity and texture information, compared to other current techniques. In particular, isolating modality-specific information reduces the loss of information significantly.

Throughout the paper, we use bold capital letters for matrices. In a matrix, the entry at the intersection of the ith row and jth column is denoted as $[\cdot]_{(i,j)}$. A single subscript $[\cdot]_i$ is used to denote a column of a matrix. For example, $[\boldsymbol{D}_1]_i$ is the ith column of the matrix $\boldsymbol{D}_1$. The Frobenius norm of a matrix is denoted by $\|\cdot\|_F$, the Euclidean norm of a vector is denoted by $\|\cdot\|_2$, and $\|\cdot\|_0$ is the operator counting the number of nonzero coefficients of a vector. The supp$\{\cdot\}$ denotes the support of a matrix. Operator $|\cdot|$ denotes the absolute value of a number. The symbol $(\cdot)^T$ denotes the transpose operation and $(\cdot)^+$ stands for the updated variable. The symbol $\perp\!\!\!\perp$ denotes the conditional independence between two variables.

The remainder of the paper is organized as follows. Section 2 presents the proposed CDL approach and SR with common supports. The proposed image decomposition method and the fusion step are explained in Sections 3 and 4, respectively. Section 5 reports the experimental results using various examples of multi-modal images. Finally, conclusions are provided in Section 6.

2. Coupled feature learning

In this section, we modify the standard CDL problem to learn the correlated features in multimodal image pairs. The standard CDL provides a pair of dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, used to jointly represent two datasets $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The underlying relationship between these datasets is captured using a common sparse coding matrix $\boldsymbol{A}$. A standard formulation of the CDL problem is given by the following minimization problem

$$\begin{aligned} &\min_{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{A}} \|\boldsymbol{D}_1\boldsymbol{A}-\boldsymbol{X}_1\|_{\mathrm{F}}^2+\|\boldsymbol{D}_2\boldsymbol{A}-\boldsymbol{X}_2\|_{\mathrm{F}}^2 \\ &\text{s.t.}\|\boldsymbol{A}_i\|_0\le T,\ \|[\boldsymbol{D}_1]_t\|_2=1,\ \|[\boldsymbol{D}_2]_t\|_2=1,\ \forall t,i \end{aligned} \tag{1}$$

where T is the maximum number of non-zero coefficients in each column of the sparse coding matrix $\boldsymbol{A}$. The constraint on the norm of the dictionary elements is used to avoid scaling ambiguities. The standard CDL is particularly suitable for tackling problems that involve image reconstruction in different feature spaces. Problem (1) has been successfully employed in numerous image processing applications, such as image super-resolution [31], single-modal image fusion [32], or photo and sketch mapping [33].

The proposed modified CDL algorithm also captures the correlated features in associated atoms of the learned dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. However, since we are dealing with images acquired from different sensors, the same underlying structure can be displayed with different levels of visibility in each modality. For example, both CT and MRI can show tendons, but they are more visible in

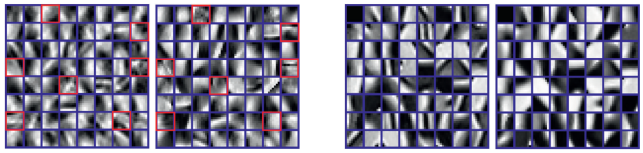


Fig. 1. Example of coupled dictionaries learned from MR and CT images using (a) the standard CDL approach of [29] and (b) the proposed SCDL method. The red frames indicate weakly correlated atoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MR images [34]. We incorporate this in the modified CDL by imposing identical supports instead of enforcing an equal sparse representation for each pair of dictionary atoms. In this way, the correlated elements can be represented with different levels of significance in each dictionary. The proposed modified CDL problem can be formulated as follows

$$\begin{aligned} \underset{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{A}_1,\boldsymbol{A}_2}{\text{minimize}} & \|\boldsymbol{D}_1\boldsymbol{A}_1-\boldsymbol{X}_1\|_{\mathrm{F}}^2+\|\boldsymbol{D}_2\boldsymbol{A}_2-\boldsymbol{X}_2\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \operatorname{supp}\{\boldsymbol{A}_1\}=\operatorname{supp}\{\boldsymbol{A}_2\} \\ & \|[\boldsymbol{A}_1]_i\|_0\le T,\ \|[\boldsymbol{A}_2]_i\|_0\le T,\ \forall i \\ & \|[\boldsymbol{D}_1]_t\|_2=1,\ \|[\boldsymbol{D}_2]_t\|_2=1,\ \forall t. \end{aligned} \tag{2}$$

Problem (1) is typically solved by alternating between a sparse coding stage and a dictionary update step [35–37]. In this work, we solve problem (2) using an alternating approach too. Specifically, after splitting the variables into two subsets $\{\boldsymbol{A}_1,\boldsymbol{A}_2\}$ and $\{\boldsymbol{D}_1,\boldsymbol{D}_2\}$, we alternate between two optimization phases detailed in the following subsections.

2.1. Sparse coding

Minimizing (2) with respect to the first set of variables $\{\boldsymbol{A}_1,\boldsymbol{A}_2\}$ requires some changes to the standard sparse coding procedure (*e.g.*, OMP [38]). First, one has to enforce common supports. The method of simultaneous orthogonal matching pursuit (SOMP) [39] is of interest here as it includes a common support constraint. Secondly, one has to consider coupled dictionaries. This can be achieved by modifying the atom selection rule of SOMP so that each of the input signals is approximated using a different dictionary instead of sharing a single one. In each iteration, this modified algorithm selects a pair of coupled atoms $\{[\boldsymbol{D}_1]_s,[\boldsymbol{D}_2]_s\}$ minimizing the sum of the squared residuals. This can be formulated as

$$s=\underset{t}{\operatorname{argmin}}\ \left\|\alpha_1^t[\boldsymbol{D}_1]_t-\boldsymbol{r}_1\right\|_2^2+\left\|\alpha_2^t[\boldsymbol{D}_2]_t-\boldsymbol{r}_2\right\|_2^2, \tag{3}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the approximation residuals of a pair of input signals (*e.g.*, $\boldsymbol{r}_1=[\boldsymbol{X}_1]_i-\boldsymbol{D}_1[\boldsymbol{A}_1]_i$ and $\boldsymbol{r}_2=[\boldsymbol{X}_2]_i-\boldsymbol{D}_2[\boldsymbol{A}_2]_i$ are the residuals corresponding to the ith columns of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively). Moreover, $\alpha_1^t=[\boldsymbol{A}_1]_{(t,i)}$ and $\alpha_2^t=[\boldsymbol{A}_2]_{(t,i)}$ are the sparse coefficients corresponding to $[\boldsymbol{D}_1]_t$ and $[\boldsymbol{D}_2]_t$, respectively. Problem (3) can be efficiently addressed by solving its equivalent maximization problem, that is

$$s=\underset{t}{\operatorname{argmax}}\ \left(\boldsymbol{r}_1^T[\boldsymbol{D}_1]_t\right)^2+\left(\boldsymbol{r}_2^T[\boldsymbol{D}_2]_t\right)^2.$$

The optimal nonzero coefficients of the sparse codes are then computed based on their associated selected atoms. As opposed to SOMP, we stop the algorithm when only one of the input signals meets the stopping criterion (*i.e.,* the Euclidean norm of one of the residuals is smaller than a user-defined threshold ϵ). This is based on the fact that the main objective of the algorithm is to estimate the correlated features, whereas any remaining noise (once the approximation of one of the signals is complete) is evidently uncorrelated with the residual of the second signal. The modified SOMP algorithm can be straightforwardly extended to multiple inputs (see Section 4.4 for details). The steps of the proposed SOMP algorithm are explained in Appendix A.

2.2. Dictionary update

The first two terms of (2) are independent with respect to the dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. They are also independent with respect to the non-zero coefficients in $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ when the supports are fixed. Therefore, the dictionaries $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be updated individually using any efficient dictionary learning algorithm. We choose to use the K-SVD method [35] because it updates the dictionary atoms and the associated sparse coefficients without changing their supports. Specifically, K-SVD uses a singular value decomposition.

The proposed CDL method alternates between the sparse coding step using the modified SOMP and the dictionary update using K-SVD. We will refer to the proposed approach as simultaneous coupled dictionary learning (SCDL). Algorithm 1 summarizes

Algorithm 1 Main steps of the SCDL method.

Inputs: multimodal data $\mathbb{X}_1$ and $\mathbb{X}_2$, and initial dictionaries $\mathbb{D}_1$ and $\mathbb{D}_2$.**Repeat** until stopping criteria are fulfilled:

1. ***Simultaneous Sparse Coding***: solve problem (2) with respect to $\mathbb{A}_1$ and $\mathbb{A}_2$ using the modified SOMP method (see Section 2.1).
2. ***Dictionary Update***: solve problem (2) with respect to $\mathbb{D}_1$ and $\mathbb{D}_2$ and the nonzero coefficients in $\mathbb{A}_1$ and $\mathbb{A}_2$ using the K-SVD method [35].

Outputs: learned coupled dictionaries $\mathbb{D}_1$ and $\mathbb{D}_2$, and sparse representations with identical supports $\mathbb{A}_1$ and $\mathbb{A}_2$.

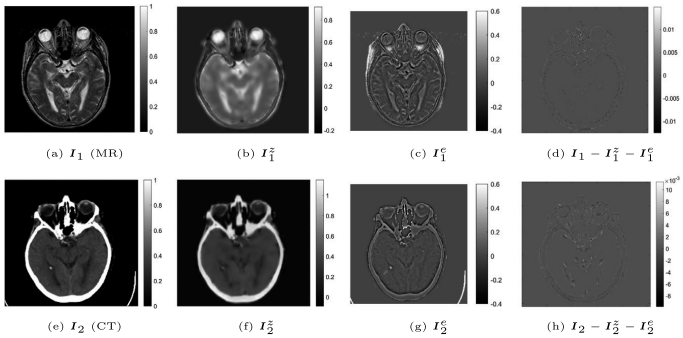
the main steps of the SCDL method. Fig. 1 shows how the proposed SCDL method captures correlated features more efficiently than the standard CDL method of [29]. Specifically, one can see how the dictionaries obtained using the standard method (Fig.) contain relatively uncorrelated atoms (framed in red), while all the atoms obtained with SCDL present a clear correlation (see Fig. 1).

3. Image decomposition via SCDL

Our fusion method decomposes the input images into correlated and uncorrelated components. The latter represent features that are unique to each modality and which we seek to preserve. This section introduces the proposed decomposition model and explains how to solve the resulting minimization problem. Throughout this section, we use examples of multimodal medical images to illustrate the properties of the proposed model.

3.1. Decomposition model

Let the input images to be fused be denoted by $\boldsymbol{I}_1\in\mathbb{R}^{M\times N}$ and $\boldsymbol{I}_2\in\mathbb{R}^{M\times N}$.[1] The correlated components are denoted by $\boldsymbol{I}_1^z\in\mathbb{R}^{M\times N}$



(a) $\boldsymbol{I}_1$ (MR) (b) $\boldsymbol{I}_1^z$ (c) $\boldsymbol{I}_1^e$ (d) $\boldsymbol{I}_1 - \boldsymbol{I}_1^z - \boldsymbol{I}_1^e$

(e) $\boldsymbol{I}_2$ (CT) (f) $\boldsymbol{I}_2^z$ (g) $\boldsymbol{I}_2^e$ (h) $\boldsymbol{I}_2 - \boldsymbol{I}_2^z - \boldsymbol{I}_2^e$

Fig. 2. A pair of MR and CT images (a,e) decomposed into their correlated (b,f) and uncorrelated (c,g) components based on the proposed model. The residuals are shown in (d,h).

and $\boldsymbol{I}_2^z \in \mathbb{R}^{M\times N}$, and the uncorrelated components by $\boldsymbol{I}_1^e \in \mathbb{R}^{M\times N}$ and $\boldsymbol{I}_2^e \in \mathbb{R}^{M\times N}$. The proposed decomposition model can be expressed as follows

$$\begin{cases} \boldsymbol{I}_1 = \boldsymbol{I}_1^z + \boldsymbol{I}_1^e \\ \boldsymbol{I}_2 = \boldsymbol{I}_2^z + \boldsymbol{I}_2^e \end{cases} \text{where } \boldsymbol{I}_1^e \perp\!\!\!\perp \boldsymbol{I}_2^e. \tag{4}$$

The SCDL method explained in Section 2 operates patch-wise. Therefore, we rewrite (4) using the matrices containing all the extracted patches as follows

$$\begin{cases} \boldsymbol{X}_1 = \boldsymbol{Z}_1 + \boldsymbol{E}_1 \\ \boldsymbol{X}_2 = \boldsymbol{Z}_2 + \boldsymbol{E}_2 \end{cases} \text{where } \boldsymbol{E}_1 \perp\!\!\!\perp \boldsymbol{E}_2, \tag{5}$$

where the matrices $\boldsymbol{X}_1 \in \mathbb{R}^{m\times p}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{m\times p}$ contain the p vectorized overlapping patches of size m extracted from the input images. The patches of the correlated components are represented by $\boldsymbol{Z}_1 \in \mathbb{R}^{m\times p}$ and $\boldsymbol{Z}_2 \in \mathbb{R}^{m\times p}$, and those of the uncorrelated components by $\boldsymbol{E}_1 \in \mathbb{R}^{m\times p}$ and $\boldsymbol{E}_2 \in \mathbb{R}^{m\times p}$.

Fig. 2 shows how the proposed decomposition model captures correlated and uncorrelated features in a pair of MR-CT images. The uncorrelated components contain edges and details that can be clearly observed in only one of the modalities, *e.g.,* sulci details in the MR image and calcification in the CT image (indicated by red arrows). The correlated components represent the underlying joint structure that can be referred to as the base or background layer. Fig. 3, shows another example of decomposition for a pair of PET and MR images. In functional-anatomical imaging, any details in the images are naturally uncorrelated. The uncorrelated components also capture any non-overlapping regions. Finally, the correlated components contain the regions where the background of the anatomical image overlaps with the biological activity information.

3.2. Minimization problem

In order to estimate the correlated and uncorrelated components in (5), we formulate a minimization problem based on the SCDL approach described in the previous section. Specifically, we seek the coupled sparse representation of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ using sparse codes with identical supports $\boldsymbol{A}_1 \in \mathbb{R}^{n\times p}$ and $\boldsymbol{A}_2 \in \mathbb{R}^{n\times p}$ and coupled dictionaries $\boldsymbol{D}_1 \in \mathbb{R}^{m\times n}$ and $\boldsymbol{D}_2 \in \mathbb{R}^{m\times n}$ (*i.e.,* $\boldsymbol{Z}_1 = \boldsymbol{D}_1\boldsymbol{A}_1$ and $\boldsymbol{Z}_2 = \boldsymbol{D}_2\boldsymbol{A}_2$). The element-wise independence of $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ is enforced by minimizing the squared Pearson correlation coefficients, where the local means μ and standard deviations σ are estimated patch-wise. The corresponding cost function is expressed as follows

$$\boldsymbol{\phi}\big([\boldsymbol{E}_1]_{(i,j)}, [\boldsymbol{E}_2]_{(i,j)}\big) = \left(\frac{([\boldsymbol{E}_1]_{(i,j)} - \mu_{1,j})([\boldsymbol{E}_2]_{(i,j)} - \mu_{2,j})}{\sigma_{1,j}\sigma_{2,j}}\right)^2$$

where j is the patch index and the subscripts 1 and 2 indicate the associated components $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$. Note that this patch-wise approach results implicitly in multiple counts of the same pixel when considering overlapping patches [43]. Combining the independence term with the proposed SCDL approach leads to the following minimization problem

$$\begin{aligned} \underset{\substack{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{A}_1,\boldsymbol{A}_2,\\ \boldsymbol{E}_1,\boldsymbol{E}_2}}{\text{minimize}} \quad & \textstyle\sum_{\substack{i=1,\dots,m\\ j=1,\dots,p}} \boldsymbol{\phi}\big([\boldsymbol{E}_1]_{(i,j)}, [\boldsymbol{E}_2]_{(i,j)}\big) \\ \text{s.t.} \quad & \boldsymbol{D}_k\boldsymbol{A}_k + \boldsymbol{E}_k = \boldsymbol{X}_k, k = 1, 2 \\ & \text{supp}\{\boldsymbol{A}_1\} = \text{supp}\{\boldsymbol{A}_2\} \\ & \|[\boldsymbol{A}_1]_i\|_0 \le T, \|[\boldsymbol{A}_2]_i\|_0 \le T, \forall i \\ & \|[\boldsymbol{D}_1]_t\|_2 = 1, \|[\boldsymbol{D}_2]_t\|_2 = 1, \forall t, \end{aligned} \tag{6}$$

where the sparsity and common support constraints introduced in (2).

3.3. Optimization

Problem (6) is challenging because of its non-convexity and the presence of multiple sets of variables. Therefore, we propose to break the optimization procedure into simpler subproblems where we consider minimization with respect to separate blocks of variables. We then alternate between these subproblems until finding a local optimum. Specifically, the minimization with respect to the sparse codes and dictionaries, and the minimization with respect to the uncorrelated components are treated separately. Furthermore, we simplify the problem by approximating the first two equality constraints in (6) by quadratic approximation terms. This leads to a new optimization problem that can be written as

$$\begin{aligned} \underset{\substack{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{A}_1,\boldsymbol{A}_2,\\ E_1,\boldsymbol{E}_2}}{\text{minimize}} \quad & \textstyle\sum_{\substack{i=1,\cdots,m\\ j=1,\cdots,p}} \boldsymbol{\phi}\big([\boldsymbol{E}_1]_{(i,j)}, [\boldsymbol{E}_2]_{(i,j)}\big) \\ & +\rho \textstyle\sum_{k=1,2} \|\boldsymbol{D}_k\boldsymbol{A}_k + \boldsymbol{E}_k - \boldsymbol{X}_k\|_F^2 \\ \text{s.t.} \quad & \text{supp}\{\boldsymbol{A}_1\} = \text{supp}\{\boldsymbol{A}_2\} \\ & \|[\boldsymbol{A}_1]_i\|_0 \le T, \|[\boldsymbol{A}_2]_i\|_0 \le T, \forall i \\ & \|[\boldsymbol{D}_1]_t\|_2 = 1, \|[\boldsymbol{D}_2]_t\|_2 = 1, \forall t \end{aligned} \tag{7}$$

where $\rho > 0$ controls the trade-off between the independence of $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ and the accuracy of the sparse representations. As mentioned above, problem (7) is tackled by alternating minimizations

[1] A typical assumption in multimodal image fusion is that the input images are accurately coregistered beforehand [40]. Multimodal image registration is considered a separate problem and effectively addressed using various available methods [41,42].

with respect to the two blocks of variables $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{D}_1, \boldsymbol{D}_2\}$ and $\{\boldsymbol{E}_1, \boldsymbol{E}_2\}$. Each resulting subproblem is described in more detail in the following.

3.3.1. Optimization with respect to $\{A_1, A_2, D_1, D_2\}$

The first optimization subproblem can be written as

$$\begin{aligned}\underset{\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{A}_1,\boldsymbol{A}_2}{\text{minimize}} & \|\boldsymbol{D}_1\boldsymbol{A}_1 - \boldsymbol{X}_1'\|_F^2 + \|\boldsymbol{D}_2\boldsymbol{A}_2 - \boldsymbol{X}_2'\|_F^2 \\ \text{s.t.} \quad & \text{supp}\{\boldsymbol{A}_1\} = \text{supp}\{\boldsymbol{A}_2\} \\ & \|[\boldsymbol{A}_1]_i\|_0 \le T, \|[\boldsymbol{A}_2]_i\|_0 \le T, \forall i \\ & \|[\boldsymbol{D}_1]_t\|_2 = 1, \|[\boldsymbol{D}_2]_t\|_2 = 1, \forall t\end{aligned}$$

where $\boldsymbol{X}_1' = \boldsymbol{X}_1 - \boldsymbol{E}_1$ and $\boldsymbol{X}_2' = \boldsymbol{X}_2 - \boldsymbol{E}_2$. This subproblem is addressed using the SCDL method explained in Section 2. The dictionaries can be initialized in the first iteration of the algorithm using a predefined dictionary, *e.g.,* based on discrete cosine transforms (DCT). In subsequent iterations, the SCDL method is initialized using the dictionaries obtained from the previous one. Furthermore, initializing the sparsity parameter T with a small value and gradually increasing it at each iteration ensures a warm start of the algorithm.

3.3.2. Optimization with respect to $\{E_1, E_2\}$

The second optimization subproblem can be written as

$$\begin{aligned}\underset{\boldsymbol{E}_1,\boldsymbol{E}_2}{\text{minimize}} & \sum_{\substack{i=1,\cdots,m \\ j=1,\cdots,p}} \boldsymbol{\phi}\big([\boldsymbol{E}_1]_{(i,j)}, [\boldsymbol{E}_2]_{(i,j)}\big) \\ & + \rho \sum_{k=1,2} \|\boldsymbol{D}_k\boldsymbol{A}_k + \boldsymbol{E}_k - \boldsymbol{X}_k\|_F^2.\end{aligned} \tag{8}$$

The estimates of $\{\mathbf{E_1}, \mathbf{E_2}\}$ are dependent on unobserved latent variables, namely the patch-wise means and standard deviations. Therefore, we propose to address this subproblem using an expectation-maximization (EM) method [44, Chap. 5.3]. The EM approach leads to the following updates for the uncorrelated component

$$\begin{aligned}[\boldsymbol{E}_1]_{(i,j)}^{+} &= \frac{\rho[\boldsymbol{X}_1 - \boldsymbol{D}_1\boldsymbol{A}_1]_{(i,j)} + \frac{([\boldsymbol{E}_2]_{(i,j)} - \mu_{2,j})^2}{\sigma_{1,j}^2\sigma_{2,j}^2}\mu_{1,j}}{\rho + \frac{([\boldsymbol{E}_2]_{(i,j)} - \mu_{2,j})^2}{\sigma_{1,j}^2\sigma_{2,j}^2}} \\ [\boldsymbol{E}_2]_{(i,j)}^{+} &= \frac{\rho[\boldsymbol{X}_2 - \boldsymbol{D}_2\boldsymbol{A}_2]_{(i,j)} + \frac{([\boldsymbol{E}_1]_{(i,j)} - \mu_{1,j})^2}{\sigma_{1,j}^2\sigma_{2,j}^2}\mu_{2,j}}{\rho + \frac{([\boldsymbol{E}_1]_{(i,j)} - \mu_{1,j})^2}{\sigma_{1,j}^2\sigma_{2,j}^2}}\end{aligned} \tag{9}$$

where the means and standard deviations are computed using the current values of $\{\boldsymbol{E_1}, \boldsymbol{E_2}\}$. Note that $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are initialized with $\boldsymbol{X}_1 - \boldsymbol{D}_1\boldsymbol{A}_1$ and $\boldsymbol{X}_2 - \boldsymbol{D}_2\boldsymbol{A}_2$, respectively. The updates are performed only for entries in columns with $\sigma_{1,j}^2\sigma_{2,j}^2 > \delta$, where $\delta > 0$ is a small constant used to avoid division by zero.

3.4. Computational complexity

The computational cost of the proposed decomposition algorithm is dominated by the first subproblem (solved using SCDL). In particular, the complexity of one SCDL iteration (including all substeps) is $\mathcal{O}(p\max(Tnm, T^2m, T^3, nm^2))$, while the complexity of the EM step is only $\mathcal{O}(mp)$. Based on the experimental findings of [45], sparse approximations are performed in this work in single precision, which improves the computational efficiency of the algorithm.

4. Multimodal fusion rule

Once the correlated and uncorrelated components are estimated, the final fused image is obtained using an appropriate fusion rule. In this work, the different components are handled separately. Specifically, the correlated components are fused because they contain redundant information, for instance, shared underlying structures in anatomical imaging or background elements in the functional-anatomical case. In contrast, the uncorrelated components are entirely preserved in the final image to avoid loss of modality-specific information (*e.g.,* the calcification inside the CT image of Fig. 2(g)).

4.1. Fusion of correlated components

According to the justifications provided above, the correlated components are fused using a binary selection where the most relevant features are chosen based on the magnitudes of the sparse coefficients. Recall that the proposed SCDL method with common supports allows correlated features to be captured with varying significance levels for each modality. This is a novel approach compared to the standard max-absolute-value rule with a single predefined basis [21,46]. Precisely, the most significant features are selected based on the sparse coefficients with the largest magnitudes as follows

$$[\boldsymbol{A}_1']_{(i,j)} = \begin{cases}[\boldsymbol{A}_1]_{(i,j)}, & \text{if} |[\boldsymbol{A}_1]_{(i,j)}| \ge |[\boldsymbol{A}_2]_{(i,j)}| \\ 0, & \text{otherwise}\end{cases},$$

$$[\boldsymbol{A}_2']_{(i,j)} = \begin{cases}[\boldsymbol{A}_2]_{(i,j)}, & \text{if} |[\boldsymbol{A}_2]_{(i,j)}| > |[\boldsymbol{A}_1]_{(i,j)}| \\ 0, & \text{otherwise}\end{cases}.$$

Then, the fused correlated component, denoted by $\mathbb{Z}^F$, is reconstructed using the selected coefficients as $\boldsymbol{Z}^F = \boldsymbol{D}_1\boldsymbol{A}_1' + \boldsymbol{D}_2\boldsymbol{A}_2'$.

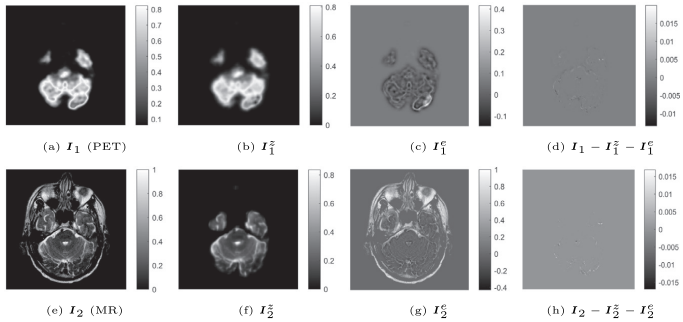
4.2. Reconstruction of final fused image

Since the uncorrelated components contain details or non-overlapping regions that should be preserved in the final image, they can be added directly to the fused correlated components, *i.e.,* $\boldsymbol{X}^F = \boldsymbol{Z}^F + \boldsymbol{E}_1 + \boldsymbol{E}_2$.

Finally, the fused image $\boldsymbol{I}^F$ is reconstructed by placing the patches of $\boldsymbol{X}^F$ at their original positions in the image and averaging the overlapping ones. The decomposition residuals (see examples in Fig. 2a–d) are negligible and are not included in the final fused image. The block diagram of the proposed method is presented in Fig. 4. Fig. 5 shows the fused image obtained with the proposed method for the MR-CT images in Fig. 2 and the MR-PET images in Fig. 3. The fused MR-CT image contains the modality-specific information captured by the uncorrelated components (*e.g.,* calcification and sulci details), as well as the most visible features selected from the correlated components. The fused MR-PET image combines the background of the MR image (Fig. 3a) with the functional information from the PET image at the overlapping regions. The details and non-overlapping regions, captured in the uncorrelated components appear unaltered in the fused image.

4.3. Color images

Multimodal image fusion can involve fusion of a color image with a greyscale one. For example, functional medical images (*e.g.,* PET) are usually displayed in a color code, as opposed to the grey-scale anatomical medical images. One common approach for dealing with the fusion of color images is to convert them from the original RGB format to the YCbCr (or YUV) color-space [13,27]. In this new color-space, component Y (*i.e.,* luminance) provides the grey-scale version of the image, which is used here for fusion. Since the full color information comes from the functional images, the remaining color components (*i.e.*, Cb and Cr) are transmitted directly to the final (grey-scale) fused image. Fig. 6 shows the block diagram of the greyscale and color image fusion method.



(a) I_1 (PET) (b) I_1^z (c) I_1^e (d) $I_1 - I_1^z - I_1^e$

(e) I_2 (MR) (f) I_2^z (g) I_2^e (h) $I_2 - I_2^z - I_2^e$

Fig. 3. A pair of PET and MR images (a,e) decomposed into their correlated (b,f) and uncorrelated (c,g) components using the proposed model. The residuals are shown in (d,h). (The grey-scale component of the PET image is used in the decomposition, as explained in Section 4.3.).

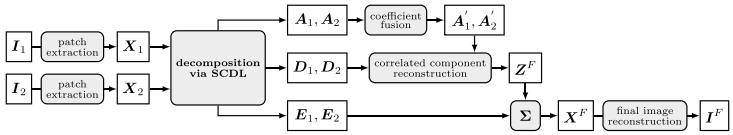
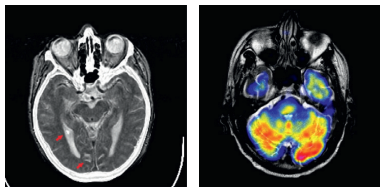


Fig. 4. Block diagram of the proposed fusion method.



(a) (b)

Fig. 5. Final fused images obtained with the proposed method for (a) the MR-CT example in Fig. 2, and (b) MR-PET example in Fig. 3. Red arrows indicate the calcification and sulci details preserved by the proposed method.

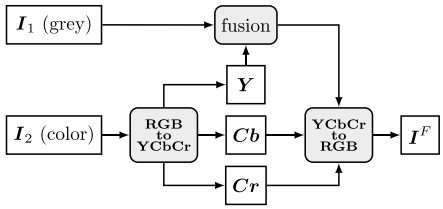


Fig. 6. Block-diagram of the grey-scale and color image fusion method.

4.4. Extension to multiple input images

Generalizing the proposed method to handle more than two input images is straightforward. Since the sparse approximation (the modified SOMP method) can be applied to any number of inputs (see Appendix A), and the dictionary update is performed over the coupled dictionaries separately, the SCDL algorithm can be directly applied to more than two images. To ensure that all correlated features between images are captured, the SOMP iterations can be stopped when all inputs except one meet the stopping criteria.

In the correlation minimization phase, one needs to minimize the correlation between each uncorrelated component and all other uncorrelated components. For K inputs, the correlation term for the l-th input can be written as

$$\sum_{\substack{k=1,\cdots,K \\ k\neq l}} \boldsymbol{\phi}\big([\boldsymbol{E}_l]_{(i,j)}, [\boldsymbol{E}_k]_{(i,j)}\big) = \sum_{\substack{k=1,\cdots,K \\ k\neq l}} \times \left(\frac{\big([\boldsymbol{E}_l]_{(i,j)}-\mu_{l,j}\big)\big([\boldsymbol{E}_k]_{(i,j)}-\mu_{k,j}\big)}{\sigma_{l,j}\sigma_{k,j}}\right)^2.$$

To optimize the objective function with respect to $\boldsymbol{E}_l$, the following problem is solved

$$\underset{\boldsymbol{E}_l}{\text{minimize}}\ \rho\|\boldsymbol{D}_l\boldsymbol{A}_l+\boldsymbol{E}_l-\boldsymbol{X}_l\|_F^2+\sum_{\substack{k=1,\cdots,K \\ k\neq l}}\sum_{\substack{i=1,\cdots,m \\ j=1,\cdots,p}} \boldsymbol{\phi}\big([\boldsymbol{E}_l]_{(i,j)}, [\boldsymbol{E}_k]_{(i,j)}\big)$$

which leads to the iteration

$$[\boldsymbol{E}_l]^+_{(i,j)} = \frac{\rho[\boldsymbol{X}_l-\boldsymbol{D}_l\boldsymbol{A}_l]_{(i,j)}+\sum_{\substack{k=1,\ldots,K \\ k\neq l}}\left(\frac{[\boldsymbol{E}_k]_{(i,j)}-\mu_{k,j}}{\sigma_{l,j}\sigma_{k,j}}\right)^2\mu_{l,j}}{\rho+\sum_{\substack{k=1,\ldots,K \\ k\neq l}}\left(\frac{[\boldsymbol{E}_k]_{(i,j)}-\mu_{k,j}}{\sigma_{l,j}\sigma_{k,j}}\right)^2}.$$

In the fusion step, similar to the case with two inputs, the coupled features with the largest sparse coefficients are selected for the fused image, and the uncorrelated components are added to the fused image directly.

5. Experiments

In this section, we evaluate the proposed method in the context of two major applications of multimodal image fusion. First, we use medical imaging data from various modalities and compare our method to state-of-the-art medical image fusion methods. Then, we conduct experiments in the context of infrared-visible images and compare with several recent methods. The evaluation is based on objective metrics, visual quality, and computational efficiency. We also discuss the parameter tuning strategy.

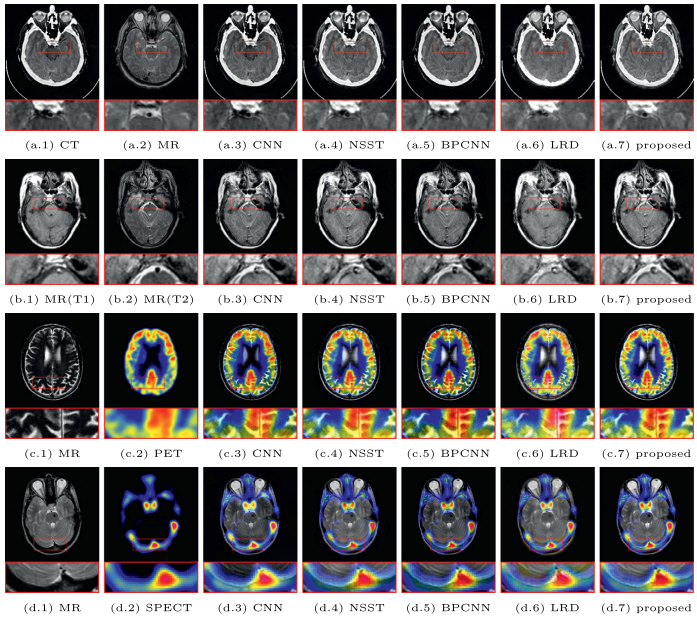


Fig. 7. Multimodal medical image fusion results.

5.1. Experiment setup

5.1.1. Datasets

We use 60 pairs of multimodal medical images from The Whole Brain Atlas database [47] and 20 infrared-visible images collected from [48]. The medical fusion data includes 20 anatomical-anatomical images (10 MR(T1)–MR(T2) images and 10 MR–CT images). The functional-anatomical fusion dataset comprises 20 MR–PET images and 20 MR–SPECT images. All of the images are accurately registered. The medical images are of size 256×256 pixels.

5.1.2. Methods used for comparison

For the experiments on multimodal medical images, the proposed fusion method is compared to seven recent medical image fusion methods, including (1) NSST [13] and (2) LLF [15] (introduced in Section 1); (3) CNN [27], which relies on convolutional neural networks and Laplacian pyramids; (4) a method using convolutional sparse coding referred to as CSR [19]; (5) a method using union Laplacian pyramids referred to as ULAP [14]; (6) a method based on boundary measured PCNN and energy attribute in NSST domain (BPCNN) [49]; a method based on Laplacian redecomposition (LRD) [18]. For the infrared-visible image fusion task, four recent methods are used for comparison: (1) a method based on deep learning that uses deep residual network and relies on zero-phase component analysis, referred to as ResNet [26]; (2) a method based on DCT in a discrete stationary wavelet transform domain, referred to as SWT [11]; (3) a method that incorporates a hierarchical Bayesian model (Bayes) [10]; (4) the CSR method [19]. All methods considered for comparison are implemented using MATLAB. All experiments are performed on a PC running an Intel(R) Core(TM) i5-8365U 1.60GHz CPU. Note that LLF is an anatomical-functional image fusion method. Therefore, it is tested in our experiments for this type of data only. To ensure a fair comparison in the case of anatomical-functional images, the grey-scale CSR method is adopted to color images using the approach explained in Section 4.3.

5.1.3. Parameter setting

For the proposed method, we use $m = 64$ (fully overlapping patches of size 8×8), $\epsilon = \sqrt{m} \times 10^{-3}$, and $\delta = 10^{-6}$ in all the experiments. In addition, we use $n = 100$ (number of atoms), $T = 6$ (sparsity level) and $\rho = 20$ for anatomical-anatomical (MR-MR and MR-CT) image fusion, $n = 16$, $T = 3$ and $\rho = 10$ for functional-anatomical (MR-PET and MR-SPECT) image fusion, and $n = 16$, $T = 3$ and $\rho = 5$ for infrared-visible image fusion. The details of the parameter selection strategy are discussed in Section 5.4. For all other methods, we use the best parameters as tuned by the authors.

5.1.4. Objective metrics

The quantitative comparison of the methods is performed based on the following metrics: the tone-mapped image quality index *TMQI* [50], which measures preservation of intensity and structural information, the similarity-based fusion quality metric Q_Y [51], the human visual system-based metric Q_{CB} [52], the feature mutual information metric *FMI* [53], the visual information fidelity metric *VIF* [54], the objective image fusion performance measure $Q_{AB/F}$ [55], the spatial frequency index *SF* [56], which measures the overall activity level in the image, the edge intensity metric *EI* [57], the structural similarity index *SSIM* [58], and standard deviation (STD). Note that high *VIF*, *TMQI* and *SSIM* values correspond to high fidelity in terms of intensity and structural features, high *FMI*, $Q_{AB/F}$ and Q_Y values indicate high structural similarities, high Q_{CB} values indicate good visual compatibility, high *EIs* indicate high quality edges, and high *SF* and *STD* values imply an improved contrast.

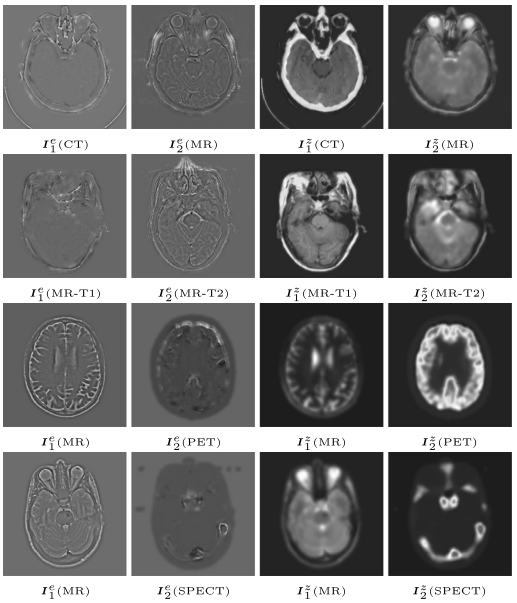


Fig. 8. Decomposition components obtained using the proposed method for multimodal medical images in Fig. 7: (first row) CT-MR images, (second row) MR(T1)-MR(T2) images, (third row) MR-PET images, and (forth row) MR-SPECT images. For better visualization, the components are converted to the standard image range.

5.2. Multimodal medical image fusion

5.2.1. Visual comparison

Fig. 7 shows the final fused image for one pair of images in each of the medical image fusion experiments. The decomposition components obtained using the proposed method for the images in Fig. 7 are shown in Fig. 8. In Fig. 7, one can see that the CNN and BPCNN methods clearly suffer from a loss of local intensity. We observed the same negative effect (even more severe) for the ULAP and CSR methods. This is because these four methods rely on an averaging-based approach for the fusion of low-resolution components, which usually contain most of the energy content of the images. For example, the intensities are significantly attenuated in regions where one of the input images is dark. The ULAP and CNN methods also use averaging for their high-resolution components. In this case, averaging results in a loss of details or texture, as can be seen in Fig. 7(c,3) and (d,3).

We observed that all the results obtained using the LLF method contain clear artifacts and visual inconsistencies. This MST-based method uses binary-selection-based fusion for both low and high-resolution features. As explained in Section 1, binary selection of resolution-based components can cause loss of details and essential information. The same effect is observed for the CSR method, which also relies on binary selection for its *detail* layer corresponding to the highest resolution. The LRD method usually preserves the local intensities but results in color distortions and blurred details.

The BPCNN and NSST methods both use 49 decomposition layers that is significantly more than in the CSR (2 layers) and LLF (3 layers) methods. This allows the BPCNN and NSST methods to capture more relevant information, including intensity and texture. Also, the use of directional filters in the NSST method improves the fusion of features with higher structural similarities. However, these NSST-based method occasionally suffers from a non-negligible amount of intensity attenuation, again due to employing a binary-selection rule for resolution-based features (see Fig. 7(a,4) and (b,4)). Moreover, the NSST method reconstructs the final image solely based on a sparse representation, which inevitably leads to a loss of texture information. For example, the magnified region of Fig. 7(d,4) shows how the texture of the MR image appears with a significantly lower contrast in the NSST and BPCNN results. Note that the corresponding regions in the SPECT image are entirely dark, meaning that the texture is expected to appear in the final image unaltered.

The proposed method provides good visual results by preserving both intensity and details. Recall that in the proposed method, the fusion rule is applied to the correlated features only, which guarantees that binary selection does not omit any modality-specific information. Moreover, the uncorrelated components isolate these modality-specific features and add them directly to the fused image without employing any additional processing that might lead to texture degradation or necessitate expensive computations. The advantage of this strategy is evident in functional-anatomical fusion, where large areas in one of the images can be flat or dark in the other image. For example, the regions of the MR image that are dark (flat with low intensity) in the SPECT image of Fig. 7d are well preserved by the proposed method, while all other methods show a loss of the local intensity or decreased contrast.

Fig. 8 shows how the uncorrelated components in MR-PET and MR-SPECT capture the high-resolution content of the MR images. In contrast, the correlated components of these images contain only low-resolution information. In MR-CT and MR(T1)-MR(T2), significant amount of information is captured by the correlated components, while each uncorrelated component contains the details of specific types of tissues.

5.2.2. Comparison using objective metrics

The results obtained based on the objective metrics are reported in Table 1. These results are in favor of the proposed method. Specifically, the ULAP and CSR methods provide low STD and SF values, which is due to the loss of contrast discussed previously. Moreover, the LLF method always shows relatively low Q_Y and Q_{CB} values, which points to the presence of visual artifacts. The objective metrics of the CNN and BPCNN methods are always lower than those of the proposed methods. Finally, the proposed method leads to the best overall results for all datasets. These findings show that the proposed method generalizes well to diverse medical imaging modalities.

5.2.3. Execution times

The average execution times of all the experiments are reported in the last row of Table 1. This table shows that the proposed method is competitive with recent multimodal fusion methods in terms of computational efficiency. Specifically, the running time of the proposed method is comparable to that of the NSST method and significantly better than those of the LRD and BPCNN methods. The ULAP method results in the shortest execution time but does not yield competitive results.

5.3. Fusion of infrared and visible images

Fig. 9 illustrates four examples of fused infrared-visible images obtained by different methods. The decomposition components obtained for the images in Figs. 9 using the proposed method are visualized in Fig. 10. The average objective evaluation results of all images in the dataset are summarized in Table 2. As can be seen in Fig. 9, the ResNet and Bayes methods result in a loss of intensity, as well as blurred textures and details. The low contrast resolution

Table 1
Objective evaluation results and average execution times for different methods using medical image datasets. The best performance is shown in bold.

Data-sets	Metrics	CSR	LLF	ULAP	CNN	NSST	LRD	BPCNN	proposed
MR(T1)-MR(T2)	*FMI*	0.7560	–	0.6739	0.7229	0.7412	0.7760	0.7125	**0.7899**
	VIF	0.6388	–	0.6461	0.7444	0.7603	0.7364	0.7140	**0.7708**
	$Q_{AB/F}$	0.5679	–	0.3664	**0.5741**	0.4750	0.5721	0.5086	0.4295
	Q_Y	0.8899	–	0.8359	0.7349	0.8651	0.8778	0.7730	**0.9223**
	Q_{CB}	0.7073	–	0.6769	0.5891	0.7107	0.6858	0.6273	**0.7309**
	TMQI	0.7745	–	0.7680	0.7694	0.7768	0.7802	0.7612	**0.7809**
	STD	56.3661	–	66.1464	64.5726	66.2631	68.2687	62.2850	**68.9236**
	SF	23.0481	–	23.4086	26.2503	25.3642	24.9977	25.1888	**27.2283**
	EI	61.6357	–	75.5990	72.4133	71.7468	67.9044	68.9969	**76.9263**
	SSIM	0.7569	–	0.7233	0.5845	0.7297	0.7368	0.6294	**0.7811**
MR-CT	*FMI*	0.6474	–	0.5989	0.6504	0.6592	0.7122	0.6433	**0.7343**
	VIF	0.3134	–	0.3525	0.3726	0.4162	0.4086	0.3565	**0.4463**
	$Q_{AB/F}$	0.4272	–	0.3747	0.5160	0.4255	**0.5628**	0.3681	0.4382
	Q_Y	0.8007	–	0.7613	0.7573	0.7621	0.8171	0.7092	**0.8754**
	Q_{CB}	0.6112	–	0.5991	0.5659	0.5725	0.5627	0.5527	**0.6160**
	TMQI	0.7285	–	0.7228	0.7101	0.7277	0.7470	0.6913	**0.7494**
	STD	65.0380	–	71.4757	84.2350	87.4317	**89.4686**	82.5867	87.8284
	SF	31.0923	–	28.4499	36.4562	**35.2533**	32.8678	34.8559	34.2227
	EI	70.6873	–	82.7175	82.1789	82.7084	75.3461	78.7288	**84.6628**
	SSIM	0.5848	–	0.5791	0.5217	0.5371	0.5834	0.4708	**0.6299**
MR-PET	*FMI*	0.7005	0.6281	0.6371	0.6844	0.6883	0.6588	0.7109	**0.7517**
	VIF	0.4067	0.5003	0.4534	0.5054	0.5376	0.5054	0.4836	**0.5649**
	$Q_{AB/F}$	0.6178	0.5616	0.3962	0.5637	0.6114	0.5637	0.6015	**0.6426**
	Q_Y	0.6388	0.8022	0.8196	0.7938	0.8216	0.8305	0.8914	**0.9012**
	Q_{CB}	0.6148	0.6661	0.6761	0.6346	0.6455	0.6105	0.6993	**0.7032**
	TMQI	0.7108	0.7429	0.7379	0.7369	0.7422	**0.7542**	0.7395	0.7484
	STD	59.0764	69.4553	59.5870	67.6949	71.8906	71.6910	67.9095	**74.9391**
	SF	26.2088	26.6573	21.3770	25.6565	26.5064	23.9968	28.9480	**29.4094**
	EI	62.2546	64.5835	63.5225	63.8979	67.2917	61.6251	71.1088	**74.0612**
	SSIM	0.7106	0.6465	0.6986	0.6388	0.6538	0.6737	0.7141	**0.7264**
MR-SPECT	*FMI*	0.6826	0.6044	0.5650	0.7266	0.7470	0.7046	0.7242	**0.7435**
	VIF	0.4452	0.5460	0.5315	0.5659	0.5432	0.5487	0.5009	**0.5833**
	$Q_{AB/F}$	**0.6346**	0.4115	0.3304	0.5855	0.6043	0.6080	0.6307	0.6002
	Q_Y	0.6117	0.7195	0.7554	0.7618	0.8448	0.8500	0.8170	**0.8599**
	Q_{CB}	0.5819	0.5617	0.5619	0.5281	0.5966	0.5500	0.5968	**0.6155**
	TMQI	0.6901	0.7157	0.7184	0.7145	0.7143	**0.7219**	0.7083	0.7190
	STD	56.4166	64.8633	54.2769	67.7034	67.7515	65.8108	64.8094	**70.7426**
	SF	19.9102	19.9478	17.2870	20.2278	20.2400	18.9368	22.0236	**22.1879**
	EI	55.3838	59.4569	56.6308	60.0496	60.2900	55.5375	62.3058	**64.9607**
	SSIM	0.6348	0.5552	0.6080	0.5434	0.6267	0.6327	0.6333	**0.6449**
Avg runtime (s)		34.57	64.58	0.11	12.69	4.62	74.95	14.26	6.10

Table 2
Objective evaluation results and average execution times using infrared and visible image dataset. The best performance is shown in bold.

	CSR	ResNet	SWT	Bayes	proposed
FMI	0.2963	0.3166	**0.4738**	0.3966	0.4206
VIF	0.2715	0.2539	0.2988	0.2117	**0.3612**
$Q_{AB/F}$	0.4774	0.3222	0.4884	0.3989	**0.4888**
Q_Y	0.8015	0.6979	**0.8090**	0.8013	0.8038
Q_{CB}	0.4894	0.5013	0.4975	**0.5377**	0.5015
TMQI	0.7038	0.6999	0.6937	0.6777	**0.7103**
STD	23.4193	22.4443	36.1532	27.4666	**36.1935**
SF	9.0527	6.0869	10.6624	7.5415	**11.1993**
EI	32.0732	23.4465	40.8319	27.3616	**43.0047**
SSIM	0.4723	0.5096	0.4790	0.5070	**0.5109**
Avg runtime (s)	140.05	3.76	65.06	2.07	18.37

and blurred edges in the results of ResNet and Bayes are also reflected in their *STD*, *SF* and Q_Y results, all of which are exceptionally low. The CSR method also uses plain averaging for the fusion of low resolution components which results in a considerable attenuation of local intensities. SWT on the other hand preserves the local intensities and edge information in the source images. However, since this method reconstructs the final fused images from transform coefficients, it often results in a loss of fine texture details. The objective evaluation metrics are always relatively high for SWT, but often inferior to those of the proposed method.

The best overall results in terms of fusion of local intensities and preservation of edge information and contrast resolution are obtained by the proposed method. This is also validated by the best overall performance in terms of objective evaluation metrics.

In particular, only the visible images contain visual details. Moreover, the details of some objects are visible only in the infrared images. The proposed method captures this information in the uncorrelated components (see Fig. 10) and directly transfers them to the fused image. For example, in Fig. 9, the texts on the sun shades (fourth row) and the bush leaves (third row) are well preserved in our results. On the other hand, the objects and surfaces visible in both source images are captured as correlated components. Thus, the most visible representation of the correlated features, such as brighter surfaces and sharper edges, are used in the fused images. This is illustrated in the third row of Fig. 9, where the proposed method captures the most visible details of the person (in the infrared image) and the roof (in the visible image).

The average execution times using all the images in the infrared-visible image dataset are given in the last rwo of Table 2. The results show that the proposed method is significantly faster than the CSR and the SWT methods but slower than ResNet and

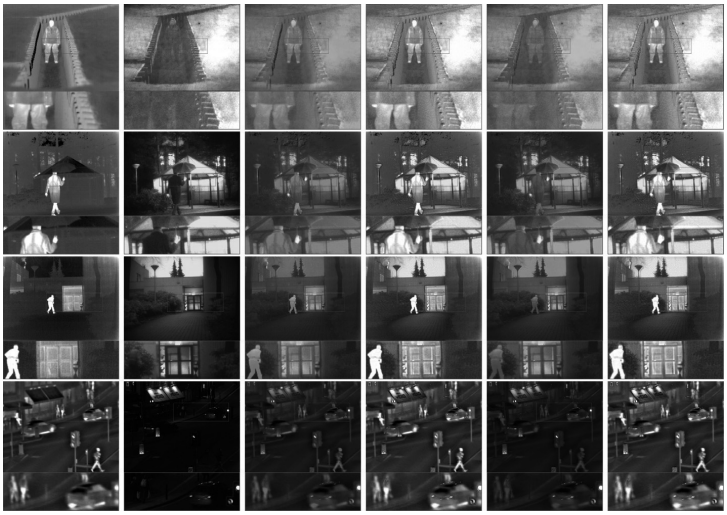


Fig. 9. Four examples of (a) infrared and (b) visible image fusion results using (c) Resnet, (d) SWT, (e) Bayes and (f) the proposed method.

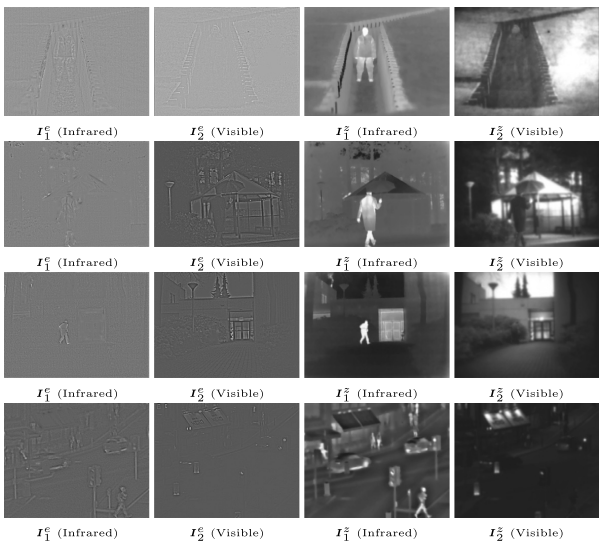


Fig. 10. Decomposition components obtained using the proposed method for the infrared-visible images in Fig. 9. For better visualization, the components are converted to the standard image range.

especially Bayes method. However, the ResNet method uses a pre-trained convolutional neural network available in MATLAB, which takes ~ 136 seconds to be loaded on the computer used in our experiments. Note that using a low-level programming language such as C++ can dramatically improve the speed of the proposed method and make it applicable to real-time tasks.

5.4. Parameter tuning

In this section, we discuss our strategy for selecting optimal parameters for the proposed decomposition model. The parameters are listed in Table 3. Optimal parameters must meet three criteria: (1) the decomposition must be accurate; (2) low correlation between the uncorrelated components must be ensured; (3) the chosen parameters must be computationally efficient (smaller n and T are preferred).

The best patch size is related to the size of local features in the input images. The patch size also impacts the computational cost of the fusion problem. We use a value of $m = 64$ in the experiments in order to achieve the best compromise between running time and effective capturing of features. In our experiments, we

Table 3
The parameters of the proposed image decomposition model.

ϵ:	the maximum allowed residual norm in sparse approximation
T:	the maximum number of nonzero entries in the sparse codes
m:	the patch size (also the size of the dictionary atoms)
n:	the number of atoms in the dictionaries
ρ:	the tuning parameter in optimization algorithm (7)
δ:	the constant used for stabilization of $\{\mathbf{E_1}, \mathbf{E_2}\}$ updates (9)

Table 4
The parameters of the image decomposition model and the resulting performances.

Multimodal Dataset	n	T	ρ	avg MSE	avg Correlation
Anatomical–Anatomical:	100	6	5	9.84×10^{-5}	0.0035
Functional–Anatomical:	16	3	10	4.73×10^{-5}	0.0054
Infrared–Visible:	16	3	5	3.22×10^{-5}	0.0092

observed that $\epsilon = \sqrt{m} \times 10^{-3}$ and $\delta = 10^{-6}$ lead to the best overall performance in all presented fusion problems. An exhaustive grid search was carried out to find the optimal values for the remaining parameters n, T and ρ. Table 4 summarizes the optimal parameters selected for each of the fusion problems in the experiments as well as the resulting performance in terms of the mean squared errors (MSE) of the decomposition and the absolute Pearson correlation coefficient between the associated entries of the uncorrelated components (both averaged over all images in each dataset). Note that average correlations are calculated using only pixels with enough variance (*i.e.*, when $\sigma_{1,j}^2 \sigma_{2,j}^2 > \delta$), so that dark/no signal regions are ignored.

The functional medical images are characterized by very low contrast-resolutions. Also, it should be noted that, infrared images usually do not contain any texture information. As a result, the correlated features associated with these two imaging modalities are very sparse and can be estimated using a small dictionary and few samples (see n and T values in Table 4). In contrast, the medical anatomical images contain high resolution information and fine texture details. Consequently, the estimation of correlated components in anatomical-anatomical fusion requires a relatively larger dictionary and more samples as can be seen in Table 4.

6. Conclusion

A novel image fusion method for multimodal images has been presented. A decomposition method separates input images into their correlated (*i.e.*, common to both images) and uncorrelated (modality-specific) components. The correlated components are captured by sparse representations with identical supports and learned coupled dictionaries. The low correlation between the uncorrelated components is enforced by the minimization of pixel-wise Pearson correlations. An alternating optimization strategy is adopted for addressing the resulting optimization problem. One particularity of the proposed method is that it applies a fusion rule to the correlated components only while fully preserving the uncorrelated components. In the experiments, this strategy has shown superior preservation of intensity and detail compared to other recent methods. Quantitative evaluation metrics and comparison of execution times have also shown the competitiveness of the proposed method.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Farshad G. Veshki: Writing – review & editing, Conceptualization, Methodology, Software, Visualization, Formal analysis, Data curation. **Nora Ouzir:** Writing – review & editing, Conceptualization, Formal analysis, Methodology. **Sergiy A. Vorobyov:** Writing – review & editing, Conceptualization, Supervision, Funding acquisition. **Esa Ollila:** Writing – review & editing, Conceptualization, Supervision.

Appendix A. Modified SOMP algorithm

The steps of the proposed modified SOMP method are explained in Algorithm A.1. Symbol $|\cdot|$ denotes the cardinality of a set (number of elements), operator $\mathbf{1}(\cdot)$ returns one if the condition is true and zero otherwise, and symbol $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse.

Algorithm 2 SOMP with coupled dictionaries.

Input: Data matrices $\boldsymbol{X}_k \in \mathbb{R}^{m\times p}$, $k = 1, \ldots, K$, coupled dictionaries $\boldsymbol{D}_k \in \mathbb{R}^{m\times n}$, $k = 1, \ldots, K$, error threshold ϵ, and maximum number of non-zero coefficients T.

Initialization: sparse representations $\boldsymbol{A}_k = \mathbf{0} \in \mathbb{R}^{n\times p}$, $k = 1, \ldots, K$.
for $i = 1, \cdots, p$ **do**
  $\mathcal{J} = \{\}$, $\boldsymbol{r}_k = [\boldsymbol{X}_k]_i$, $k = 1, \ldots, K$
  while $|\mathcal{J}| < T$ **and** $\sum_{k=1}^{K} \mathbf{1}\big(\|\boldsymbol{r}_k\|_2 \geq \epsilon\big) > 1$ **do**
    $s = \underset{t}{\operatorname{argmax}} \sum_{k=1}^{K} (\boldsymbol{r}_k^T [\boldsymbol{D}_k]_t)^2$ ▷ Simultaneous atom selection
    $\mathcal{J} = \{\mathcal{J}, s\}$ ▷ Updating the support
    $[\boldsymbol{A}_k]_{(\mathcal{J},i)} = [\boldsymbol{D}_k]_{(\cdot,\mathcal{J})}^{\dagger} [\boldsymbol{X}_k]_i, \quad k = 1, \ldots, K$ ▷ Orthogonal projection
    $\boldsymbol{r}_k = [\boldsymbol{X}_k]_i - \boldsymbol{D}_k [\boldsymbol{A}_k]_i, \quad k = 1, \ldots, K$ ▷ Updating the residuals
  end while
end for

Output: Sparse representations with common supports $\boldsymbol{A}_k, k = 1, \ldots, K$

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sigpro.2022.108637

References

[1] X. Zhang, P. Ye, H. Leung, K. Gong, G. Xiao, Object fusion tracking based on visible and infrared images: a comprehensive review, Inf. Fusion 63 (2020) 166–187.
[2] L. Ren, Z. Pan, J. Cao, H. Zhang, H. Wang, Infrared and visible image fusion based on edge-preserving guided filter and infrared feature decomposition, Signal Process. 186 (2021) 108108.
[3] F.I. Arnous, R.M. Narayanan, B.C. Li, Application of multidomain data fusion, machine learning and feature learning paradigms towards enhanced image-based SAR class vehicle recognition, in: Radar Sensor Technology XXV, 11742, 2021, pp. 35–46.
[4] R. Dian, S. Li, B. Sun, A. Guo, Recent advances and new guidelines on hyperspectral and multispectral image fusion, Inf. Fusion 69 (2021) 40–51.
[5] Y. Peng, W. Li, X. Luo, J. Du, Y. Gan, X. Gao, Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporalspectral fusion of remote sensing images, Inf. Fusion 65 (2021) 21–36.
[6] L.-J. Deng, M. Feng, X.-. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-laplacian prior, Inf. Fusion 52 (2019) 76–89.
[7] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: a survey of the state of the art, Inf. Fusion 33 (2017) 100–112.
[8] B. Huang, F. Yang, M. Yin, X. Mo, C. Zhong, A review of multimodal medical image fusion techniques, Comput. Math. Methods. Med. 2020 (2020).

[9] H. Hermessi, O. Mourali, E. Zagrouba, Multimodal medical image fusion review: theoretical background and recent advances, Signal Process. 183 (2021).
[10] Z. Zhao, S. Xu, C. Zhang, J. Liu, J. Zhang, Bayesian fusion for infrared and visible images, Signal Process. 177 (2020) 1–12.
[11] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, S. Lee, K. He, Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain, Infrared Phys. Technol. 88 (2018) 1–12.
[12] G. Li, Y. Lin, X. Qu, An infrared and visible image fusion method based on multi-scale transformation and norm optimization, Inf. Fusion 71 (2021) 109–129.
[13] M. Yin, X. Liu, Y. Liu, X. Chen, Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain, IEEE Trans. Instrum. Meas. 68 (1) (2019) 49–64.
[14] J. Du, W. Li, B. Xiao, Q. Nawaz, Union Laplacian pyramid with multiple features for medical image fusion, Neurocomputing 194 (2016) 326–339.
[15] J. Du, W. Li, B. Xiao, Anatomical-functional image fusion by information of interest in local Laplacian filtering domain, IEEE Trans. Image Process. 26 (12) (2017) 5855–5865.
[16] J. Du, W. Li, B. Xiao, Fusion of anatomical and functional images using parallel saliency features, Inf. Sci. 430–431 (2018) 567–576.
[17] J. Du, W. Li, B. Xiao, Q. Nawaz, Medical image fusion by combining parallel features on multi-scale local extrema scheme, Knowl. Based Syst. 113 (2016) 4–12.
[18] X. Li, X. Guo, P. Han, X. Wang, H. Li, T. Luo, Laplacian redecomposition for multimodal medical image fusion, IEEE Trans. Instrum. Meas. 69 (9) (2020) 6880–6890.
[19] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett. 23 (12) (2016) 1882–1886.
[20] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Medical image fusion via convolutional sparsity based morphological component analysis, IEEE Signal Process. Lett. 26 (3) (2019) 485–489.
[21] Y. Jiang, M. Wang, Image fusion with morphological component analysis, Inf. Fusion 18 (2014) 107–118.
[22] H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognit. 79 (2018) 130–146.
[23] S. Singh, R.S. Anand, Multimodal medical image sensor fusion model using sparse k-SVD dictionary learning in nonsubsampled shearlet domain, IEEE Trans. Instrum. Meas. 69 (2) (2020) 593–607.
[24] D.P. Bavirisetti, G. Xiao, G. Liu, Multi-sensor image fusion based on fourth order partial differential equations, in: Int. Conf. Inf. Fusion, 2017, pp. 1–9.
[25] Y. Yang, S. Cao, S. Huang, W. Wan, Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency, IEEE Trans. Instrum. Meas. 70 (2021) 1–16.
[26] H. Li, X. Wu, T.S. Durrani, Infrared and visible image fusion with resnet and zero-phase component analysis, Infrared Phys. Technol. 102 (2019) 103039.
[27] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, in: Proc. 20th Int. Conf. Inf. Fusion, Xi'an, China, 2017, pp. 1–7.
[28] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, X. Guan, Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform, Expert Syst. Appl. 171 (2021) 114574.
[29] F.G. Veshki, S.A. Vorobyov, An efficient coupled dictionary learning method, IEEE Signal Process. Lett. 26 (10) (2019) 1441–1445.
[30] F.G. Veshki, 2021, (https://github.com/FarshadGVeshki/CFL_for_MMIF).
[31] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, IEEE Trans. Image Process. 21 (8) (2012) 3467–3478.
[32] F.G. Veshki, N. Ouzir, S.A. Vorobyov, Image fusion using joint sparse representations and coupled dictionary learning, in: Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Barcelona, Spain, 2020, pp. 8344–8348.
[33] S. Wang, L. Zhang, Y. Liang, Q. Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI, USA, 2012, pp. 2216–2223.
[34] K. Sharma, S.K. Yadav, B. Valluru, L. Liu, Significance of MRI in the diagnosis and differentiation of clear cell sarcoma of tendon and aponeurosis (CCSTA), Medicine 97 (31) (2018) e111012.
[35] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[36] K. Engan, S.O. Aase, J.H. Husoy, Method of optimal directions for frame design, in: Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Phoenix, AZ, USA, 1999, pp. 2443–2446.
[37] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: Proc. ACM Int. Conf. Mach. Learn., Montreal, QC, Canada, 2009, pp. 689–696.
[38] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, USA, 1993, pp. 40–44.
[39] J. Tropp, A. Gilbert, M. Strauss, Algorithms for simultaneous sparse approximation. part I: greedy pursuit, Signal Process. 86 (2006) 572588.
[40] Y. Liu, S. Liu, A. Wang, A general framework for image fusion based on multi--scale transform and sparse representation, Inf. Fusion 24 (2015) 1047–1064.
[41] S. Klein, M. Staring, K. Murphy, M. Viergever, J. Pluim, Elastix: a toolbox for intensity-based medical image registration, IEEE Trans. Med. Imaging 29 (1) (2010) 196–205.
[42] D. Stein, K. Fritzsche, M. Nolden, H. Meinzer, I. Wolf, The extensible open–source rigid and affine image registration module of the medical imaging interaction toolkit (mitk), Comput. Methods Programs Biomed. 100 (1) (2010) 79–86.
[43] J. Sulam, M. Elad, Expected patch log likelihood with a sparse prior, in: Proc. Int. Workshop Energy Minimization Methods Comput. Vision Pattern Recognit. (EMMCVPR), Hong Kong, China, 2015, pp. 99–111.
[44] R.B. Millar, Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB, Wiley, 2011.
[45] B. Wohlberg, Efficient algorithms for convolutional sparse representations, IEEE Trans. Image Process. 25 (1) (2016) 301–315.
[46] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuits, Inf. Fusion 13 (2012) 10–19.
[47] Harvard Medical School, The whole brain atlas, ????, (http://www.med.harvard.edu/AANLIB/). [Online; accessed 9-may-2021].
[48] H. Li, ????, (https://github.com/hli1221/imagefusion_resnet50/tree/master/IV_images). [Online; accessed 9-may-2021].
[49] W. Tan, P.T.H.M. Pandey, C. Moreira, A.K. Jaiswal, Multimodal medical image fusion algorithm in the era of big data, Neural Comput. Appl. (2020).
[50] H. Yeganeh, Z. Wang, Objective quality assessment of tone mapped images, IEEE Trans. Image Process. 22 (2) (2013) 657–667.
[51] C. Yang, J.Q. Zhang, X.R. Wang, X. Liu, A novel similarity based quality metric for image fusion, Inf. Fusion 9 (2) (2008) 156–160.
[52] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (10) (2009) 1421–1432.
[53] A.A. M. B. A. Haghighat, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features, Comput. Electr. Eng. 37 (5) (2011) 744–756.
[54] Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity, Inf. Fusion 14 (2) (2013) 127135.
[55] C. Xydeas, V. Petrovic, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308–309.
[56] A. Eskicioglu, P. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 43 (12) (1995) 2959–2965.
[57] B. Rajalingam, R. Priya, R. Bhavani, Hybrid multimodal medical image fusion using combination of transform techniques for disease analysis, Procedia Comput. Sci. 152 (2019) 150–157.
[58] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

Publication V

F. G. Veshki and S. A. Vorobyov. Coupled Feature Learning Via Structured Convolutional Sparse Coding for Multimodal Image Fusion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, pp. 2500-2504, May 2022.

COUPLED FEATURE LEARNING VIA STRUCTURED CONVOLUTIONAL SPARSE CODING FOR MULTIMODAL IMAGE FUSION

Farshad G. Veshki and Sergiy A. Vorobyov

Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

ABSTRACT

A novel method for learning correlated features in multimodal images based on convolutional sparse coding with applications to image fusion is presented. In particular, the correlated features are captured as coupled filters in convolutional dictionaries. At the same time, the shared and independent features are approximated using separate convolutional sparse codes and a common dictionary. The resulting optimization problem is addressed using alternating direction method of multipliers. The coupled filters are fused based on a maximum-variance rule, and a maximum-absolute-value rule is used to fuse the sparse codes. The proposed method does not entail any prelearning stage. The experimental evaluations using medical and infrared-visible image datasets demonstrate the superiority of our method compared to state-of-the-art algorithms in terms of preserving the details and local intensities as well as improving objective metrics.

Index Terms— Multimodal image fusion, convolutional sparse coding, structured dictionary learning.

1. INTRODUCTION

Multimodal image fusion aims at merging the information from multiple images acquired using different imaging modalities into a single image, without introducing noise or artifacts [1, 2]. For instance, in medical image fusion, different information about the anatomies of tissues or the levels of biological activities captured using various medical imaging modalities are aggregated in a single fused image [1]. In surveillance applications, combining the visual information in optical images and the thermal information captured using infrared imaging techniques yield more informative images, and has applications, for example, in night vision [2].

A common approach for addressing the multimodal image fusion problem is to decompose the input images into multiscale or morphologically distinct components. This is usually done by employing deterministic mathematical models such as multiscale transforms [3–5]. Other techniques used for a similar purpose include subspace learning [6], dictionary learning [7, 8], and deep learning [9, 10]. An assumption made by all aforementioned fusion techniques is that the features (components) with similar structural properties convey the same type of information. Therefore, they are appropriate for fusion. However, the multimodal images may not obey this assumption. For example, in medical imaging, computed tomography (CT) captures hard tissues and structures (e.g., bones and implants) with a higher resolution, while using magnetic resonance (MR) imaging, the details of soft tissues (e.g., fat and bone marrow) are reflected more effectively [1]. In infrared-visible images, the details in each input image provide different types of information. Thus, a fusion based on the similarity of structural properties can lead to the omission of important information.

In a recent work [11], we demonstrated that the fusion performance can be considerably improved by replacing the conventional deterministic feature-extraction techniques with a data-driven approach for extracting correlated features in multimodal images. Specifically, a method based on coupled dictionary learning [12] and a Pearson correlation constraint has been developed to decompose the multimodal images into their correlated and independent components. In particular, the correlated features have been captured as pairs of atoms in the coupled dictionaries. Then, the fusion is performed using the most significant representations of the coupled atoms. Since the information in the independent components is specific to each modality, these components are transferred to the fused image directly. This approach has shown to be superior in terms of preserving important information while yielding an improved contrast resolution [11].

In this paper, we present a coupled feature learning (CFL) method based on convolutional sparse coding (CSC) and dictionary learning. CSC incorporates a global single-valued model that, unlike standard sparse approximation, does not require patch extraction and enables shift-invariant dictionary learning [13]. In addition, instead of minimizing linear correlations between independent components (as in [11]), we incorporate a more general model that promotes statistical independence. We also propose novel schemes for fusion of correlated features and reconstruction of the final fused image. Experimental evaluations using multimodal medical and infrared-visible image datasets show that the proposed method significantly improves the performance of state-of-the-art multimodal fusion techniques. A MATLAB implementation of our fusion method is available at `https://github.com/FarshadGVeshki/ConvCFL-MMIF`.

2. CONVOLUTIONAL COUPLED FEATURE LEARNING

The proposed model decomposes n input multimodal images $\boldsymbol{s}^i \in \mathbb{R}^N$, $i = 1, \ldots, n$, where N is the number of pixels in the images, into their correlated, shared and independent components. For simplicity of notations, we use one-dimensional arrays to represent the images. Generalization to multi-dimensional arrays is mathematically straightforward.

2.1. Problem Formulation

The correlated components are captured using a set of common sparse feature maps $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times K}$ and coupled convolutional dictionaries $\boldsymbol{D}^i \in \mathbb{R}^{M\times K}$, $i = 1, \ldots, n$. The shared and independent components are represented using a common dictionary $\boldsymbol{C} \in \mathbb{R}^{M\times L}$ and separate sparse feature maps $\boldsymbol{X}^i \in \mathbb{R}^{N\times L}$, $i = 1, \ldots, n$. The decomposition problem can then be formulated as the following optimization problem

$$\underset{\substack{\{\boldsymbol{D}^i\}_{i=1}^n, \boldsymbol{C},\\ \{\boldsymbol{X}^i\}_{i=1}^n, \boldsymbol{\Gamma}}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 \quad \text{s.t.} \quad \boldsymbol{C}_l, \boldsymbol{D}_k^i, \in \boldsymbol{\mathcal{D}} \quad \forall k, l, i, \quad (1)$$

where $\boldsymbol{\mathcal{D}} = \{\boldsymbol{d} \in \mathbb{R}^M | \|\boldsymbol{d}\|_2 \leq 1\}$ is the set of dictionary filters and $\lambda_1 > 0$ and $\lambda_2 > 0$ are regularization parameters. Subscripts are used to denote the columns of matrices.

The overlapping nonzero entries in $\{\boldsymbol{X}^i\}_{i=1}^n$ indicate that one of the dictionary filters $\{\boldsymbol{C}_l\}_{l=1}^L$ is used for approximation of multiple images at the same location, thus, it represents a shared feature. In addition, when only one of the entries in $\{\boldsymbol{X}^i\}_{i=1}^n$ is nonzero at one location, it means that one of the filters in $\{\boldsymbol{C}_l\}_{l=1}^L$ is used for only one source image. Thus, it represents an independent feature. Note that the convolutional filters are assumed to be statistically independent. As first demonstrated in [14], dictionary learning promotes statistical independence. The proof relies on the fact that accurate sparse codes preserve the information (joint entropy) in the source signal. Moreover, the sparsity regularization minimizes the entropy in each of the sparse codes (simply by maximizing the probability of one event (being zero) and minimizing the probability of all other events (being nonzero)). Therefore, by enforcing the equality of the joint entropy and the sum of the entropies of the individual sparse codes, sparse dictionary learning promotes statistical independence.

2.2. Optimization

Problem (1) is typically solved by alternating between minimization over the sparse codes and the dictionary filters. Since we address both steps in Fourier domain (using [15]), we first zero-pad all of the dictionary filters to the size of the sparse coefficient maps ($\mathbb{R}^N$).

2.2.1. Sparse Coding Step

Using the consensus ADMM method [16], (1) can be addressed with respect to the sparse feature maps $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ by solving the following optimization problem

$$\underset{\{\boldsymbol{X}^i\}_{i=1}^n, \boldsymbol{\Gamma}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Theta}_k^i + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{Y}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 \quad \text{s.t.} \quad \boldsymbol{\Gamma} = \boldsymbol{\Theta}^i, \ \boldsymbol{X}^i = \boldsymbol{Y}^i \quad i = 1, \ldots, n.$$

Using scaled Lagrangian multipliers $\boldsymbol{U}^i \in \mathbb{R}^{N\times K}$ and $\boldsymbol{V}^i \in \mathbb{R}^{N\times L}$,$i = 1, \ldots, n$, the augmented Lagrangian is written as

$$\frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Theta}_k^i + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{Y}_l^i - \boldsymbol{s}^i\Big\|_2^2 + \lambda_1\sum_{k=1}^{K}\|\boldsymbol{\Gamma}_k\|_1 + \lambda_2\sum_{i=1}^{n}\sum_{l=1}^{L}\|\boldsymbol{X}_k^i\|_1 + \frac{\rho}{2}\sum_{i=1}^{n}\Big(\sum_{k=1}^{K}\|\boldsymbol{\Theta}_k^i - \boldsymbol{\Gamma}_k + \boldsymbol{U}_k^i\|_2^2 + \sum_{l=1}^{L}\|\boldsymbol{Y}_l^i - \boldsymbol{X}_l^i + \boldsymbol{V}_l^i\|_2^2\Big), \quad (2)$$

where $\rho > 0$ is the penalty parameter. The ADMM iterations consist of minimizing (2) alternatively with respect to $\{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}_{i=1}^n$, $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ and $\{\boldsymbol{U}^i, \boldsymbol{V}^i\}_{i=1}^n$. The details of each subproblem are explained in the following. Denoting $\boldsymbol{Z}^i = \{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}$, $\boldsymbol{F}^i = \{\boldsymbol{D}^i, \boldsymbol{C}\}$ and $\boldsymbol{W}^i = \{\boldsymbol{\Gamma} - \boldsymbol{U}^i, \boldsymbol{X}^i - \boldsymbol{V}^i\}$, we can update $\{\boldsymbol{\Theta}^i, \boldsymbol{Y}^i\}_{i=1}^n$ by solving n parallel optimization problems, which can be written as follows

$$(\boldsymbol{Z}^i)^+ = \underset{\boldsymbol{Z}^i}{\text{argmin}} \frac{1}{2}\Big\|\sum_{p=1}^{P}\boldsymbol{F}_p^i * \boldsymbol{Z}_p^i - \boldsymbol{s}^i\Big\|_2^2 + \frac{\rho}{2}\sum_{p=1}^{P}\|\boldsymbol{Z}_p^i - \boldsymbol{W}_p^i\|_2^2, \quad (3)$$

where $P = K + L$ and $(\cdot)^+$ denotes the updated variables. The problem in (3) is a standard convolutional fitting problem that can be addressed using available CSC methods (e.g., [15]).

Updating $\{\boldsymbol{\Gamma}, \{\boldsymbol{X}^i\}_{i=1}^n\}$ can be efficiently addressed in an elementwise manner using the shrinkage operator $\mathcal{S}_\kappa(a) = \text{sign}(a)\max(0, |a| - \kappa)$. The updates are written as follows

$$\boldsymbol{\Gamma}^+ = \mathcal{S}_{\lambda_1/\rho}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Theta}^i + \boldsymbol{U}^i\Big), \ (\boldsymbol{X}^i)^+ = \mathcal{S}_{\lambda_2/\rho}(\boldsymbol{Y}^i + \boldsymbol{V}^i), \ i = 1, \ldots, n.$$

Finally, the updates for the scaled Lagrangian variables $\{\boldsymbol{U}^i, \boldsymbol{V}^i\}_{i=1}^n$ are given as

$$(\boldsymbol{U}^i)^+ = \boldsymbol{\Theta}^i - \boldsymbol{\Gamma} + \boldsymbol{U}^i, \quad (\boldsymbol{V}^i)^+ = \boldsymbol{Y}^i - \boldsymbol{X}^i + \boldsymbol{V}^i, \quad i = 1, \ldots, n.$$

2.2.2. Dictionary Update Step

Using the consensus ADMM, (1) can be reformulated with respect to the dictionary filters $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ as

$$\underset{\{\boldsymbol{D}^i\}_{i=1}^n, \boldsymbol{C}}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{G}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{H}_l^i * \boldsymbol{X}^i - \boldsymbol{s}^i\Big\|_2^2 + \boldsymbol{\Omega}\Big(\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}\Big) \quad \text{s.t.} \quad \boldsymbol{C} = \boldsymbol{H}^i, \quad \boldsymbol{D}^i = \boldsymbol{G}^i, \quad i = 1, \ldots, n$$

where $\boldsymbol{\Omega}(\cdot)$ is an indicator function of the constraint set in (1). The augmented Lagrangian is written as follows

$$\frac{1}{2}\sum_{i=1}^{n}\Big\|\sum_{k=1}^{K}\boldsymbol{G}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{H}_l^i * \boldsymbol{X}^i - \boldsymbol{s}^i\Big\|_2^2 + \boldsymbol{\Omega}(\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}) + \frac{\sigma}{2}\sum_{i=1}^{n}\Big(\sum_{k=1}^{K}\|\boldsymbol{G}_k^i - \boldsymbol{D}_k^i + \boldsymbol{R}_k^i\|_2^2 + \sum_{l=1}^{L}\|\boldsymbol{H}_l^i - \boldsymbol{C}_l + \boldsymbol{T}_l^i\|_2^2\Big), \quad (4)$$

where $\boldsymbol{R}^i \in \mathbb{R}^{N\times K}$ and $\boldsymbol{T}^i \in \mathbb{R}^{N\times L}$,$i = 1,\ldots,n$, are scaled Lagrangian variables. The ADMM iterations consist of minimizing (4) alternatively with respect to $\{\boldsymbol{G}^i, \boldsymbol{H}^i\}_{i=1}^n$, $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ and $\{\boldsymbol{R}^i, \boldsymbol{T}^i\}_{i=1}^n$.

Indeed, denote $\boldsymbol{E}^i = \{\boldsymbol{G}^i, \boldsymbol{H}^i\}, \boldsymbol{S}^i = \{\boldsymbol{\Gamma}, \boldsymbol{X}^i\}$ and $\boldsymbol{Q}^i = \{\boldsymbol{D}^i - \boldsymbol{R}^i, \boldsymbol{C} - \boldsymbol{T}^i\}, i = 1,\ldots,n,$. Then updating $\{\boldsymbol{G}^i, \boldsymbol{H}^i\}_{i=1}^n$ can be addressed by solving n parallel optimization problems

$$(\boldsymbol{E}^i)^+ = \underset{\boldsymbol{E}^i}{\operatorname{argmin}} \frac{1}{2}\|\sum_{p=1}^{P} \boldsymbol{E}_p^i * \boldsymbol{S}_p^i - \boldsymbol{s}^i\|_2^2 + \frac{\sigma}{2}\sum_{p=1}^{P}\|\boldsymbol{E}_p^i - \boldsymbol{Q}_p^i\|_2^2. \tag{5}$$

Problem (5) is similar to (3) and can be efficiently addressed using available CSC methods (e.g., [15]).

Updating $\{\boldsymbol{C}, \{\boldsymbol{D}^i\}_{i=1}^n\}$ is performed as

$$(\boldsymbol{D}^i)^+ = \operatorname{proj}_{\boldsymbol{\mathcal{D}}}(\boldsymbol{G}^i + \boldsymbol{R}^i), i = 1,\ldots,n, \quad \boldsymbol{C}^+ = \operatorname{proj}_{\boldsymbol{\mathcal{D}}}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{H}^i + \boldsymbol{T}^i\Big),$$

where $\operatorname{proj}_{\boldsymbol{\mathcal{D}}}(\cdot)$ denotes the orthogonal projection onto the set $\boldsymbol{\mathcal{D}}$. This can be done by mapping the entries outside the constraint support to zero and then projecting the filters on the unit-ball.

The updates for scaled Lagrangian variables are given as

$$(\boldsymbol{R}^i)^+ = \boldsymbol{G}^i - \boldsymbol{D}^i + \boldsymbol{R}^i, \quad (\boldsymbol{T}^i)^+ = \boldsymbol{H}^i - \boldsymbol{C} + \boldsymbol{T}^i. \quad i = 1,\ldots,n.$$

We perform the sparse coding and the dictionary update steps in an interleaved manner (one iteration of each step is executed before passing the variables to the next). The updated ADMM variables (auxiliary variables and scaled Lagrangian multipliers) are used to initialize the next iteration.

2.3. Projection on the Sparse Support

After the convolutional CFL stage, we can still significantly improve the approximation accuracy by orthogonalizing the residuals on the supports (the set of indices of nonzero entries) of the sparse coefficient maps. For this purpose, we use a gradient descent (GD) approach. Based on the convolution theorem, the GD iterations are found as

$$[\boldsymbol{\Gamma}_k^+]_{\mathcal{S}(\boldsymbol{\Gamma}_k)} = [\boldsymbol{\Gamma}_k]_{\mathcal{S}(\boldsymbol{\Gamma}_k)} - \alpha\Big[\mathrm{DFT}^{-1}\Big(\sum_{i=1}^{n}\bar{\hat{\boldsymbol{D}}}_k^i \odot \hat{\boldsymbol{r}}^i\Big)\Big]_{\mathcal{S}(\boldsymbol{\Gamma}_k)}, \forall k,$$

$$[(\boldsymbol{X}_l^i)^+]_{\mathcal{S}(\boldsymbol{X}_l^i)} = [\boldsymbol{X}_l^i]_{\mathcal{S}(\boldsymbol{X}_l^i)} - \alpha\Big[\mathrm{DFT}^{-1}\Big(\bar{\hat{\boldsymbol{C}}}_l^i \odot \hat{\boldsymbol{r}}^i\Big)\Big]_{\mathcal{S}(\boldsymbol{X}_l^i)}, \forall l, i,$$

where $\hat{(\cdot)}$ denotes the discrete Fourier transform and $\mathrm{DFT}^{-1}(\cdot)$ represents its inverse, $\bar{(\cdot)}$ denotes the complex-conjugate, $\odot$ is the elementwise multiplication and operator $\mathcal{S}(\cdot)$ returns the support of an array. In addition, α is the stepsize and $\boldsymbol{r}^i$ represents the residuals associated with $\boldsymbol{s}^i$, that is

$$\boldsymbol{r}^i = \sum_{k=1}^{K}\boldsymbol{D}_k^i * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^i - \boldsymbol{s}^i, i = 1,\ldots,n.$$

3. MULTIMODAL IMAGE FUSION ALGORITHM

In this section, the steps of the proposed fusion method are explained. Note that the images are considered as one-dimensional arrays. The elementwise operations are applied to all pixels.

3.1. Low-pass Filtering

The input images are first decomposed into base-layers $\{\boldsymbol{s}_b^i\}_{i=1}^n$ and details-layers $\{\boldsymbol{s}_d^i\}_{i=1}^n$ using low-pass filtering. This is done using the *lowpass* function from the SPORCO library [17] (the regularization parameter is set to 10).

3.2. Fusion of the Details-layers

The details-layers $\{\boldsymbol{s}_d^i\}_{i=1}^n$ are decomposed into the correlated, shared and independent components using the convolutional CFL method explained in Section 2.

3.2.1. Fusion of Coupled Features

The coupled features are fused based on the highest visual significance, which can be measured, for example, using *variance* (denoted as $\mathrm{var}(\cdot)$). This can be formulated as follows

$$\boldsymbol{D}_k^F = \boldsymbol{D}_k^{i^*}, \quad i^* = \underset{i=1,\ldots,n}{\operatorname{argmax}}\Big(\mathrm{var}(\boldsymbol{D}_k^i)\Big), \quad k = 1,\ldots,K,$$

where $\boldsymbol{D}^F$ is the dictionary of fused coupled features.

3.2.2. Fusion of Shared and Independent Components

The fusion of shared and independent component $\boldsymbol{X}^F$ is found by combining the redundant sparse codes $\{\boldsymbol{X}^i\}_{i=1}^n$ using maximum-absolute-value rule. This can be written as

$$\boldsymbol{X}_l^F(j) = \boldsymbol{X}_l^{i^*}(j), i^* = \underset{i=1,\ldots,n}{\operatorname{argmax}}\Big(|(\boldsymbol{X}_l^i(j)|\Big), j = 1,\ldots,N, l = 1,\ldots,L.$$

This allows to transfer the independent features along with the shared features with the most significant representation coefficients into the fused image.

The fused details-layer $\boldsymbol{s}_d^F$ is then reconstructed using

$$\boldsymbol{s}_d^F = \sum_{k=1}^{K}\boldsymbol{D}_k^F * \boldsymbol{\Gamma}_k + \sum_{l=1}^{L}\boldsymbol{C}_l * \boldsymbol{X}_l^F.$$

3.3. Fusion of the Base-layers

We form two images $\boldsymbol{s}_b^{max}$ and $\boldsymbol{s}_b^{min}$ representing the maximum and the minimum allowed local intensities, using

$$\boldsymbol{s}_b^{max} = \max_{i=1,\ldots,n}(\boldsymbol{s}_b^i), \quad \boldsymbol{s}_b^{min} = \omega\Big(\max_{i=1,\ldots,n}(\boldsymbol{s}_b^i)\Big) + (1-\omega)\Big(\min_{i=1,\ldots,n}(\boldsymbol{s}_b^i)\Big),$$

where $0 \le \omega \le 1$, and $\max(\cdot)$ and $\min(\cdot)$ are the elementwise maximum and minimum operators, respectively.

It is favorable to incorporate $\boldsymbol{s}_b^{max}$ into the final fused image. However, this can cause a loss of information due to the limited range (0 to 1) of the standard images. To achieve a compromise between contrast resolutions and local intensities, we propose the following approach. First, the difference between the local maximum and minimum intensities (local variations) of $\boldsymbol{s}_d^F$ (for example, in a 3×3 neighborhood) is stored in $\boldsymbol{s}_d^v$. Then the fused base-layer $\boldsymbol{s}_b^F$ is computed as

$$\boldsymbol{s}_b^F(j) = \begin{cases} \boldsymbol{s}_b^{max}(j), & \text{if } \boldsymbol{s}_d^v(j) \leq 1 - \boldsymbol{s}_b^{max}(j) \\ \boldsymbol{s}_b^{min}(j), & \text{if } \boldsymbol{s}_d^v(j) \geq 1 - \boldsymbol{s}_b^{min}(j) \\ 1 - \boldsymbol{s}_b^v(j), & \text{if otherwise} \end{cases}, \quad j = 1, \ldots, N.$$

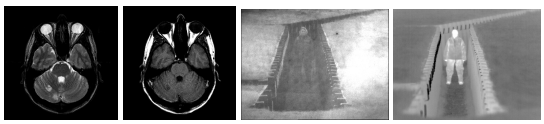
A Gaussian filter may be used to smooth $\boldsymbol{s}_b^F$ so that discontinuities are not introduced.

The final fused image $\boldsymbol{s}^F$ is then reconstructed as

$$\boldsymbol{s}^F = \boldsymbol{s}_b^F + \boldsymbol{s}_d^F.$$

4. EXPERIMENTAL RESULTS

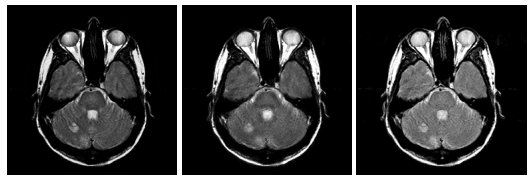
We compare our method to four recent multimodal fusion methods both visually and using objective evaluation metrics. We use two medical image fusion methods: a method based on the non-subsampled shearlet transform (NSST) [4] and a method based on Laplacian redecomposition (LRD) [5]. We also use two infrared-visible image fusion method: a method that incorporates a hierarchical Bayesian model (Bayes) [18] and a method based on deep learning (Resnet) [9]. The multimodal medical image dataset consists of 20 pairs of images collected from [19], and the infrared-visible image dataset includes 21 pairs of images taken from `https://github.com/hli1221/imagefusion_resnet50/tree/master/IV_images`. Four metrics are used for objective evaluations, the objective image fusion performance measure $\mathrm{Q_{AB/F}}$ [20], the information measure for performance of image fusion $\mathrm{Q_{IM}}$ [21], spatial frequency (SF) [22] and the structural similarity index (SSIM) [23]. The algorithm parameters are λ_1=λ_2= 0.01, K=8, L=12, ρ=σ=10, α=0.01 and $\omega = 0.9$. Moreover, we use 150 ADMM iterations, 100 GD iterations and 8×8 filters (M in two-dimensional case).



(a) MR-T1 (b) MR-T2 (c) Visible (d) Infrared

Fig. 1: Examples of multimodal images.

Fig. 1 shows a pair of images from each dataset used. The results obtained using different methods are shown in Fig. 2. Table 1 compares the average results for objective evaluation metrics for each dataset. The results show that the LRD method leads to low contrast-resolutions, which is reflected in very low SF and $\mathrm{Q_{AB/F}}$ values for this method. NSST also loses/blurs high-resolution information, while this information is well preserved using our method (see Figs. 2a and 2b, for example). Results obtained using Resnet and Bayes show an inferior fusion of local intensities, which results in low visibility of the details in the fused images (see Figs. 2d and 2e, for example). Overall, the proposed method results in the best performance in terms of the fusion of the high-resolution information as well as the local intensities (for example, see Figs. 2c and 2f). These observations can be validated by the objective evaluation results in Table 1, where our method obtains the best results in all cases.



(a) NSST (b) LRD (c) proposed

(d) Bayes (e) Resnet (f) proposed

Fig. 2: The fusion results for the multimodal images in Fig. 1 using different methods.

Metrics	Medical			Infrared-Visible		
	NSST	LRD	proposed	Bayes	Resnet	proposed
$\mathrm{Q_{AB/F}}$	0.5646	0.5278	**0.5667**	0.4561	0.3520	**0.4941**
$\mathrm{Q_{IM}}$	0.6778	0.7341	**0.7424**	0.4044	0.3167	**0.4311**
SF	30.02	28.93	**31.93**	7.63	6.13	**11.11**
SSIM	0.5501	0.6601	**0.7047**	0.5040	0.5072	**0.5121**

Table 1: Average objective evaluation results for each dataset using different methods. The Best results are shown in bold.

5. CONCLUSION

A novel multimodal image fusion method based on convolutional sparse coding has been developed. A convolutional coupled feature learning algorithm has been proposed for the decomposition of multimodal images into correlated, shared, and independent features. Appropriate schemes have been proposed for the fusion of extracted features and reconstruction of the final image. The experimental results show significant improvements by the proposed method compared to the state-of-the-art multimodal image fusion methods.

6. REFERENCES

[1] B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, "A review of multimodal medical image fusion techniques," *Comput. Math. Methods. Med.*, vol. 2020, 2020.

[2] X. Zhang, P. Ye, H. Leung, K. Gong, and G. Xiao, "Object fusion tracking based on visible and infrared images: A comprehensive review," *Inf. Fusion*, vol. 63, pp. 166–187, 2020.

[3] G. Li, Y. Lin, and X. Qu, "An infrared and visible image fusion method based on multi-scale transformation and norm optimization," *Inf. Fusion*, vol. 71, pp. 109–129, 2021.

[4] W. Tan, P. Tiwari H. M. Pandey, C. Moreira, and A. K. Jaiswal, "Multimodal medical image fusion algorithm in the era of big data," *Neural Comput. Appl.*, 2020.

[5] X. Li, X. Guo, P. Han, X. Wang, H. Li, and T. Luo, "Laplacian re-decomposition for multimodal medical image fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6880–6890, 2020.

[6] Y. Yang, S. Cao, S. Huang, and W. Wan, "Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–16, 2021.

[7] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 485–489, 2019.

[8] H. Li, X. He, D. Tao, Y. Tang, and R. Wang, "Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning," *Pattern Recognit.*, vol. 79, pp. 130–146, 2018.

[9] H. Li, X. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," *Infrared Physics and Technology*, vol. 102, 2019.

[10] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform," *Expert Systems with Applications*, vol. 171, 2021.

[11] F. G. Veshki, N. Ouzir, S. A. Vorobyov, and E. Ollila, "Coupled feature learning for multimodal medical image fusion," *arXiv:2102.08641*, 2021.

[12] F. G. Veshki and S. A. Vorobyov, "An efficient coupled dictionary learning method," *IEEE Signal Process. Lett.*, vol. 26, no. 10, pp. 1441–1445, 2019.

[13] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[14] B. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[15] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389–393, 2022.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] B. Wohlberg, "SParse Optimization Research COde (SPORCO)," Software library available from `http://purl.org/brendt/software/sporco`, 2017.

[18] Z. Zhao, S. Xu, C. Zhang, J. Liu, and J. Zhang, "Bayesian fusion for infrared and visible images," *Signal Process.*, vol. 177, pp. 1–12, 2020.

[19] Harvard Medical School, "The Whole Brain Atlas ," `http://www.med.harvard.edu/AANLIB/`, [Online; accessed 16-sep-2021].

[20] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[21] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, no. 7, pp. 313–315, 2002.

[22] A.M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

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

Publication VI

F. G. Veshki and S. A. Vorobyov. Convolutional Simultaneous Sparse Approximation with Applications to RGB-NIR Image Fusion. In *Proceedings of the 56th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, November 2022.

Convolutional Simultaneous Sparse Approximation with Applications to RGB-NIR Image Fusion

Farshad G. Veshki
Dept. Signal Processing and Acoustics
Aalto University
Espoo, Finland
farshad.ghorbaniveshki@aalto.fi

Sergiy A. Vorobyov
Dept. Signal Processing and Acoustics
Aalto University
Espoo, Finland
sergiy.vorobyov@aalto.fi

***Abstract*—Simultaneous sparse approximation (SSA) seeks to represent a set of dependent signals using sparse vectors with identical supports. The SSA model has been used in various signal and image processing applications involving multiple correlated input signals. In this paper, we propose algorithms for convolutional SSA (CSSA) based on the alternating direction method of multipliers. Specifically, we address the CSSA problem with different sparsity structures and the convolutional feature learning problem in multimodal data/signals based on the SSA model. We evaluate the proposed algorithms by applying them to multimodal and multifocus image fusion problems.**

***Index Terms*—Simultaneous sparse approximation, convolutional sparse coding, dictionary learning, image fusion**

I. Introduction

Simultaneous sparse approximation (SSA) aims to reconstruct multiple input signals using sparse representations (SRs) with identical supports, i.e., using different linear combinations of the same subset of atoms in a dictionary [1], [2]. The SSA problem can be written as follows

$$\begin{aligned} \underset{\{\boldsymbol{x}_n\}_{n=1}^N}{\text{minimize}} \; & \sum_{n=1}^{N} \left(\frac{1}{2} \left\| \boldsymbol{D}\boldsymbol{x}_n - \boldsymbol{s}_n \right\|_2^2 + \lambda \left\| \boldsymbol{x}_n \right\|_0 \right) \\ \text{s.t.} \quad & \operatorname{Supp}(\boldsymbol{x}_l) = \operatorname{Supp}(\boldsymbol{x}_m), \quad l, m = 1, \ldots, N, \end{aligned} \tag{1}$$

where $\boldsymbol{D}$, $\{\boldsymbol{x}_n\}_{n=1}^N$ and $\{\boldsymbol{s}_n\}_{n=1}^N$ represent the dictionary, the SRs with identical supports, and the input signals, respectively. Moreover, $\lambda > 0$ is the sparsity regularization parameter, $\|\cdot\|_2$ is the Euclidean norm, $\|\cdot\|_0$ is an operator that counts the nonzero entries of a vector, and $\operatorname{Supp}(\cdot)$ denotes the support of an array. The simultaneous sparsity model has been used in a wide range of signal and image processing applications involving multiple dependent input signals. For example, multi measurement vectors (MMV) problems [3], [4], image fusion [5], [6], anomaly detection [7], and blind source separation [8].

Problem (1) is non-convex and, in general, intractable in polynomial time. A common approach for addressing the SSA problem is convex relaxation using mixed-norms [2], [9]. For a matrix $\boldsymbol{A} \in \mathbb{R}^{R \times C}$, the mixed $\ell_{p,q}$-norm, $p, q \geq 1$, is defined as

$$\|\boldsymbol{A}\|_{p,q} = \left(\sum_{r=1}^{R} \|\boldsymbol{A}(r,\cdot)\|_p^q \right)^{\frac{1}{q}}$$

where $\boldsymbol{A}(r,\cdot)$ is the rth row of $\boldsymbol{A}$, and $\|\cdot\|_p$ denotes the p-norm of a vector. For example, the $\ell_{2,1}$ and the $\ell_{\infty,1}$-norms have been used for addressing the SSA problem in [10] and [2], respectively. An unconstrained convex relaxation of (1) using the $\ell_{2,1}$-norm can be written as

$$\underset{\boldsymbol{X}}{\text{minimize}} \; \frac{1}{2} \sum_{n=1}^{N} \|\boldsymbol{D}\boldsymbol{x}_n - \boldsymbol{s}_n\|_2^2 + \lambda \|\boldsymbol{X}\|_{2,1}, \tag{2}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_N]$. Solving (2) entails minimizing the ℓ_2-norm of the rows (enforcing dense rows) and the sum of the ℓ_2-norms of the rows (promoting all-zero rows) of $\boldsymbol{X}$. Thus, the resulting $\boldsymbol{X}$ is expected to be mostly zeros with only few non-zero and dense rows. This structure is referred to as *row-sparse* structure. A *row-sparse* structure with sparse rows can be enforced by embedding an additional ℓ_1-norm regularization term in the objective function of (2) [9]

$$\underset{\boldsymbol{X}}{\text{minimize}} \frac{1}{2} \sum_{n=1}^{N} \|\boldsymbol{D}\boldsymbol{x}_n - \boldsymbol{s}_n\|_2^2 + \gamma_1 \sum_{n=1}^{N} \|\boldsymbol{x}_n\|_1 + \gamma_2 \|\boldsymbol{X}\|_{2,1}, \tag{3}$$

where $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ are the element-sparsity and row-sparsity regularization parameters, respectively.

In this paper, we extend the SSA problem to the convolutional sparse approximation (CSA) framework. Unlike its conventional counterpart, CSA allows local processing of large signals without first breaking them into vectorized overlapping blocks. Thus, it provides a global, single-valued, and shift-invariant model. Specifically, CSA uses a sum of convolutions instead of the matrix-vector product as in the standard sparse approximation model [11].

We first address the convolutional SSA (CSSA) problem with *row-sparse* structure using the $\ell_{2,1}$-norm regularization (the convolutional extension of problem (2)). Then, we discuss variations of the proposed method for solving problem (3) and SSA with $\ell_{\infty,1}$-norm regularization in the CSA framework. We use the alternating direction method of multipliers (ADMM) as a base optimization approach for solving the corresponding problems. We investigate convolutional dictionary learning (CDL), and coupled feature learning in multimodal data based on CSSA. We evaluate the proposed CSSA and CDL algorithms by applying them to the multifocus image fusion and

the near infrared (NIR) and visible light (VL) image fusion problems. Specifically, a novel NIR-VL image fusion method is proposed. MATLAB implementations of the proposed algorithms are available at https://github.com/FarshadGVeshki/ConvSSA-IF.

II. Convolutional Simultaneous Sparse Approximation

We aim to approximate the input signals $\boldsymbol{s}^{(n)} \in \mathbb{R}^P,\ n = 1,\dots,N$, using the sparse feature maps with identical supports $\boldsymbol{X}^{(n)} \in \mathbb{R}^{P\times K},\ n = 1,\dots,N$, and the dictionary $\boldsymbol{D} \in \mathbb{R}^{Q\times K}$. The columns of $\boldsymbol{X}^{(n)}$ and $\boldsymbol{D}$ are the convolutional SR elements and the convolutional filters, respectively. For simplicity, we consider the case where the input signals are one-dimensional arrays. The proposed method can be straightforwardly generalized to handling multi-dimensional arrays.

A. Problem Formulation

The CSSA problem is formulated as follows

$$\underset{\{\boldsymbol{X}^{(n)}\}_{n=1}^N}{\text{minimize}} \frac{1}{2}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k^{(n)} - \boldsymbol{s}^{(n)}\right\|_2^2 + \lambda\sum_{n=1}^{N}\sum_{k=1}^{K}\left\|\boldsymbol{X}_k^{(n)}\right\|_0 \quad \text{s.t.}\quad \operatorname{Supp}\big(\boldsymbol{X}^{(m)}\big) = \operatorname{Supp}\big(\boldsymbol{X}^{(n)}\big),\quad m,n = 1,\dots,N. \tag{4}$$

Using the $\ell_{2,1}$-norm[1], a convex relaxation of (4) can be written as

$$\underset{\{\boldsymbol{X}^{(n)}\}_{n=1}^N}{\text{minimize}} \frac{1}{2}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k^{(n)} - \boldsymbol{s}^{(n)}\right\|_2^2 + \lambda\sum_{k=1}^{K}\left\|\boldsymbol{\mathcal{X}}^{(k)}\right\|_{2,1} \tag{5}$$

where $\boldsymbol{\mathcal{X}}^{(k)}(p,\cdot) = [\boldsymbol{X}_k^{(1)}(p)\ \cdots\ \boldsymbol{X}_k^{(N)}(p)],\ p = 1,\dots,P$.

B. Optimization Procedure

The ADMM formulation of (5) can be written as

$$\underset{\{\boldsymbol{X}^{(n)},\boldsymbol{Y}^{(n)}\}_{n=1}^N}{\text{minimize}} \frac{1}{2}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{Y}_k^{(n)} - \boldsymbol{s}^{(n)}\right\|_2^2 + \lambda\sum_{k=1}^{K}\left\|\boldsymbol{\mathcal{X}}^{(k)}\right\|_{2,1} \quad \text{s.t.}\quad \boldsymbol{X}^{(n)} = \boldsymbol{Y}^{(n)},\ n = 1,\cdots,N. \tag{6}$$

Then the ADMM iterations are given as

$$(\boldsymbol{Y}^{(n)})^{i+1} = \underset{\boldsymbol{Y}^{(n)}}{\operatorname{argmin}} \frac{1}{2}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{Y}_k^{(n)} - \boldsymbol{s}^{(n)}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{Y}^{(n)} - (\boldsymbol{X}^{(n)})^i + (\boldsymbol{U}^{(n)})^i\right\|_{\mathrm{F}}^2,\ n = 1,\dots,N \tag{7}$$

$$(\{\boldsymbol{X}^{(n)}\}_{n=1}^N)^{i+1} = \underset{\{\boldsymbol{X}^{(n)}\}_{n=1}^N}{\operatorname{argmin}} \lambda\sum_{k=1}^{K}\left\|\boldsymbol{\mathcal{X}}^{(k)}\right\|_{2,1} + \frac{\rho}{2}\sum_{n=1}^{N}\left\|(\boldsymbol{Y}^{(n)})^{i+1} - \boldsymbol{X}^{(n)} + (\boldsymbol{U}^{(n)})^i\right\|_{\mathrm{F}}^2 \tag{8}$$

$$(\boldsymbol{U}^{(n)})^{i+1} = (\boldsymbol{Y}^{(n)})^{i+1} - (\boldsymbol{X}^{(n)})^{i+1} + (\boldsymbol{U}^{(n)})^i, n = 1,\dots,N,$$

[1]The mixed $\ell_{2,1,\dots,1}$-norm can be used for multi-dimensional input signal.

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix, $\{\boldsymbol{U}^{(n)}\}_{n=1}^N$ are the scaled Lagrangian multipliers, and $\rho > 0$ is the ADMM penalty parameter. The Y-update step (7) entails N convolutional regression subproblems which can be addressed using existing CSA methods (e.g., [11]).

Since the $\ell_{2,1}$-norm is a separable sum of the ℓ_2-norms of the rows, (8) can be addressed in a row-wise manner using the proximal operator of the Euclidean norm. Using $\boldsymbol{W}^{(n)} = (\boldsymbol{Y}^{(n)})^{i+1} + (\boldsymbol{U}^{(n)})^i$, the solution to (8) can be calculated as

$$\Big([\boldsymbol{X}_k^{(1)}(p)\ \cdots\ \boldsymbol{X}_k^{(N)}(p)]\Big)^{i+1} = \operatorname{prox}_{\frac{\lambda}{\rho}\|\cdot\|_2}\Big([\boldsymbol{W}_k^{(1)}(p)\ \cdots\ \boldsymbol{W}_k^{(N)}(p)]\Big),\quad k = 1,\dots,K,\ p = 1,\dots,P, \tag{9}$$

with

$$\operatorname{prox}_{\tau\|\cdot\|_2}(\boldsymbol{a}) = \left(1 - \frac{\tau}{\max(\|\boldsymbol{a}\|_2,\tau)}\right)\boldsymbol{a}. \tag{10}$$

C. Other Convex Formulations of CSSA

Problem (4) can be alternatively relaxed using the $\ell_{\infty,1}$-norm. To address the resulting optimization problem, we only need to modify the X-update step of the ADMM algorithm explained in Subsection II-B. Specifically, in (9), we need to replace $\operatorname{prox}_{\frac{\lambda}{\rho}\|\cdot\|_2}(\cdot)$ with the proximal operator of the ℓ_∞-norm $\operatorname{prox}_{\frac{\lambda}{\rho}\|\cdot\|_\infty}(\cdot)$, which is given as

$$\operatorname{prox}_{\tau\|\cdot\|_\infty}(\boldsymbol{a}) = \boldsymbol{a} - \tau\Pi_{(\|\cdot\|_1\le 1)}\left(\frac{\boldsymbol{a}}{\tau}\right), \tag{11}$$

where $\Pi_{(\|\cdot\|_1\le 1)}(\cdot)$ denotes the projection on the unit ℓ_1-norm ball. Solving (11) requires iterative methods and it is more computationally expensive compared to computing (10).

The CSSA problem corresponding to (3) can be written as

$$\underset{\{\boldsymbol{X}^{(n)}\}_{n=1}^N}{\text{minimize}} \frac{1}{2}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k^{(n)} - \boldsymbol{s}^{(n)}\right\|_2^2 + \sum_{k=1}^{K}\left(\gamma_1\left\|\boldsymbol{\mathcal{X}}^{(k)}\right\|_{1,1} + \gamma_2\left\|\boldsymbol{\mathcal{X}}^{(k)}\right\|_{2,1}\right). \tag{12}$$

Problem (12) can be addressed using the method in Subsection II-B after modifying the X-update step (9) by replacing $\operatorname{prox}_{\frac{\lambda}{\rho}\|\cdot\|_2}(\cdot)$ with $\operatorname{prox}_{\frac{\gamma_1}{\rho}\|\cdot\|_1 + \frac{\gamma_2}{\rho}\|\cdot\|_2}(\cdot)$, which can be calculated using

$$\operatorname{prox}_{\tau\|\cdot\|_1 + \kappa\|\cdot\|_2}(\boldsymbol{a}) = \operatorname{prox}_{\kappa\|\cdot\|_2}\big(\mathcal{S}_\tau(\boldsymbol{a})\big), \tag{13}$$

where the (elementwise) shrinkage operator $\mathcal{S}_\tau(a) = \operatorname{sign}(a)\max(0,|a|-\tau)$ is a proximal operator of the ℓ_1-norm.

III. Convolutional Dictionary Learning in Simultaneous Sparse Approximation Setup

Given T sets of N dependent input signals and their simultaneous SRs ($\{\boldsymbol{s}^{(t,n)}\}_{n=1}^N$ and $\{\boldsymbol{X}^{(t,n)}\}_{n=1}^N,\ t = 1,\dots,T$), the CDL problem can be formulated as follows

$$\underset{\boldsymbol{D}}{\text{minimize}} \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k^{(t,n)} - \boldsymbol{s}^{(t,n)}\right\|_2^2 \quad \text{s.t.}\quad \boldsymbol{D}_k, \in \boldsymbol{\mathcal{D}},\ k = 1,\dots,K, \tag{14}$$

where $\mathcal{D} = \left\{ \boldsymbol{d} \in \mathbb{R}^Q \mid \|\boldsymbol{d}\|_2 \leq 1 \right\}$. Problem (14) is a standard CDL problem and can be addressed using available batch [11] or online [12] CDL methods. Batch CDL requires all training data to be available at once, while online CDL is useful when the training samples are observed sequentially over time. Online CDL is also more computationally efficient when the total number of training samples (here $T \times N$) is larger than the number of filters in the dictionary (here K) [12].

Convolutional Feature Learning in Multimodal Data

If the input signals are multimodal and the order of modalities is fixed in all T sets of training samples, we can extend the CDL problem (14) to learning multimodal convolutional dictionaries. This can be formulated as

$$\begin{aligned} \underset{\{\boldsymbol{D}^{(n)}\}_{n=1}^{N}}{\text{minimize}} \; & \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} \sum_{n=1}^{N} \Big\| \sum_{k=1}^{K} \boldsymbol{D}_k^{(n)} * \boldsymbol{X}_k^{(t,n)} - \boldsymbol{s}^{(t,n)} \Big\|_2^2 \\ \text{s.t.} \quad & \boldsymbol{D}_k^{(n)}, \in \mathcal{D}, \; k = 1, \ldots, K, \; n = 1, \ldots, N, \end{aligned} \tag{15}$$

which can be addressed as N separate CDL problems. Problem (15) can be interpreted as learning correlated (coupled) features in multimodal data using the corresponding filters in the multimodal dictionaries $\{\boldsymbol{D}^{(n)}\}_{n=1}^{N}$.

IV. NIR-VL Image Fusion based on CSSA

The NIR images are characterized by high contrast resolutions, for example, in capturing vegetation scenes and imaging in low-visibility atmospheric conditions such as fog or haze [16]. Based on these characteristics, the NIR images are used for enhancing outdoor VL images. In this section, we propose a NIR-VL image fusion method based on CSSA and CDL. The CSSA is performed using both ℓ_1 and $\ell_{2,1}$ regularizations and also multimodal dictionaries. The steps of the proposed method for the fusion of a pair of NIR and VL images (denoted as $\boldsymbol{s}_{\rm n}$ and $\boldsymbol{s}_{\rm v}$, respectively) of the same sizes are explained as follows.

Since, the NIR images are presented in greyscale, they can be fused with the intensity components of the VL images which are usually available in the RGB (red-green-blue) format. Hence, in the first step, the VL image is converted to a color space (e.g., YCbCr), where the intensity (greyscale) component, denoted by $\boldsymbol{s}_{\rm v,g}$, is isolated from the color components of the image. Next, $\boldsymbol{s}_{\rm n}$ and $\boldsymbol{s}_{\rm v,g}$ are decomposed into their low-resolution components $\boldsymbol{s}_{\rm n}^{\rm b}$ and $\boldsymbol{s}_{\rm v,g}^{\rm b}$, and high-resolution components $\boldsymbol{s}_{\rm n}^{\rm h}$ and $\boldsymbol{s}_{\rm v,g}^{\rm h}$, for example, using low-pass filtering (more details are given in Subsection V-B).

Using the proposed CSSA method and a pair of pre-learned multimodal NIR-VL dictionaries (denoted as $\boldsymbol{D}^{\rm n}$ and $\boldsymbol{D}^{\rm v}$), the convolutional SRs $\boldsymbol{X}^{\rm n}$ and $\boldsymbol{X}^{\rm v}$ are obtained for $\boldsymbol{s}_{\rm n}^{\rm h}$ and $\boldsymbol{s}_{\rm v,g}^{\rm h}$, respectively. The convolutional SRs are fused using the max-absolute-value fusion rule. This can be formulated as follows

$$\begin{aligned} \boldsymbol{F}_k^{\rm v}(i,j) &= \begin{cases} \boldsymbol{X}_k^{\rm v}(i,j), & \text{if } |\boldsymbol{X}_k^{\rm v}(i,j)| \geq |\boldsymbol{X}_k^{\rm n}(i,j)| \\ 0, & \text{otherwise} \end{cases}, \\ \boldsymbol{F}_k^{\rm n}(i,j) &= \begin{cases} \boldsymbol{X}_k^{\rm n}(i,j), & \text{if } |\boldsymbol{X}_k^{\rm n}(i,j)| > |\boldsymbol{X}_k^{\rm v}(i,j)| \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{16}$$

where $\boldsymbol{F}_k^{\rm n}$ and $\boldsymbol{F}_k^{\rm v}$ are the fused convolutional SRs containing only the most significant representation coefficients at each entry. Moreover, the points (i,j) represent the locations of all pixels in $\boldsymbol{s}_{\rm n}^{\rm h}$ and $\boldsymbol{s}_{\rm v,g}^{\rm h}$, $|\cdot|$ denotes the absolute value of a number, and $k = 1, \ldots, K$ (number of filters in the dictionaries). The fused greyscale high-resolution component $\boldsymbol{s}_{\rm f,g}^{\rm h}$ is then reconstructed using

$$\boldsymbol{s}_{\rm f,g}^{\rm h} = \sum_{k=1}^{K} \boldsymbol{F}_k^{\rm n} * \boldsymbol{D}_k^{\rm n} + \sum_{k=1}^{K} \boldsymbol{F}_k^{\rm v} * \boldsymbol{D}_k^{\rm v}.$$

The fused greyscale image $\boldsymbol{s}_{\rm f,g}$ is formed using $\boldsymbol{s}_{\rm f,g}^{\rm h}$ and the low-resolution component of the VL image

$$\boldsymbol{s}_{\rm f,g} = \boldsymbol{s}_{\rm v,g}^{\rm b} + \boldsymbol{s}_{\rm f,g}^{\rm h}.$$

Finally, the (YCbCr) image with $\boldsymbol{s}_{\rm f,g}$ as the intensity component and the color components of the VL image is converted back to the RGB format to form the fused color image $\boldsymbol{s}_{\rm f}$.

V. Experiments

We first use the proposed CSSA methods with different sparsity structures for sparse approximation of a pair of NIR-VL images and compare the obtained SRs. Next, we use the proposed methods in multifocus and multimodal image fusion tasks and compare the results with existing image fusion methods. The convolutional dictionaries used in the experiments contain 32 filters of size 8×8 and are learned using the online CDL method of [12]. The training data consists of a NIR-VL image dataset and a multifocus image dataset, each containing 10 pairs of images. The NIR-VL and multifocus images are collected from the RGB-NIR Scene dataset [15] and the Lytro dataset [14], respectively. The fusion results are evaluated both visually and based on objective evaluation metrics. Five metrics are used for objective evaluations: average entropy (EN), average peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM) [13], spatial frequency (SF) [18], and edge intensity (EI) [19].

(a) VL image (b) NIR image

Fig. 1: A pair of VL and NIR images.

A. Performance Comparison

We investigate the performances of the proposed CSSA methods in capturing the underlying structures of the NIR-VL images in Fig. 1 in terms of sparsity, the overlap between supports of the SRs, and the residual power. We compare the results also with those obtained using the unstructured CSA

	CSA			CSSA-1			CSSA-2a			CSSA-2b				
λ	Sparsity	Com. supp.	App. err.	Sparsity	Com. supp.	App. err.	Sparsity	Com. supp.	App. err.	Sparsity	Com. supp.	App. err.	γ_1	γ_2
0.001	0.0159	2.92%	4.2326	0.0345	100%	2.5245	0.0209	33.52%	6.5948	0.0248	32.91%	2.8497	0.001	0.001
0.01	0.0094	3.48%	40.2714	0.0183	100%	36.7564	0.0165	87.66%	40.9878	0.0201	87.30%	24.9402	0.001	0.01
0.05	0.0038	4.22%	148.9280	0.0073	100%	137.0919	0.0091	37.61%	72.8397	0.0110	35.31%	51.6590	0.01	0.01
0.1	0.0020	4.34%	241.0831	0.0040	100%	221.8418	0.0040	98.57%	223.7260	0.0045	98.74%	214.8775	0.001	0.1
0.5	0.0001	1.14%	657.2492	0.0004	100%	625.3224	0.0034	87.96%	240.1183	0.0038	87.77%	233.0870	0.01	0.1

TABLE I: Comparison of the convolutional SRs of the multimodal images in Fig. 1 obtained using the (unstructured) CSA method of [11], the proposed CSSA method with $\ell_{2,1}$ regularization (CSSA-1), and the proposed CSSA method with $\ell_{2,1}$ and ℓ_1 regularizations using a single dictionary (CSSA-2a) and two (multimodal) dictionaries (CSSA-2b) in terms of ratio of the nonzero entries (sparsity), the percentage of overlapping nonzero entries (Com. supp.), and the residuals power (App. err.). The convolutional dictionaries used consist of 32 filters of size 8×8 and are learned using a set of 10 pairs of NIR-VL images.

method of [11]. The results obtained using different values of the sparsity regularization parameters are summarized in Table I. As can be seen, the unstructured CSA leads to inconsiderable overlaps between the supports of the convolutional SRs, indicating the fact that CSA with no structure cannot effectively capture the existing correlations between the input images. The CSSA method with $\ell_{2,1}$ regularization (CSSA-1) results in convolutional SRs with identical supports (100% overlap). However, the imposed structure leads to lower sparsity in the SRs and higher approximation errors.

CSSA using ℓ_1 and $\ell_{2,1}$ regularizations (CSSA-2a) allows to relax the identical supports constraint. Specifically, the use of a larger element-sparsity parameter γ_1 allows for a smaller overlap between the supports of the SRs. This approximation model is more appropriate when the correlated input signals can contain (or lack) specific features. For example, in NIR-VL images, some details are visible only in one of the input images. This model can be also extended to learn the nonlinear local relationships in the multimodal data in terms of a set of multimodal dictionaries (CSSA-2b). The results in Table I show that the use of multimodal dictionaries leads to considerably more accurate approximations while achieving SRs with the same level of sparsity compared to the case where a single dictionary is used for both modalities.

B. NIR-VL Image Fusion Results

We benchmark the performance of the proposed NIR-VL image fusion method by comparing our results with those obtained using the fusion method of [16]. There are 51 pairs of outdoor NIR-VL images labeled as "*country*" in the RGB-NIR Scene dataset. We use 10 pairs of these images for CDL, and the remainder 41 images are used as the test dataset. The CSSA is performed using parameters $\rho = 10$, $\gamma_1 = 0.001$ and $\gamma_2 = 0.01$. The lowpass filtering is performed using the *lowpass* function from the SPORCO library [20] with the regularization parameter of 5.

Fig. 2 shows the fusion results for the NIR-VL images in Fig. 1. The average objective evaluation results obtained for the entire test dataset are reported in Table II. As it can be seen in Fig. 2, the proposed fusion method achieves higher contrast resolutions, which is also reflected in larger entropy, spatial frequency, and edge intensity values in Table II. However, method of [16] results in better SSIM and PSNR. This can be explained by the fact that in the proposed method, the fused images are reconstructed from sparse approximations, while the original pixel values are used in [16].

(a) The method of [16]

(b) CSSA

Fig. 2: Visible light and near infrared image fusion results.

C. Multifocus Image Fusion

In this section, we modify the multifocus image fusion method of [17] to incorporate CSSA instead of using unconstrained CSA and compare the resulting performances. The test dataset contains 10 pairs of multifocus images (different from

	Multifocus		NIR-VL	
Metrics	The method of [17]	CSSA	The method of [16]	CSSA
EN	7.4273	**7.4371**	7.0630	**7.2649**
SF	16.6709	**16.8536**	18.1657	**20.0136**
EI	60.6051	**61.9919**	59.7521	**73.0752**
SSIM	0.8491	**0.8498**	**0.7629**	0.7574
PSNR	**27.8952**	27.5893	**20.3470**	19.3590

TABLE II: Average objective evaluation results using different methods. The best results are shown in bold.

the training dataset) and 4 sets of triple multifocus images. The CSSA is performed using only the ℓ_1-norm regularization with parameters $\rho = 10$ and $\lambda = 0.01$. The method of [17] uses the max-ℓ_1-norm rule for fusing the convolutional SRs. In the modified fusion method, we fuse the convolutional SRs (with identical supports) using the elementwise maximum absolute value rule to generate the fused convolutional SRs. All other steps of the two algorithms are identical. The obtained fusion results show that the use of CSSA leads to considerable improvements in terms of higher contrast resolutions and better fusion of multifocus edges (boundaries where one side is in-focus and the other side is out of focus). Fig. 3 shows an example of fusion results obtained using the two methods. The objective evaluation results in Table II also indicate that CSSA improves on the overall performance of the CSA-based multifocus image fusion method of [17] .

(a) Input 1 (b) Input 2 (c) Input 3

(d) The method of [17] (e) CSSA

Fig. 3: Multifocus image fusion results.

VI. Conclusion

Algorithms for convolutional simultaneous sparse approximation with different sparsity structures based on the alternating direction method of multipliers have been proposed. We have evaluated the effectiveness of the proposed methods by using them in two different categories of image fusion problems and compared the obtained results with those of existing image fusion methods. In particular, a novel near infrared and visible light image fusion method based on convolutional simultaneous sparse approximation has been proposed.

References

[1] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572-588, 2006.

[2] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589-602, 2006.

[3] F. Boßmann, S. Krause-Solberg, J. Maly and N. Sissouno, "Structural sparsity in multiple measurements," *IEEE Trans. Signal Process.*, vol. 70, pp. 280-291, 2022.

[4] B. Zheng, C. Zeng, S. Li, and M. G. Liao, "The MMV tail null space property and DOA estimations by tail-$\ell_{2,1}$ minimization," *Signal Process.*, vol. 194, pp. 108450, 2022.

[5] B. Yang, and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10-19, 2012.

[6] F. G. Veshki, N. Ouzir, S. A. Vorobyov, and E. Ollila, "Coupled feature learning for multimodal medical image fusion," *arXiv:2102.08641*, 2021.

[7] J. Li, H. Zhang, L. Zhang and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2523-2533, 2015.

[8] S. H. Fouladi, S. Chiu, B. D. Rao and I. Balasingham, "Recovery of independent sparse sources from linear mixtures using sparse bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6332-6346, 2018.

[9] W. Chen, D. Wipf, Y. Wang, Y. Liu and I. J. Wassell, "Simultaneous bayesian sparse approximation with structured sparse models," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6145-6159, 2016.

[10] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: from theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053-4085, 2011.

[11] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389-393, 2022.

[12] Y. Wang, Q. Yao, J. T. Kwok and L. M. Ni, "Scalable online convolutional sparse coding," *IEEE Trans. Signal Process.*, vol. 27, no. 10, pp. 4850-4859, 2018.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004.

[14] M. Nejati, S. Samavi and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion,* vol. 25, pp. 72-84, 2015.

[15] EPFL RGB-NIR scene dataset. [Online]. Available: https://www.epfl.ch/labs/ivrl/research/downloads/rgb-nir-scene-dataset/. [Accessed: Feb-2022].

[16] M. Herrera-Arellano, H. Peregrina-Barreto and I. Terol-Villalobos, "Visible-NIR image fusion based on top-hat transform," *IEEE Trans. Image Process.*, vol. 30, pp. 4962-4972, 2021.

[17] Y. Liu, X. Chen, R. K. Ward and Z. Jane Wang, "Image fusion With convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882-1886, 2016.

[18] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959-2965, 1995.

[19] B. Rajalingam and R. Priya, "Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis," *Int. J. Eng. Sci.*, vol. 2, pp. 52-60, 2018.

[20] B. Wohlberg, "SParse Optimization Research COde (SPORCO)," Software library available from http://purl.org/brendt/software/sporco, 2017.

Publication VII

F. G. Veshki and S. A. Vorobyov. Efficient Online Convolutional Dictionary Learning Using Approximate Sparse Components. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes island, Greece, June 2023.

EFFICIENT ONLINE CONVOLUTIONAL DICTIONARY LEARNING USING APPROXIMATE SPARSE COMPONENTS

Farshad G. Veshki and Sergiy A. Vorobyov

Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

ABSTRACT

Most available convolutional dictionary learning (CDL) methods use a batch-learning strategy, which consists of alternating optimization of the dictionary and the sparse representations using a training dataset. The computational efficiency of CDL can be improved using an online-learning approach, where the dictionary is optimized incrementally following a sparse approximation of each training sample. However, the existing online CDL (OCDL) methods are still computationally costly when learning large dictionaries. In this paper, we propose an OCDL approach that incorporates decomposed sparse approximations instead of the training samples and substantially improves the computational costs of the existing CDL methods. The resulting optimization problem is addressed using the alternating direction method of multipliers (ADMM).

1. INTRODUCTION

Sparse representations have been widely used in various signal processing and machine learning applications [1–6]. In this model, a signal is approximated using a product of a matrix, called dictionary, and a vector with only a few non-zero entries, i.e., a sparse representation vector. In the context of sparse representations, *dictionary learning* refers to the process of finding a dictionary that leads to sparser representations and more accurate approximations for a large collection of data [7–10].

Convolutional sparse representations (CSRs) provide a shift-invariant model that can be applied to the entire high-dimensional image [11, 12]. In the CSR model, the signal $\boldsymbol{s} \in \mathbb{R}^P$ is approximated using a sum of convolutions of the dictionary filters $\{\boldsymbol{d}_k \in \mathbb{R}^m\}_{k=1}^K$ and convolutional sparse representations $\{\boldsymbol{x}_k \in \mathbb{R}^P\}_{k=1}^K$, i.e., using $\boldsymbol{s} = \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k$, where $*$ is the convolution operator. The convolutional dictionary learning (CDL) problem is commonly addressed using alternating optimization with respect to the CSRs and the dictionary filters using a training dataset [13–17]. This approach is referred to as batch CDL. Optimization of the dictionary filters over a batch of N samples $\{\boldsymbol{s}^n \in \mathbb{R}^P\}_{n=1}^N$ can be formulated as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2N} \sum_{n=1}^{N} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{s}^n \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \quad (1)$$

where $\|\cdot\|_2$ represents the Euclidean norm of a vector and $\boldsymbol{\Omega}(\cdot)$ is the indicator function associated with the constraint on the dictionary filters, that is,

$$\boldsymbol{\Omega}\left(\boldsymbol{d}\right) = \begin{cases} 0, & \text{if} \quad \|\boldsymbol{d}\|_2 \leq 1 \\ \infty, & \text{otherwise} \end{cases}.$$

When the number of training samples N is large, batch CDL becomes extremely computationally expensive. The state-of-the-art batch CDL algorithms have a space (memory) complexity of $\mathcal{O}(KNP)$. The computational efficiency[1] of CDL can be improved using an online-learning approach, where the dictionary is optimized incrementally after sparse approximation of each training sample [18–20]. The online CDL (OCDL) methods are also useful when the data is observed gradually over time. Available OCDL methods have achieved a space complexity of $\mathcal{O}(K^2P)$ which is independent from the number of the data samples. Nevertheless, the computational costs of these methods can be still excessively large when learning large dictionaries or using high-dimensional data samples.

This paper proposes a novel OCDL method that substantially improves the computational efficiency of the state-of-the-art algorithms. The space complexity of the proposed method is of $\mathcal{O}(KP)$. In particular, an approximate sparse components decomposition is used to decentralize the CDL problem with respect to the convolutional filters.

2. BACKGROUNDS

The most efficient solutions to the CDL problem are based on the convolution theorem [16–19]. In the frequency (Fourier) domain, problem (1) can be reformulated as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2NP} \sum_{n=1}^{N} \left\| \sum_{k=1}^{K} \hat{\boldsymbol{d}}_k \odot \hat{\boldsymbol{x}}_k^n - \hat{\boldsymbol{s}}^n \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \quad (2)$$

[1] Computational efficiency accounts for overall algorithmic complexities in terms of both time and space.

where $\hat{(\cdot)}$ denotes the discrete Fourier transform (DFT) and $\odot$ is the elementwise multiplication operator. The filters $\{\boldsymbol{d}_k\}_{k=1}^K$ are zero-padded prior to DFT to be of the same size as the CSRs. By defining $\boldsymbol{\delta}_p \triangleq [\hat{\boldsymbol{d}}_1(p), \cdots, \hat{\boldsymbol{d}}_K(p)]^T$ and $\boldsymbol{\zeta}_p^n \triangleq [\hat{\boldsymbol{x}}_1^n(p), \cdots, \hat{\boldsymbol{x}}_K^n(p)]^T$, problem (2) can be rewritten as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2NP} \sum_{p=1}^{P} \sum_{n=1}^{N} \left\| (\boldsymbol{\zeta}_p^n)^T \boldsymbol{\delta}_p - \hat{\boldsymbol{s}}^n(p) \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k), \quad (3)$$

where $(\cdot)^T$ denotes the non-conjugate transpose operator. Efficient solutions to problem (3) have been proposed based on ADMM and the fast iterative shrinkage-thresholding algorithm (FISTA) [18, 19]. The time and the space complexities of these algorithms are both of $\mathcal{O}(KNP)$.

An online reformulation of problem (3) can be written as [18, 19]

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2} \sum_{p=1}^{P} \boldsymbol{\delta}_p^H \boldsymbol{A}_p^N \boldsymbol{\delta}_p - \sum_{p=1}^{P} \boldsymbol{\delta}_p^T \boldsymbol{b}_p^N + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k), \quad (4)$$

where $(\cdot)^H$ denotes the Hermitian transpose, and the history arrays $\boldsymbol{A}_p^N \in \mathbb{R}^{K \times K}$ and $\boldsymbol{b}_p^N \in \mathbb{R}^K$, $p = 1, \ldots, P$, are defined as $\boldsymbol{A}_p^N \triangleq \frac{1}{NP} \sum_{n=1}^{N} (\boldsymbol{\zeta}_p^n)^* (\boldsymbol{\zeta}_p^n)^T$, $\boldsymbol{b}_p^N \triangleq \frac{1}{NP} \sum_{n=1}^{N} \hat{\boldsymbol{s}}^n(p)^* \boldsymbol{\zeta}_p^n$, with $(\cdot)^*$ representing the elementwise complex-conjugate of an array. The history arrays can be updated incrementally after observing each data sample and its sparse representations. The update rules are written as

$$\begin{aligned} \boldsymbol{A}_p^N &= \frac{1}{NP} (\boldsymbol{\zeta}_p^N)^* (\boldsymbol{\zeta}_p^N)^T + \frac{N-1}{N} \boldsymbol{A}_p^{N-1}, \ p = 1, \ldots, P, \\ \boldsymbol{b}_p^N &= \frac{1}{NP} \hat{\boldsymbol{s}}^N(p)^* \boldsymbol{\zeta}_p^N + \frac{N-1}{N} \boldsymbol{b}_p^{N-1}, \ p = 1, \ldots, P, \end{aligned} \quad (5)$$

where $\boldsymbol{A}_p^0$ and $\boldsymbol{b}_p^0$ are initialized using zero arrays. In OCDL, the dictionary filters are optimized by solving problem (4) once updated history arrays are available. In this way, the space complexity of CDL is reduced[2] to $\mathcal{O}(K^2 P)$. The state-of-the-art OCDL algorithms have a time complexity of $\mathcal{O}(K^2 NP)$ [18, 19].

3. THE PROPOSED METHOD

The proposed method entails compressing the old (already processed once CDL-wise) data samples and their CSRs in a pair of compact history arrays which are used for regularizing the optimization of the dictionary with respect to the new data sample and its CSRs ($\boldsymbol{s}^N$ and $\{\boldsymbol{x}_k^N\}_{k=1}^K$). Hence, the CDL problem (1) is rewritten as

$$\begin{aligned} \underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} & \frac{1}{2N} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^N - \boldsymbol{s}^N \right\|_2^2 \\ & + \frac{1}{2N} \sum_{n=1}^{N-1} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{s}^n \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k). \end{aligned} \quad (6)$$

[2]The efficiency of OCDL is justified only when the number of training samples is larger than that of dictionary filters, i.e., when $N > K$.

In OCDL the CSRs are calculated only once. Thus, the best achievable approximations of the data samples $\{\boldsymbol{s}^n\}_{n=1}^{N-1}$ can be calculated as $\boldsymbol{t}^n = \sum_{k=1}^{K} \boldsymbol{c}_k^n * \boldsymbol{x}_k^n$, $n = 1, \ldots, N-1$, where $\{\boldsymbol{c}_k^n\}_{k=1}^K$ is the optimal dictionary for the single data sample $\boldsymbol{s}^n$ with CSRs $\{\boldsymbol{x}_k^n\}_{k=1}^K$. It can be found by solving

$$\{\boldsymbol{c}_k^n\}_{k=1}^K = \underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\arg\min} \frac{1}{2} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{s}^n \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k). \quad (7)$$

Problem (7) can be efficiently addressed (with complexity of $\mathcal{O}(KP)$) using the existing CDL algorithms. The solution to (6) can be then approximated by replacing the original data samples $\{\boldsymbol{s}^n\}_{n=1}^{N-1}$ with their best achievable approximations $\{\boldsymbol{t}^n\}_{n=1}^{N-1}$. This leads to the following optimization problem

$$\begin{aligned} \underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} & \frac{1}{2N} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^N - \boldsymbol{s}^N \right\|_2^2 \\ & + \frac{1}{2N} \sum_{n=1}^{N-1} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{t}^n \right\|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k). \end{aligned} \quad (8)$$

3.1. CDL Based on Approximate Sparse Components

We further approximate the solution to problem (8) using the following optimization problem

$$\begin{aligned} \underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} & \frac{1}{2N} \left\| \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^N - \boldsymbol{s}^N \right\|_2^2 \\ & + \frac{1}{2N} \sum_{n=1}^{N-1} \sum_{k=1}^{K} \| \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{t}_k^n \|_2^2 + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{d}_k), \end{aligned} \quad (9)$$

where the approximate sparse components (ASCs) are calculated using $\boldsymbol{t}_k^n = \boldsymbol{c}_k^n * \boldsymbol{x}_k^n$. To demonstrate an approximate equivalency between (8) and (9), we need to show that the second quadratic terms in the two problems are approximately equal. Let us denote the approximation residuals of ASCs in (9) as $\boldsymbol{r}_k^n = \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{t}_k^n$. We also denote the approximation residuals of $\{\boldsymbol{t}^n\}_{n=1}^{N-1}$ in (8) as $\boldsymbol{r}^n = \sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{t}^n$. Then, we have

$$\frac{1}{2} \| \boldsymbol{r}^n \|_2^2 = \frac{1}{2} \left\| \sum_{k=1}^{K} \boldsymbol{r}_k^n \right\|_2^2 = \underbrace{\frac{1}{2} \sum_{k=1}^{K} \| \boldsymbol{r}_k^n \|_2^2}_{r_1} + \underbrace{\sum_{k=1}^{K-1} \sum_{l=k+1}^{K} (\boldsymbol{r}_k^n)^T \boldsymbol{r}_l^n}_{r_2}. \quad (10)$$

Since the squared Euclidean norms of the approximation residuals are minimized in (9), we can assume that residuals $\boldsymbol{r}_k^n$ have zero mean Gaussian distributions.[3] Moreover, since the approximation of ASCs is addressed using separate terms with no couplings in the objective function of (9), we can assume that $\{\boldsymbol{r}_k^n\}_{k=1}^K$ are statistically independent. Based on

[3]This is a standard assumption made based on the fact that minimization of squared Euclidean norm of the residuals is equivalent to maximum likelihood estimation when the residuals are assumed to be zero mean Gaussian distributed. See [8, Sec. 3.B], for example.

the aforementioned assumptions, the term r_2 can be disregarded in (10) (because it involves a sum of inner products of uncorrelated zero mean variables). Thus, we can use the following approximation

$$\frac{1}{2}\left\|\boldsymbol{r}^n\right\|_2^2 \simeq \frac{1}{2}\sum_{k=1}^{K}\left\|\boldsymbol{r}_k^n\right\|_2^2. \tag{11}$$

Note that the two sides of the approximation in (11) are the second quadratic terms in (8) and (9). For example, if we assume that the entries of $\boldsymbol{r}_k^n$, $k=1,\ldots,K$, are independent and identically distributed Gaussian random variables with zero mean and variance σ_n^2, then the value of r_1 (see (10)) has a generalized chi-squared distribution with mean $\mu_{r_1} = 2KP\sigma_n^2$ and variance $\sigma_{r_1}^2 = 4KP\sigma_n^4$ (here, note that σ_{r_1} is considerably smaller than μ_{r_1}, which means that r_1 is expected to be centered around its mean value). Based on the same assumption, it can be also shown that the value of r_2 has a Gussian distribution with zero mean and variance $\sigma_{r_2}^2 = K(K-1)P\sigma_n^4/2$. This ensures that the standard deviation of r_2 is drastically smaller than the expected value of r_1, specifically, $\sigma_{r_2} < \mu_{r_1}/2\sqrt{2P}$. This means, for example, for a small image of size 128×128 pixels ($2\sqrt{2P}=362.0387$), that $r_2 < \frac{r_1}{180}$ with a probability larger than 95%.

3.2. Problem Formulation

In this section, we recast problem (9) as an OCDL problem. In the Fourier domain, problem (9) can be reformulated as

$$\begin{aligned}\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ &\frac{1}{2NP}\left\|\sum_{k=1}^{K}\hat{\boldsymbol{d}}_k\odot\hat{\boldsymbol{x}}_k^N-\hat{\boldsymbol{s}}^N\right\|_2^2\\ &+\frac{1}{2NP}\sum_{n=1}^{N-1}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{d}}_k\odot\hat{\boldsymbol{x}}_k^n-\hat{\boldsymbol{t}}_k^n\right\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right),\end{aligned}$$

and then rewritten as

$$\begin{aligned}\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ &\frac{1}{2NP}\sum_{p=1}^{P}\left\|(\boldsymbol{\zeta}_p^N)^T\boldsymbol{\delta}_p-\hat{\boldsymbol{s}}^N(p)\right\|_2^2+\frac{1}{2P}\sum_{p=1}^{P}\boldsymbol{\alpha}_p^{N-1}\odot\boldsymbol{\delta}_p^*\odot\boldsymbol{\delta}_p\\ &-\frac{1}{P}\sum_{p=1}^{P}\boldsymbol{\beta}_p^{N-1}\odot\boldsymbol{\delta}_p+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right),\end{aligned} \tag{12}$$

with history arrays $\boldsymbol{\alpha}_p^N\in\mathbb{R}^K$ and $\boldsymbol{\beta}_p^N\in\mathbb{R}^K$, $p=1,\ldots,P$, defined as

$$\boldsymbol{\alpha}_p^N \triangleq \frac{1}{(N+1)}\sum_{n=1}^{N}\boldsymbol{\zeta}_p^n\odot(\boldsymbol{\zeta}_p^n)^*,\quad \boldsymbol{\beta}_p^N \triangleq \frac{1}{(N+1)}\sum_{n=1}^{N}(\boldsymbol{\tau}_p^n)^*\odot\boldsymbol{\zeta}_p^n,$$

where $\boldsymbol{\tau}_p^n \triangleq [\hat{\boldsymbol{t}}_1^n(p),\cdots,\hat{\boldsymbol{t}}_K^n(p)]^T$. The incremental updates for the history arrays are given by

$$\begin{aligned}\boldsymbol{\alpha}_p^N &= \frac{1}{(N+1)}\boldsymbol{\zeta}_p^N\odot(\boldsymbol{\zeta}_p^N)^*+\frac{N}{N+1}\boldsymbol{\alpha}_p^{N-1},\ p=1,\ldots,P,\\ \boldsymbol{\beta}_p^N &= \frac{1}{(N+1)}(\boldsymbol{\tau}_p^N)^*\odot\boldsymbol{\zeta}_p^N+\frac{N}{N+1}\boldsymbol{\beta}_p^{N-1},\ p=1,\ldots,P.\end{aligned}$$

3.3. Optimization Procedure

We address problem (12) using the ADMM approach. The ADMM formulation of problem (12) can be written as

$$\begin{aligned}\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ &\frac{1}{2NP}\sum_{p=1}^{P}\left\|(\boldsymbol{\zeta}_p^N)^T\boldsymbol{\delta}_p-\hat{\boldsymbol{s}}^N(p)\right\|_2^2\\ &+\frac{1}{2P}\sum_{p=1}^{P}\boldsymbol{\alpha}_p^{N-1}\odot\boldsymbol{\delta}_p^*\odot\boldsymbol{\delta}_p-\frac{1}{P}\sum_{p=1}^{P}\boldsymbol{\beta}_p^{N-1}\odot\boldsymbol{\delta}_p\\ &+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{g}_k\right)\quad\text{s.t.}\quad\boldsymbol{d}_k=\boldsymbol{g}_k,\forall k.\end{aligned}$$

The ADMM iterations are then given as

$$\begin{aligned}\left(\{\boldsymbol{d}_k\}_{k=1}^K\right)^{t+1} =\ &\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\arg\min}\ \frac{1}{2NP}\sum_{p=1}^{P}\left\|(\boldsymbol{\zeta}_p^N)^T\boldsymbol{\delta}_p-\hat{\boldsymbol{s}}^N(p)\right\|_2^2\\ &+\frac{1}{2P}\sum_{p=1}^{P}\boldsymbol{\alpha}_p^{N-1}\odot\boldsymbol{\delta}_p^*\odot\boldsymbol{\delta}_p-\frac{1}{P}\sum_{p=1}^{P}\boldsymbol{\beta}_p^{N-1}\odot\boldsymbol{\delta}_p\\ &+\frac{\rho}{2}\sum_{k=1}^{K}\left\|\boldsymbol{d}_k-(\boldsymbol{g}_k)^t+(\boldsymbol{h}_k)^t\right\|_2^2\end{aligned} \tag{13}$$

$$\begin{aligned}(\{\boldsymbol{g}_k\}_{k=1}^K)^{t+1} =&\underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\arg\min}\ \sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{g}_k\right)\\ &+\frac{\rho}{2}\sum_{k=1}^{K}\left\|(\boldsymbol{d}_k)^{t+1}-\boldsymbol{g}_k+(\boldsymbol{h}_k)^t\right\|_2^2\\ (\boldsymbol{h}_k)^{t+1} =&(\boldsymbol{d}_k)^{t+1}-(\boldsymbol{g}_k)^{t+1}+(\boldsymbol{h}_k)^t,\quad k=1,\ldots,K.\end{aligned} \tag{14}$$

where $\rho>0$ is the penalty parameter and $\{\boldsymbol{h}_k\}_{k=1}^K$ are the scaled Lagrangian variables.

3.3.1. The d-update step

Defining $(\boldsymbol{w}_k)^t \triangleq (\boldsymbol{g}_k)^t-(\boldsymbol{h}_k)^t$, $k=1,\ldots,K$, and $\boldsymbol{\omega}_p \triangleq [\hat{\boldsymbol{w}}_1(p),\cdots,\hat{\boldsymbol{w}}_K(p)]^T$, the solution to update (13) can be obtained by solving P separate problems

$$\begin{aligned}\underset{\boldsymbol{\delta}_p}{\text{minimize}}\ &\frac{1}{2N}\left\|(\boldsymbol{\zeta}_p^N)^T\boldsymbol{\delta}_p-\hat{\boldsymbol{s}}^N(p)\right\|_2^2\\ &+\frac{1}{2}\boldsymbol{\alpha}_p^{N-1}\odot\boldsymbol{\delta}_p^*\odot\boldsymbol{\delta}_p-\boldsymbol{\beta}_p^{N-1}\odot\boldsymbol{\delta}_p+\frac{\rho}{2}\left\|\boldsymbol{\delta}_p-(\boldsymbol{\omega}_p)^t\right\|_2^2.\end{aligned} \tag{15}$$

Based on the Sherman-Morrison formula, (15) can be efficiently solved with complexity of $\mathcal{O}(K)$ as

$$\begin{aligned}(\boldsymbol{\delta}_p)^{t+1} =&\left(\boldsymbol{\gamma}_p-\frac{\boldsymbol{\gamma}_p^{\odot2}\odot|\boldsymbol{\zeta}_p^N|^{\odot2}}{N+\sum_{k=1}^{K}|\boldsymbol{\zeta}_p^N(k)|^2\boldsymbol{\gamma}_p(k)}\right)\\ &\odot\left(\frac{1}{N}(\boldsymbol{\zeta}_p^N)^*\hat{\boldsymbol{s}}^N(p)+(\boldsymbol{\beta}_p^{N-1})^*+\rho(\boldsymbol{\omega}_p)^t\right),\end{aligned} \tag{16}$$

with $\boldsymbol{\gamma}_p=\left(\rho+\boldsymbol{\alpha}_p^{N-1}\right)^{\odot-1}$, where $(\cdot)^{\odot a}$ denotes elementwise exponentiation to the power of a. After finding $\left(\{\boldsymbol{\delta}_p\}_{p=1}^P\right)^{t+1}$, the filters $\left(\{\boldsymbol{d}_k\}_{k=1}^K\right)^{t+1}$ are found using an inverse DFT.

3.3.2. The g-update step

Problem (14) is addressed simply by projecting $(\boldsymbol{d}_k)^{t+1}+(\boldsymbol{h}_k)^t$ on to the unit ball after mapping the entries outside the support to zero (recall that the filters are zero padded).

4. EXPERIMENTAL RESULTS

Compared methods: We compare our algorithm to the following state-of-the-art OCDL methods: the method of [18], which is based on ADMM and uses the iterative Sherman-Morrison formula for updating the history arrays (the ISM method); the frequency-domain-based OCDL method proposed in [19] which is based on FISTA (the FISTA method). We set the maximum number of iterations to 200 in all algorithms. We use convolutional filters of size 8×8 in all experiments. In all algorithms, the CSRs are obtained using the method of [17]. The experiments using the Flowers and SIPI datasets use $\lambda = 0.01$ and $\lambda = 0.1$, respectively. To reduce the statistical dependencies on the initial dictionaries and the order of appearance of the images, experiments are repeated 5 times using different random generator seeds and the average and standard deviation values are reported. All algorithms are implemented using MATLAB. All experiments are conducted on a PC equipped with an Intel(R) Core(TM) i5-8365U 1.60GHz CPU and 16GB memory.

Datasets: the experiments are conducted using two datasets:

- SIPI: 37 random images of size 256×256 taken from the USC-SIPI database [21]. The training and test datasets contain 32 and 5 images, respectively.
- Flowers: 210 images of flowers of size 200×200 taken from Oxford Flower Datasets [22]. The training and test datasets contain 200 and 10 images, respectively.

The original images are converted to grey-scale and resized. Conventionally, the images used for CDL are high-pass filtered [11,16,19]. In our experiments, the low frequency components of all images are removed using the *lowpass* function of the SPORCO toolbox [23] with a regularization parameter of 10.

Comparison criteria: The methods are compared using *peak signal to noise ratio* (PSNR) of the reconstructed images and the average objective functional values (fval). In addition, the evolution of the test functional values using the dictionaries learned by different methods are compared.

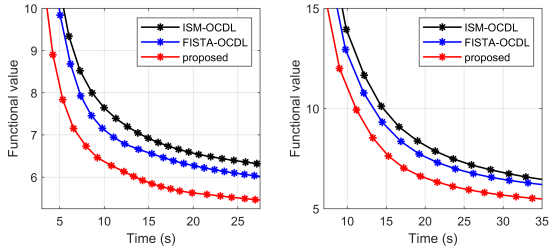
4.1. Performance Comparisons

Table 1 reports the objective functional values and the PSNR results obtained using the methods tested for datasets SIPI and Flowers. As can be seen, the proposed method results in the best performance in terms of both the smallest objective functional values, which shows the effectiveness of the proposed optimization method of solving the CDL problem, and the largest PSNR values, which indicates more accurate reconstructed images. More significant improvements obtained by the proposed method can be observed in our results for larger dataset Flowers, where the proposed method results in significantly shorter training times.

As in [18] and [19], the performances of the learned dictionaries are compared based on their effects on the evolution of objective functional values over the test datasets. The results presented in Fig. 1 show that the proposed method leads to the best performance.

Table 1: The results obtained using the SIPI dataset with $K = 32$, and the Flowers dataset with $K = 64$ and $K = 128$. The best results are shown in bold.

Methods	test fval	test PSNR	training runtime
	SIPI ($K = 32$)		
ISM	96.62 ± 0.12	30.46 ± 0.05	1044 ± 17
FISTA	96.09 ± 0.36	30.23 ± 0.05	2518 ± 157
proposed	$\mathbf{93.18 \pm 0.69}$	$\mathbf{31.10 \pm 0.05}$	$\mathbf{986 \pm 25}$
	Flowers ($K = 64$)		
ISM	5.35 ± 0.02	43.88 ± 0.06	8053 ± 38
FISTA	5.19 ± 0.01	44.44 ± 0.02	7117 ± 635
proposed	$\mathbf{4.71 \pm 0.01}$	$\mathbf{47.57 \pm 0.03}$	$\mathbf{6041 \pm 24}$
	Flowers ($K = 128$)		
ISM	$5.31 \pm +0.01$	42.72 ± 0.03	31302 ± 156
FISTA	5.15 ± 0.03	43.32 ± 0.06	20364 ± 2657
proposed	$\mathbf{4.67 \pm 0.01}$	$\mathbf{45.42 \pm 0.10}$	$\mathbf{14417 \pm 99}$



(a) Flowers ($K = 64$) (b) Flowers ($K = 128$)

Fig. 1: Evolution of *test* functional values for dictionaries learned using different methods.

5. CONCLUSIONS

An efficient online convolutional dictionary learning (OCDL) method has been presented. The proposed method is based on a novel formulation of the CDL problem that incorporates approximate sparse decomposition of training data samples. The proposed formulation assumes that the residuals of the approximate sparse components are statistically independent. The proposed algorithm substantially improves the space and time complexities of the state-of-the-art CDL algorithms. Experimental evaluations demonstrate that the proposed method outperforms the existing OCDL algorithms.

6. REFERENCES

[1] E. Dohmatob, A. Mensch, G. Varoquaux, and B. Thirion, "Learning brain regions via large-scale online structured sparse dictionary learning," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[2] T. Dupré la Tour, T. Moreau, M. Jas, and A. Gramfort, "Multivariate convolutional sparse coding for electromagnetic brain signals," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[3] E. Cheng, K. Chou, S. Rajora, B. Jin, M. Tanveer, C. Lin, K. Young, W. Lin, and M. Prasad, "Deep sparse representation classifier for facial recognition and detection system," *Pattern Recognit. Lett.*, vol. 125, no. 1, pp. 71–77, 2019.

[4] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195–208, 2015.

[5] H. Chang, J. Han, C. Zhong, A. M. Snijders and J. - H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse Coding for biomedical applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1182-1194, 2018.

[6] Z. Wang, X. Cheng, G. Sapiro, and Q. Qiu, "A dictionary approach to domain-invariant learning in deep networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6595–6605, 2020.

[7] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, USA, March. 1999, pp. 2443–2446.

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[9] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. Bach, "Supervised dictionary learning," *Advances in Neural Information Processing Systems*, 2008.

[10] N. Chatterji, and P. L. Bartlett, "Alternating minimization for dictionary learning with random initialization," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: a comparative review and new algorithms," *IEEE Trans. Comput. Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[12] V. Papyan, Y. Romano, M. Elad, and J. Sulam, "Convolutional dictionary learning via local processing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5306–5314.

[13] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, OR, USA, June 2013, pp. 391–398.

[14] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, June. 2015, pp. 5135–5143.

[15] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4290–4298.

[16] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, 2016.

[17] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389-393, 2022.

[18] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Scalable online convolutional sparse coding," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4850–4859, 2018.

[19] J. Liu, C. Garcia-Cardona, B. Wohlberg, and W. Yin, "First-and second-order methods for online convolutional dictionary learning," *SIAM J. Imaging Sci.*, vol. 11, no. 2, pp. 1589–1628, 2018.

[20] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. 1, 2010.

[21] `http://sipi.usc.edu/database/`

[22] `https://www.robots.ox.ac.uk/~vgg/data/flowers/`

[23] B. Wohlberg, "SParse Optimization Research COde (SPORCO)," *Software library available from* `http://purl.org/brendt/software/sporco`, 2017.

Publication VIII

F. G. Veshki and S. A. Vorobyov. An Efficient Approximate Method for Online Convolutional Dictionary Learning. *Submitted for publication*, 2023.

An Efficient Approximate Method for Online Convolutional Dictionary Learning

Farshad G. Veshki and Sergiy A. Vorobyov, *Fellow, IEEE*

***Abstract*—Most existing convolutional dictionary learning (CDL) algorithms are based on batch learning, where the dictionary filters and the convolutional sparse representations are optimized in an alternating manner using a training dataset. When large training datasets are used, batch CDL algorithms become prohibitively memory-intensive. An online-learning technique is used to reduce the memory requirements of CDL by optimizing the dictionary incrementally after finding the sparse representations of each training sample. Nevertheless, learning large dictionaries using the existing online CDL (OCDL) algorithms remains highly computationally expensive. In this paper, we present a novel approximate OCDL method that incorporates sparse decomposition of the training samples. The resulting optimization problems are addressed using the alternating direction method of multipliers. Extensive experimental evaluations using several image datasets show that the proposed method substantially reduces computational costs while preserving the effectiveness of the state-of-the-art OCDL algorithms.**

***Index Terms*—Convolutional sparse coding, online convolutional dictionary learning.**

I. Introduction

SPARSE representations have become increasingly prevalent as a result of their wide use in diverse applications such as signal and image processing, machine learning, and computer vision [1]–[4]. The sparse representation model approximates a signal using a product of a matrix called a dictionary and a vector that only has a few non-zero entries (sparse representation). There are numerous applications where the use of the sparse representation model coupled with a learned dictionary results in remarkably improved performance. A learned dictionary aims to produce sparser representations and more accurate approximations of its domain signals [5]–[7].

Typically, dictionary learning and sparse approximation are used to extract local patterns and features from high-dimensional signals (such as images). Therefore, a prior decomposition of the original signals into vectorized overlapping blocks is usually required (e.g., patch extraction in image processing). However, relations between neighboring blocks are ignored, which results in multi-valued sparse representations and dictionaries composed of similar (shifted) atoms.

Convolutional sparse coding (CSC) provides a single-valued and shift-invariant model for describing high-dimensional signals [8]–[11]. This model replaces the matrix-vector product used in the standard sparse approximation by a sum of convolutions of dictionary filters $\{\boldsymbol{d}_k \in \mathbb{R}^m\}_{k=1}^K$ and convolutional sparse representations (CSRs) $\{\boldsymbol{x}_k \in \mathbb{R}^P\}_{k=1}^K$ (also called sparse feature maps). The convolutional sparse approximation problem can be formulated as follows

$$\underset{\{\boldsymbol{x}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2} \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k - \boldsymbol{s} \right\|_2^2 + \lambda \sum_{k=1}^K \|\boldsymbol{x}_k\|_1, \tag{1}$$

where $\boldsymbol{s} \in \mathbb{R}^P$ is the signal, $\lambda > 0$ is the regularization parameter that controls the sparsity of the representations, $*$ denotes the convolution operator (here, with "*same*" padding), and $\|\cdot\|_1$ and $\|\cdot\|_2$ represent the ℓ_1-norm and the Euclidean norm of a vector, respectively.

The convolutional dictionary learning (CDL) problem is typically addressed using a batch approach in which the sparse representations and the dictionary filters are optimized alternately (batch CDL) [11]–[17]. The following is the formulation of the dictionary optimization problem over a batch of N training signals $\{\boldsymbol{s}^n \in \mathbb{R}^P\}_{n=1}^N$,

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2N} \sum_{n=1}^N \left\| \sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{s}^n \right\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \tag{2}$$

where $\boldsymbol{\Omega}(\cdot)$ represents the indicator function of the constraint set for the dictionary filters, that is,

$$\boldsymbol{\Omega}\left(\boldsymbol{d}\right) = \begin{cases} 0, & \text{if} \quad \|\boldsymbol{d}\|_2 \leq 1 \\ \infty, & \text{otherwise.} \end{cases}$$

The existing batch CDL methods require access to all training signals and their CSRs at once. As a result, memory of the order of NPK is required [18], which can be extremely expensive when using large training datasets, i.e., when $N \gg K$. It is reminded that K is the number of dictionary filters, N is the number of training signals (the batch size), and P is the dimension of the training signals, for example, the number of pixels in an image (usually $P \gg K$ and $P \gg N$). The memory requirement of CDL can be reduced using an online-learning approach, where the dictionary is optimized incrementally after observing each training signal and finding its sparse representations [7]. The online CDL (OCDL) methods are also useful when the training signals are not available all at once, but they are observed gradually over time. The state-of-the-art OCDL methods have achieved memory requirements of the order of K^2P [19], [20], which is independent of the number of training signals. Nevertheless, when learning large dictionaries or using high-dimensional signals, these methods can still incur excessive computational costs.

This paper presents a novel approximate OCDL method that significantly improves the computational efficiency of

The authors are with the Department of Information and Communications Engineering, Aalto University, Espoo, Finland (e-mail: farshad.ghorbaniveshki@aalto.fi; sergiy.vorobyov@aalto.fi)

the state-of-the-art algorithms while providing competitive performance compared to the existing methods. As a result, we propose a method that requires a memory of the order of KP only. More specifically, our method approximates the OCDL problem by minimizing an upper bound of the objective function, where the dictionary optimization problem is decentralized with respect to the convolutional filters. We then solve the resulting optimization problem using the *alternating direction method of multipliers* (ADMM). MATLAB implementations of the proposed algorithms are available at *https://github.com/FarshadGVeshki/Approximate-Online-Convolutional-Dictionary-Learning*.

The rest of the paper is organized as follows. Section II briefly reviews CDL in the Fourier domain. The proposed CDL method and derivation of the algorithms are presented in detail in Section III. Thorough experimental evaluation results in terms of convergence properties and reconstruction accuracy based on multiple image datasets of varying sizes are presented in Section IV. The conclusions are provided in Section V.

II. OCDL in the Fourier Domain

Most efficient CDL methods are based on the Fourier transform [11], [17], [19], [20]. In the frequency (Fourier) domain, problem (2) is equivalent to

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}} \frac{1}{2NP}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\hat{\boldsymbol{d}}_k\odot\hat{\boldsymbol{x}}_k^n-\hat{\boldsymbol{s}}^n\right\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \tag{3}$$

where $\hat{(\cdot)}$ and $\odot$ denote the discrete Fourier transform (DFT) and the elementwise multiplication operator, respectively. The filters $\{\boldsymbol{d}_k\}_{k=1}^K$ are zero-padded prior to DFT, so that $\{\hat{\boldsymbol{d}}_k\}_{k=1}^K$ are of the same size as the CSRs.

Defining $\boldsymbol{\delta}_p \triangleq [\hat{\boldsymbol{d}}_1(p),\cdots,\hat{\boldsymbol{d}}_K(p)]^T$ and $\boldsymbol{\chi}_p^n \triangleq [\hat{\boldsymbol{x}}_1^n(p),\cdots,\hat{\boldsymbol{x}}_K^n(p)]^T$, problem (3) can be rewritten as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\frac{1}{2NP}\sum_{p=1}^{P}\sum_{n=1}^{N}\left\|(\boldsymbol{\chi}_p^n)^T\boldsymbol{\delta}_p-\hat{\boldsymbol{s}}^n(p)\right\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \tag{4}$$

where $(\cdot)^T$ is the transpose operator. The most efficient solutions to problem (4) (the batch CDL problem) have been proposed based on ADMM, and the *fast iterative shrinkage-thresholding algorithm* (FISTA) [17], [18]. The complexities of these algorithms are of $\mathcal{O}(KNP)$ and they require memory of the order of KNP. As a result, when the training dataset is large, batch CDL becomes excessively computationally demanding in practice.

OCDL alleviates the problem of large required memory by storing sufficient statistics of the training signals and their CSRs in compact history arrays. An online reformulation of problem (4) can be written as

$$\underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{minimize}}\ \frac{1}{2}\sum_{p=1}^{P}\boldsymbol{\delta}_p^H\boldsymbol{A}_p^N\boldsymbol{\delta}_p-\sum_{p=1}^{P}\boldsymbol{\delta}_p^T\boldsymbol{b}_p^N+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right), \tag{5}$$

where $(\cdot)^H$ is the Hermitian transpose operator, and the history arrays $\boldsymbol{A}_p^N\in\mathbb{R}^{K\times K}$ and $\boldsymbol{b}_p^N\in\mathbb{R}^K$, $p=1,\dots,P$, are defined as

$$\boldsymbol{A}_p^N\triangleq\frac{1}{NP}\sum_{n=1}^{N}(\boldsymbol{\chi}_p^n)^*(\boldsymbol{\chi}_p^n)^T,\quad \boldsymbol{b}_p^N\triangleq\frac{1}{NP}\sum_{n=1}^{N}\hat{\boldsymbol{s}}^n(p)^*\boldsymbol{\chi}_p^n, \tag{6}$$

with $(\cdot)^*$ standing for the element-wise complex conjugate of an array vector. After observing each training signal and finding its sparse representations, the history arrays are recalculated incrementally using the following formulas

$$\begin{aligned}\boldsymbol{A}_p^N&=\frac{1}{NP}(\boldsymbol{\chi}_p^N)^*(\boldsymbol{\chi}_p^N)^T+\frac{N-1}{N}\boldsymbol{A}_p^{N-1},\ p=1,\dots,P,\\ \boldsymbol{b}_p^N&=\frac{1}{NP}\hat{\boldsymbol{s}}^N(p)^*\boldsymbol{\chi}_p^N+\frac{N-1}{N}\boldsymbol{b}_p^{N-1},\ p=1,\dots,P.\end{aligned} \tag{7}$$

The history arrays are initialized using zero arrays. In OCDL, the dictionary is optimized by solving problem (5) only after the updated history arrays are available. As a result, a memory requirement of K^2P and a complexity of $\mathcal{O}(K^2NP)$ are achieved [19], [20].

III. The Proposed Method

In the proposed method, the training signals are approximated in a distributed manner using N distinct dictionaries $\{\boldsymbol{c}_k^n\in\mathbb{R}^m\}_{k=1}^K$. A fusion of the separately optimized dictionaries based on the respective CSRs is used to calculate the dictionary $\{\boldsymbol{d}_k\}_{k=1}^K$. Specifically, the quadratic term in CDL problem (2) is approximated using the following upper-bound estimate

$$\begin{aligned}&\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k*\boldsymbol{x}_k^n-\boldsymbol{s}^n\right\|_2^2\\ &=\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{d}_k*\boldsymbol{x}_k^n-\sum_{k=1}^{K}\boldsymbol{c}_k^n*\boldsymbol{x}_k^n+\sum_{k=1}^{K}\boldsymbol{c}_k^n*\boldsymbol{x}_k^n-\boldsymbol{s}^n\right\|_2^2\\ &\le\sum_{n=1}^{N}\sum_{k=1}^{K}\|\boldsymbol{d}_k*\boldsymbol{x}_k^n-\boldsymbol{c}_k^n*\boldsymbol{x}_k^n\|_2^2+\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{c}_k^n*\boldsymbol{x}_k^n-\boldsymbol{s}^n\right\|_2^2,\end{aligned} \tag{8}$$

where the inequality is due to the triangle inequality. Accordingly, the proposed approximate CDL problem is formulated as

$$\begin{aligned}\underset{\substack{\{\boldsymbol{d}_k\}_{k=1}^K,\\ \{\{\boldsymbol{c}_k^n\}_{k=1}^K\}_{n=1}^N}}{\text{minimize}}\ &\frac{1}{2N}\sum_{n=1}^{N}\sum_{k=1}^{K}\|\boldsymbol{d}_k*\boldsymbol{x}_k^n-\boldsymbol{c}_k^n*\boldsymbol{x}_k^n\|_2^2+\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right)\\ &+\frac{1}{2N}\sum_{n=1}^{N}\left\|\sum_{k=1}^{K}\boldsymbol{c}_k^n*\boldsymbol{x}_k^n-\boldsymbol{s}^n\right\|_2^2+\sum_{n=1}^{N}\sum_{k=1}^{K}\boldsymbol{\Omega}\left(\boldsymbol{c}_k^n\right).\end{aligned} \tag{9}$$

In the following, two ADMM-based online methods for addressing (9) are presented. The first algorithm uses a standard approach for optimization of $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\boldsymbol{c}_k^N\}_{k=1}^K$, while the second algorithm incorporates pragmatic modifications to the first algorithm to improve the effectiveness of the proposed approximation method and lower computational costs.

A. Algorithm 1

Optimization problem (9) is jointly convex with respect to $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\{\boldsymbol{c}_k^n\}_{k=1}^K\}_{n=1}^N$. Thus, using the OCDL framework, problem (9) can be addressed for the joint optimization variables $\{\boldsymbol{c}_k^N, \boldsymbol{d}_k\}_{k=1}^K$ after observing the Nth training signal $\boldsymbol{s}^N$ and obtaining its CSRs $\{\boldsymbol{x}_k^N\}_{k=1}^K$. Compact history arrays are used to store sufficient statistics of $\{\{\boldsymbol{c}_k^n\}_{k=1}^K\}_{n=1}^{N-1}$ and $\{\{\boldsymbol{x}_k^n\}_{k=1}^K\}_{n=1}^{N-1}$.

The following ADMM formulation is used to solve (9) for $\{\boldsymbol{c}_k^N, \boldsymbol{d}_k\}_{k=1}^K$

$$\begin{aligned}\underset{\substack{\{\boldsymbol{c}_k^N,\boldsymbol{d}_k\}_{k=1}^K,\\ \{\boldsymbol{f}_k^N,\boldsymbol{g}_k\}_{k=1}^K}}{\text{minimize}}\ & \frac{1}{2N}\sum_{n=1}^N\sum_{k=1}^K\|\boldsymbol{g}_k*\boldsymbol{x}_k^n-\boldsymbol{f}_k^n*\boldsymbol{x}_k^n\|_2^2+\sum_{k=1}^K\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right)\\ &+\frac{1}{2N}\sum_{n=1}^N\left\|\sum_{k=1}^K\boldsymbol{f}_k^n*\boldsymbol{x}_k^n-\boldsymbol{s}^n\right\|_2^2+\sum_{n=1}^N\sum_{k=1}^K\boldsymbol{\Omega}\left(\boldsymbol{c}_k^n\right)\\ &\text{s.t.}\quad \boldsymbol{g}_k=\boldsymbol{d}_k,\quad \boldsymbol{f}_k^N=\boldsymbol{c}_k^N,\ k=1,\ldots,K,\end{aligned} \tag{10}$$

where $\{\boldsymbol{f}_k^N, \boldsymbol{g}_k\}_{k=1}^K$ are the (joint) ADMM auxiliary variables. The ADMM iterations consist of the following three steps.

The $\{\boldsymbol{f},\boldsymbol{g}\}$-update step: In this step the auxiliary variables $\{\boldsymbol{f}_k^N, \boldsymbol{g}_k\}_{k=1}^K$ are updated as

$$\begin{aligned}\left(\{\boldsymbol{f}_k^N\}_{k=1}^K\right)^{t+1} &= \underset{\{\boldsymbol{f}_k^N\}_{k=1}^K}{\text{argmin}}\frac{1}{2N}\sum_{k=1}^K\left\|\boldsymbol{f}_k^N*\boldsymbol{x}_k^N-\boldsymbol{z}_k^N\right\|_2^2\\ &+\frac{1}{2N}\left\|\sum_{k=1}^K\boldsymbol{f}_k^N*\boldsymbol{x}_k^N-\boldsymbol{s}^N\right\|_2^2+\frac{\rho}{2}\sum_{k=1}^K\left\|\boldsymbol{f}_k^N-(\boldsymbol{c}_k^N)^t+(\boldsymbol{u}_k)^t\right\|_2^2,\end{aligned} \tag{11}$$

$$\begin{aligned}\left(\{\boldsymbol{g}_k\}_{k=1}^K\right)^{t+1} &= \underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\text{argmin}}\frac{1}{2N}\sum_{n=1}^N\sum_{k=1}^K\|\boldsymbol{g}_k*\boldsymbol{x}_k^n-\boldsymbol{t}_k^n\|_2^2\\ &+\frac{\rho}{2}\sum_{k=1}^K\left\|\boldsymbol{g}_k-(\boldsymbol{d}_k)^t+(\boldsymbol{v}_k)^t\right\|_2^2,\end{aligned} \tag{12}$$

where $\{\boldsymbol{u}_k, \boldsymbol{v}_k\}_{k=1}^K$ are the scaled Lagrangian variables, $\rho > 0$ is the ADMM penalty parameter, $\boldsymbol{z}_k^N \triangleq (\boldsymbol{g}_k)^t * \boldsymbol{x}_k^N$ and $\boldsymbol{t}_k^n \triangleq (\boldsymbol{f}_k^n)^{t+1} * \boldsymbol{x}_k^n$.

The $\{\boldsymbol{c},\boldsymbol{d}\}$-update step: In this step $\{\boldsymbol{c}_k^N, \boldsymbol{d}_k\}_{k=1}^K$ is updated as

$$\begin{aligned}\left(\{\boldsymbol{c}_k^N\}_{k=1}^K\right)^{t+1} &= \underset{\{\boldsymbol{c}_k^N\}_{k=1}^K}{\text{argmin}}\sum_{k=1}^K\boldsymbol{\Omega}\left(\boldsymbol{c}_k^N\right)\\ &+\frac{\rho}{2}\sum_{k=1}^K\left\|(\boldsymbol{f}_k^N)^{t+1}-\boldsymbol{c}_k^N+(\boldsymbol{u}_k)^t\right\|_2^2,\end{aligned} \tag{13}$$

$$\begin{aligned}\left(\{\boldsymbol{d}_k\}_{k=1}^K\right)^{t+1} &= \underset{\{\boldsymbol{d}_k\}_{k=1}^K}{\text{argmin}}\sum_{k=1}^K\boldsymbol{\Omega}\left(\boldsymbol{d}_k\right)\\ &+\frac{\rho}{2}\sum_{k=1}^K\left\|(\boldsymbol{g}_k)^{t+1}-\boldsymbol{d}_k+(\boldsymbol{v}_k)^t\right\|_2^2.\end{aligned} \tag{14}$$

Updating the scaled Lagrangian parameters: Finally, the scaled Lagrangian variables are updates as

$$\begin{aligned}(\boldsymbol{u}_k)^{t+1} &= (\boldsymbol{f}_k^N)^{t+1}-(\boldsymbol{c}_k^N)^{t+1}+(\boldsymbol{u}_k)^t,\quad k=1,\ldots,K,\\ (\boldsymbol{v}_k)^{t+1} &= (\boldsymbol{g}_k)^{t+1}-(\boldsymbol{d}_k)^{t+1}+(\boldsymbol{v}_k)^t,\quad k=1,\ldots,K.\end{aligned} \tag{15}$$

The $\{\boldsymbol{c},\boldsymbol{d}\}$-update step involves projecting $(\boldsymbol{f}_k^N)^{t+1}+(\boldsymbol{u}_k)^t$ (in (13)) and $(\boldsymbol{g}_k)^{t+1}+(\boldsymbol{v}_k)^t$ (in (14)) onto the constraint set. First, the entries outside the support ($\mathbb{R}^m$) are mapped to zero (recall that the filters are zero-padded), followed by projection onto the unit ℓ_2-norm ball.

In the $\{\boldsymbol{f},\boldsymbol{g}\}$-update step, solving problem (11) is equivalent to solving the following optimization problem

$$\begin{aligned}&\underset{\{\boldsymbol{f}_k^N\}_{k=1}^K}{\text{minimize}}\ \frac{1}{2N}\sum_{k=1}^K\left\|\hat{\boldsymbol{f}}_k^N\odot\hat{\boldsymbol{x}}_k^N-\hat{\boldsymbol{z}}_k^N\right\|_2^2\\ &+\frac{1}{2N}\left\|\sum_{k=1}^K\hat{\boldsymbol{f}}_k^N\odot\hat{\boldsymbol{x}}_k^N-\hat{\boldsymbol{s}}^N\right\|_2^2+\frac{\rho}{2}\sum_{k=1}^K\left\|\hat{\boldsymbol{f}}_k^N-\hat{\boldsymbol{q}}_k\right\|_2^2,\end{aligned} \tag{16}$$

where $\boldsymbol{q}_k \triangleq (\boldsymbol{c}_k^N)^t - (\boldsymbol{u}_k)^t$. By equating the derivative of the objective in (16) to zero and using the Sherman-Morrison (SM) formula, the solution to the $\boldsymbol{f}$-update step is found as

$$\begin{aligned}\left(\hat{\boldsymbol{f}}_k^N(p)\right)^{t+1} &= \left(a_p^k+\frac{(a_p^k)^2|\hat{\boldsymbol{x}}_k^N(p)|^2}{1+\sum_{k=1}^K a_p^k|\hat{\boldsymbol{x}}_k^N(p)|^2}\right)\\ &\times\left((\hat{\boldsymbol{x}}_k^N(p))^*\left(\hat{\boldsymbol{z}}_k^N(p)+\hat{\boldsymbol{s}}^N(p)\right)+N\rho\hat{\boldsymbol{q}}_k(p)\right),\end{aligned} \tag{17}$$

where $a_p^k \triangleq (|\hat{\boldsymbol{x}}_k^N(p)|^2+N\rho)^{-1}$. Using precalculated values of $\sum_{k=1}^K a_p^k|\hat{\boldsymbol{x}}_k^N(p)|^2$, the $\boldsymbol{f}$-update step can be carried out with the complexity of $\mathcal{O}(KP)$ using (17).

Problem (12) can be addressed via solving the following optimization problem

$$\underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}}\frac{1}{2N}\sum_{n=1}^N\sum_{k=1}^K\left\|\hat{\boldsymbol{g}}_k\odot\hat{\boldsymbol{x}}_k^n-\hat{\boldsymbol{t}}_k^n\right\|_2^2+\frac{\rho}{2}\sum_{k=1}^K\|\hat{\boldsymbol{g}}_k-\hat{\boldsymbol{w}}_k\|_2^2, \tag{18}$$

where $\boldsymbol{w}_k \triangleq (\boldsymbol{d}_k)^t - (\boldsymbol{v}_k)^t$.

The solution to (18) can be found as

$$(\hat{\boldsymbol{g}}_k(p))^{t+1} = \frac{\boldsymbol{\beta}_k^N(p)+\hat{\boldsymbol{w}}_k(p)}{\boldsymbol{\alpha}_k^N+\rho},\ p=1,\ldots,P,\ k=1,\ldots,K, \tag{19}$$

where history arrays $\boldsymbol{\alpha}_k^N \in \mathbb{R}^P$ and $\boldsymbol{\beta}_k^N \in \mathbb{R}^P$, $k=1,\ldots,K$, are defined as

$$\boldsymbol{\alpha}_k^N \triangleq \frac{1}{N}\sum_{n=1}^N(\hat{\boldsymbol{x}}_k^n)^*\odot\hat{\boldsymbol{x}}_k^n,\ \boldsymbol{\beta}_k^N \triangleq \frac{1}{N}\sum_{n=1}^N(\hat{\boldsymbol{x}}_k^n)^*\odot\hat{\boldsymbol{t}}_k^n. \tag{20}$$

The history arrays are incrementally updated using

$$\boldsymbol{\alpha}_k^N = \frac{N-1}{N}\boldsymbol{\alpha}_k^{N-1}+\frac{1}{N}(\hat{\boldsymbol{x}}_k^N)^*\odot\hat{\boldsymbol{x}}_k^N, \tag{21}$$

$$\boldsymbol{\beta}_k^N = \frac{N-1}{N}\boldsymbol{\beta}_k^{N-1}+\frac{1}{N}(\hat{\boldsymbol{x}}_k^N)^*\odot\hat{\boldsymbol{t}}_k^N. \tag{22}$$

Algorithm 1 summarizes the main steps of the proposed approximate OCDL algorithm detailed in this section. Unit norm Gaussian distributed random arrays can be used as initial

dictionary $\{\boldsymbol{d}_k^0\}_{k=1}^K$. At the first iteration, dictionary $\{\boldsymbol{d}_k\}_{k=1}^K$ can be used to initialize $\{\boldsymbol{c}_k^n\}_{k=1}^K$ and $\{\boldsymbol{g}_k\}_{k=1}^K$. Note that, before each iteration of the ADMM algorithm, $\{\boldsymbol{\beta}_k^n\}_{k=1}^K$ needs to be recalculated using (22) based on the latest values of $\{\boldsymbol{f}_k^n\}_{k=1}^K$.

Algorithm 1 OCDL method proposed in Subsection III-A

Input: Training signals $\{\boldsymbol{s}^n \in \mathbb{R}^P\}_{n=1}^N$, initial dictionary $\{\boldsymbol{d}_k^0 \in \mathbb{R}^m\}_{k=1}^K$, sparsity regularization parameter λ;
Initialisation : History arrays $\boldsymbol{\alpha}_k^0 \in \mathbb{R}^P$ and $\boldsymbol{\beta}_k^0 \in \mathbb{R}^P$, $k = 1, \ldots, K$ as zero arrays, $\{\boldsymbol{d}_k\}_{k=1}^K = \{\boldsymbol{d}_k^0\}_{k=1}^K$;
1: **for** $n = 1$ to N **do**
2: Find $\{\boldsymbol{x}_k^n\}_{k=1}^K$ for $\boldsymbol{s}^n$ using $\{\boldsymbol{d}_k\}_{k=1}^K$ and λ by solving (1);
3: Calculate $\{\boldsymbol{\alpha}_k^n\}_{k=1}^K$ using (21);
4: Optimize $\{\boldsymbol{c}_k^n, \boldsymbol{d}_k\}_{k=1}^K$ using the ADMM-based method in Subsection III-A (recalculate $\{\boldsymbol{\beta}_k^n\}_{k=1}^K$ using (22) in every iteration);
5: **end for**
6: **return** Learned convolutional dictionary $\{\boldsymbol{d}_k\}_{k=1}^K$.

B. Algorithm 2

To improve the performance of the proposed OCDL algorithm, dictionary optimization can be performed *exactly* for the latest observed signal $\boldsymbol{s}^N$, while the proposed approximation method is used for $\{\boldsymbol{s}^n\}_{n=1}^{N-1}$. Thus, the modified approximate CDL problem is now formulated as

$$\begin{aligned}\underset{\substack{\{\boldsymbol{d}_k\}_{k=1}^K,\\ \{\{\boldsymbol{c}_k^n\}_{k=1}^K\}_{n=1}^N}}{\text{minimize}} \; & \frac{1}{2N}\left\|\sum_{k=1}^K \boldsymbol{d}_k * \boldsymbol{x}_k^N - \boldsymbol{s}^N\right\|_2^2 \\ & + \frac{1}{2N}\sum_{n=1}^{N-1}\sum_{k=1}^K \|\boldsymbol{d}_k * \boldsymbol{x}_k^n - \boldsymbol{c}_k^n * \boldsymbol{x}_k^n\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{d}_k\right) \\ & + \frac{1}{2N}\sum_{n=1}^{N-1}\left\|\sum_{k=1}^K \boldsymbol{c}_k^n * \boldsymbol{x}_k^n - \boldsymbol{s}^n\right\|_2^2 + \sum_{n=1}^N\sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{c}_k^n\right). \quad (23)\end{aligned}$$

The alternating procedure for addressing (23) consists of the following steps.

1) Optimization of $\{\boldsymbol{d}_k\}_{k=1}^K$: Solving (23) with respect to $\{\boldsymbol{d}_k\}_{k=1}^K$ can be addressed using the following ADMM formulation

$$\begin{aligned}\underset{\{\boldsymbol{d}_k\}_{k=1}^K, \{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}} \; & \frac{1}{2N}\left\|\sum_{k=1}^K \boldsymbol{g}_k * \boldsymbol{x}_k^N - \boldsymbol{s}^N\right\|_2^2 \\ & + \frac{1}{2N}\sum_{n=1}^{N-1}\sum_{k=1}^K \|\boldsymbol{g}_k * \boldsymbol{x}_k^n - \boldsymbol{r}_k^n\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{d}_k\right) \\ & \text{s.t.} \quad \boldsymbol{g}_k = \boldsymbol{d}_k, \; k = 1, \ldots, K. \quad (24)\end{aligned}$$

where $\boldsymbol{r}_k^n \triangleq \boldsymbol{c}_k^n * \boldsymbol{x}_k^n$.

The ADMM iterations consist of the following steps:

(i) the $\boldsymbol{g}$-update step: a convolutional least-squares fitting problem);

(ii) the $\boldsymbol{d}$-update step: projection on the constraint set (similar to (14));

(iii) updating the Lagrangian multipliers (similar to (15)).

The $\boldsymbol{g}$-update step requires solving the optimization problem in the form of

$$\begin{aligned}\underset{\{\boldsymbol{g}_k\}_{k=1}^K}{\text{minimize}} \; & \frac{1}{2N}\left\|\sum_{k=1}^K \hat{\boldsymbol{g}}_k \odot \hat{\boldsymbol{x}}_k^N - \hat{\boldsymbol{s}}^N\right\|_2^2 \\ & + \frac{1}{2N}\sum_{n=1}^{N-1}\sum_{k=1}^K \|\hat{\boldsymbol{g}}_k \odot \hat{\boldsymbol{x}}_k^n - \hat{\boldsymbol{r}}_k^n\|_2^2 + \frac{\rho}{2}\sum_{k=1}^K \|\hat{\boldsymbol{g}}_k - \hat{\boldsymbol{e}}_k\|_2^2. \quad (25)\end{aligned}$$

Equating the derivative to zero and using the SM formula, optimization problem (25) can be solved as

$$\begin{aligned}\left(\hat{\boldsymbol{g}}_k^N(p)\right)^{t+1} = & \left(b_p^k + \frac{(b_p^k)^2 |\hat{\boldsymbol{x}}_k^N(p)|^2}{N + \sum_{k=1}^K b_p^k |\hat{\boldsymbol{x}}_k^N(p)|^2}\right) \\ & \times \left(\frac{1}{N}(\hat{\boldsymbol{x}}_k^N(p))^* \hat{\boldsymbol{s}}^N(p) + \tilde{\boldsymbol{\beta}}_k^{N-1}(p) + \rho \hat{\boldsymbol{e}}_k(p)\right), \quad (26)\end{aligned}$$

with $b_p^k \triangleq (\tilde{\boldsymbol{\alpha}}_k^{N-1}(p) + \rho)^{-1}$, where history arrays $\tilde{\boldsymbol{\alpha}}_k^N \in \mathbb{R}^P$ and $\tilde{\boldsymbol{\beta}}_k^N \in \mathbb{R}^P$, $k = 1, \ldots, K$, are defined as

$$\tilde{\boldsymbol{\alpha}}_k^N \triangleq \frac{1}{N+1}\sum_{n=1}^N (\hat{\boldsymbol{x}}_k^n)^* \odot \hat{\boldsymbol{x}}_k^n, \; \tilde{\boldsymbol{\beta}}_k^N \triangleq \frac{1}{N+1}\sum_{n=1}^N (\hat{\boldsymbol{x}}_k^n)^* \odot \hat{\boldsymbol{r}}_k^n. \quad (27)$$

The incremental update rules for $\tilde{\boldsymbol{\alpha}}_k^N$ and $\tilde{\boldsymbol{\beta}}_k^N$ can be found as

$$\tilde{\boldsymbol{\alpha}}_k^N = \frac{N}{N+1}\tilde{\boldsymbol{\alpha}}_k^{N-1} + \frac{1}{N+1}(\hat{\boldsymbol{x}}_k^N)^* \odot \hat{\boldsymbol{x}}_k^N, \quad (28)$$

$$\tilde{\boldsymbol{\beta}}_k^N = \frac{N}{N+1}\tilde{\boldsymbol{\beta}}_k^{N-1} + \frac{1}{N+1}(\hat{\boldsymbol{x}}_k^N)^* \odot \hat{\boldsymbol{r}}_k^N. \quad (29)$$

The $\boldsymbol{g}$-update (26) can be performed with the complexity of $\mathcal{O}(KP)$ using precalculated values of $\sum_{k=1}^K b_p^k |\hat{\boldsymbol{x}}_k^N(p)|^2$.

2) Optimization of $\{\boldsymbol{c}_k^N\}_{k=1}^K$: In the modified algorithm, dictionary $\{\boldsymbol{c}_k^N\}_{k=1}^K$ is optimized only to provide a more accurate approximation of $\boldsymbol{s}^N$ (in comparison with the approximation provided using $\{\boldsymbol{d}_k\}_{k=1}^K$). It means that the second quadratic term in (23) is ignored in the step of $\{\boldsymbol{c}_k^N\}_{k=1}^K$ optimization. Here we rely on the fact that CSRs $\{\boldsymbol{x}_k^N\}_{k=1}^K$ are direct products of $\{\boldsymbol{d}_k\}_{k=1}^K$. As a result, considering that the approximation is based on $\{\boldsymbol{x}_k^N\}_{k=1}^K$, the resulting $\{\boldsymbol{c}_k^N\}_{k=1}^K$ cannot unfavorably deviate from $\{\boldsymbol{d}_k\}_{k=1}^K$. Problem (23), which needs to be solved now for $\{\boldsymbol{c}_k^N\}_{k=1}^K$ only, is then reduced to the following optimization problem

$$\underset{\{\boldsymbol{c}_k^N\}_{k=1}^K}{\text{minimize}} \frac{1}{2P}\left\|\sum_{k=1}^K \boldsymbol{c}_k^N * \boldsymbol{x}_k^N - \boldsymbol{s}^N\right\|_2^2 + \sum_{k=1}^K \boldsymbol{\Omega}\left(\boldsymbol{c}_k^N\right), \quad (30)$$

which is a CDL problem involving a single training signal, which can be addressed using the existing CDL methods (e.g., [17]).

The main steps of the presented approximate OCDL algorithm are summarized in Algorithm 2. Optimization of dictionaries $\{\boldsymbol{d}_k\}_{k=1}^K$ and $\{\boldsymbol{c}_k^n\}_{k=1}^K$ (lines 3 and 4) can be initialized using the existing $\{\boldsymbol{d}_k\}_{k=1}^K$.

Algorithm 2 OCDL method proposed in Subsection III-B

Input: Training signals $\{\boldsymbol{s}^n \in \mathbb{R}^P\}_{n=1}^N$, initial dictionary $\{\boldsymbol{d}_k^0 \in \mathbb{R}^m\}_{k=1}^K$, sparsity regularization parameter λ;

Initialisation : History arrays $\tilde{\boldsymbol{\alpha}}_k^0 \in \mathbb{R}^P$ and $\tilde{\boldsymbol{\beta}}_k^0 \in \mathbb{R}^P$, $k = 1, \ldots, K$ as zero arrays, $\{\boldsymbol{d}_k\}_{k=1}^K = \{\boldsymbol{d}_k^0\}_{k=1}^K$;

for $n = 1$ to N **do**
  Find $\{\boldsymbol{x}_k^n\}_{k=1}^K$ for $\boldsymbol{s}^n$ and $\{\boldsymbol{d}_k\}_{k=1}^K$ by solving (1);
  Optimize $\{\boldsymbol{d}_k\}_{k=1}^K$ as in Subsection III-B1;
  Optimize $\{\boldsymbol{c}_k^n\}_{k=1}^K$ as in Subsection III-B2;
  Calculate $\{\tilde{\boldsymbol{\alpha}}_k^n\}_{k=1}^K$ and $\{\tilde{\boldsymbol{\beta}}_k^n\}_{k=1}^K$ using (28), (29);
end for
return learned convolutional dictionary $\{\boldsymbol{d}_k\}_{k=1}^K$.

C. Memory Requirements and Computational Complexity

The largest arrays used in the proposed algorithms are of size KP. The most computationally expensive steps of performing updates (17) and (26) both have a complexity of $\mathcal{O}(KP)$, which is slightly dominated by the complexity of DFT that is of $\mathcal{O}(KP\log(P))$ when performed using *Fast Fourier Transform*. Thus, the computational complexity of the proposed algorithm is of the order of KP sequentially performed N times (once for each signal in the training dataset).

IV. Experimental Results

A. Compared Methods

The performance of the proposed algorithms is benchmarked against the following state-of-the-art OCDL methods:

OCSC The ADMM-based OCDL method of [19], which uses the iterative Sherman-Morrison formula for updating the history arrays;

FISTA The FISTA-based OCDL method of [20] that uses gradient calculated in the Fourier domain.

In addition, we compare the OCDL methods to the following batch-CDL algorithm,

ADMM-cns The batch-CDL method of [17] that is based on consensus-ADMM.

Algorithms 1-2 are referred to as "proposed-1" and "proposed-2", respectively.

B. Datasets

The experiments are conducted using the following 5 image datasets:

Fruit and **City** Two small datasets, each composed of 10 images of size 100×100. These datasets are typically used as benchmarks for CSC and CDL [12], [13], [19];

SIPI A dataset composed of 20 training images and 5 test images all of size 256×256 collected from the UCS-SIPI image database *http://sipi.usc.edu/database/*.

Flicker A dataset composed of 40 training images and 5 test images all of size 256×256 collected from the MIRFLICKR-1M image dataset *https://press.liacs.nl/mirflickr/mirdownload.html*.

Flicker-large A dataset composed of 1000 training images and 50 test images all of size 256×256 collected from the MIRFLICKR-1M image dataset.

The initial images are transformed into greyscale and the 8-bit pixel values are normalized to a range of 0-1 by dividing by 255. Images from the MIRFLICKR-1M and USC-SIPI datasets are then cropped and resized. As the CSC model is not capable of effectively handling low-frequency signals, it is a common practice to use high-pass filtered images for CDL [11], [18], [20]. In the experiments, the low-frequency components of all images are eliminated using the *lowpass* function of the SPORCO toolbox [21] with a regularization parameter of 5.

C. Implementation Details

The proposed algorithms employ the unconstrained convolutional sparse approximation method of [17]. In all ADMM-based algorithms (both sparse approximation and dictionary learning) the maximum number of iterations is set to 300, and stopping criteria discussed in [22, Subsection 3.3] with absolute and relative tolerance values of 10^{-4} are used. We use dictionary filters of size 8×8 in all experiments.

All ADMM-based algorithms except OCSC use ADMM extensions *over-relaxation* [22, Subsection 3.4.3] and *varying penalty parameter* [22, Subsection 3.4.1] with initial penalty parameter $\rho = 10$ (the same parameters are used in all methods). The OCSC method incorporates the ADMM penalty parameter ρ in the history arrays. Thus, this method cannot use *varying penalty parameter* extension. For the OCSC method, we use the default parameters set by the authors of the paper (the stopping criteria are modified to be uniform with other algorithms compared).

In all experiments, we use $\lambda = 0.1\lambda_{\max}$, where $\lambda_{\max}$ is the smallest value that results in all-zero sparse representations and can be obtained using ℓ_∞-norm of the gradient of the objective of convolutional sparse approximation problem (1) at $\{\boldsymbol{x}_k\}_{k=1}^K = \boldsymbol{0}$. Here, the value of $\lambda_{\max}$ is calculated only once using the first image in the training datasets.

All algorithms are implemented using MATLAB. All experiments are performed using a PC equipped with an Intel(R) Core(TM) i5-8365U 1.60GHz CPU and 16GB memory.

D. Comparison Criteria

The effectiveness of the CDL algorithms is typically evaluated based on the objective values of the convolutional sparse approximation problem (1) averaged over the entire test datasets [19], [20], [23]. A lower objective value indicates a better performance. For the small datasets *Fruit* and *City*, since there is no test data, the average training objective values are reported to compare the effectiveness of the optimization algorithms [12]. Using visualized learned dictionary filters, the OCDL algorithms are evaluated for their ability to extract (learn) visual features. The efficiency of the algorithms is measured using the training times.

Fig. 1. Datsets *Fruit* (first row) and *City* (second row).

E. Small Datasets Fruit and City

Fig. 1 shows the images in the small datasets *Fruit* and *City*. Tables I and II report the average training objective values and the training times obtained using the methods tested for these two datasets. To facilitate comparison, the results are presented as bar plots in Fig. 2. The experiments based on datasets *Fruit* and *City* are performed using dictionary size $K = 64$.

TABLE I
AVERAGE TRAINING OBJECTIVE VALUES AND TRAINING TIMES OBTAINED USING THE METHODS COMPARED FOR DATASET *Fruit*.

	Objective	Training Time (s)
Initial dictionary	19.5422	-
FISTA [20]	16.0159	167
OCSC [19]	14.5529	530
Proposed-1	16.6867	39
Proposed-2	14.3059	33
ADMM-cns (batch) [17]	11.8088	122

TABLE II
AVERAGE TRAINING OBJECTIVE VALUES AND TRAINING TIMES OBTAINED USING THE METHODS COMPARED FOR DATASET *City*.

	Objective	Training Time (s)
Initial dictionary	33.9411	-
FISTA [20]	28.6235	190
OCSC [19]	24.5472	462
Proposed-1	30.1463	42
Proposed-2	25.2740	32
ADMM-cns (batch) [17]	18.9411	153

As can be observed, the ADMM-cns batch CDL algorithm yields the lowest objective function values. However, this method is not suitable for large datasets as mentioned earlier. The proposed methods produce objective values that are comparable to other OCDL algorithms tested. In particular, algorithm 2 (proposed-2) results in the smallest objective for the *Fruit* dataset among all OCDL algorithms. For the *City* dataset, the OCSC method has the lowest objective compared to other OCDL methods (slightly better than that of proposed-2), but shows a longer training time. As shown in Tables I and II, the proposed algorithms result in substantially shorter training times, especially Algorithm 2, which is noticeably faster than Algorithm 1.

The convolutional dictionaries learned based on datasets *Fruit* and *City* using the methods tested are visualized in Figs. 3 and 4, respectively.

Acquiring *valid* (as opposed to *noisy* and *random*) visual features is crucial in many image and signal processing tasks

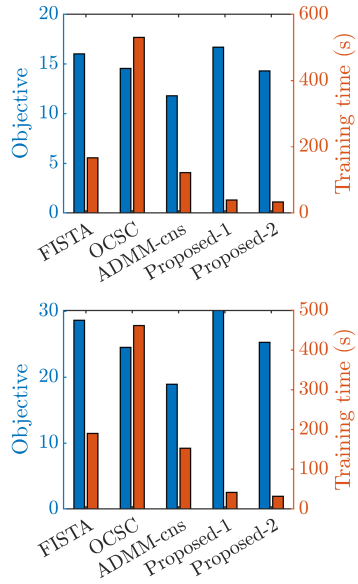


Fig. 2. Comparison of training objective values and training times obtained using all methods compared for datasets *Fruit* (top) and *City* (bottom).

that utilize dictionary learning, such as image denoising, image inpainting, and image fusion. By examining the dictionaries shown in Figs. 3 and 4, it can be seen that the dictionaries learned using the proposed method contain fewer noisy and random filters compared to those learned using OCSC and FISTA. The filters in the dictionaries learned using ADMM-cns (batch CDL) appear *crisper* and *sharper*, while those learned using the proposed algorithms seem *smoother*. This can be explained by the fact that in the proposed method, the dictionaries are, in a way, learned from the sparse approximation of the original images.

F. Datasets SIPI and Flickr

Figs. 5 and 6 depict 10 images randomly selected from the *SIPI* and *Flickr* datasets, respectively. The experiments for *SIPI* dataset are carried out using a dictionary size of $K = 80$. A dictionary size of $K = 100$ is used for the experiments based on *Flickr* dataset. The average test objective values and

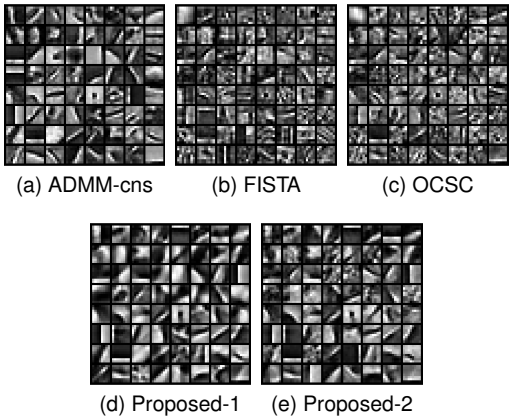


Fig. 3. Dictionaries learned ($K = 64$) using the methods compared for dataset *Fruit*.

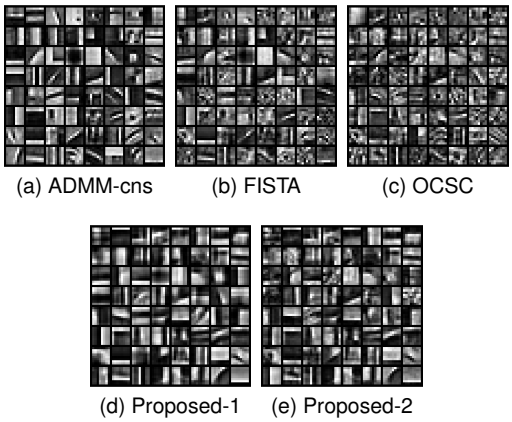


Fig. 4. Dictionaries learned ($K = 64$) using the methods compared for dataset *City*.

the training times obtained using all methods tested for these two datasets are reported in Tables III and IV, and displayed in bar charts in Fig. 7.

TABLE III
AVERAGE TEST OBJECTIVE VALUES AND TRAINING TIMES OBTAINED USING THE METHODS COMPARED FOR DATASET *SIPI*.

	Objective	Training Time (s)
Initial dictionary	103.0952	-
FISTA [20]	63.6088	4904
OCSC [19]	67.2540	5598
Proposed-1	65.0985	685
Proposed-2	63.4867	513
ADMM-cns (batch) [17]	61.1713	2248

As can be seen in Tables III and IV, the ADMM-cns method achieves the lowest test objective values. However, its advantage over the OCDL methods is not as noticeable as in the case of experiments on small datasets *Fruit* and *City*. Specifically, in the experiments on the larger dataset *Flickr*, ADMM-cns performs only slightly better than FISTA and proposed-2, while requiring the longest training time. Among the OCDL methods, FISTA results in the smallest test objective in the experiments on *Flickr*, although it takes the longest training time. The proposed methods result in comparable test objective values to other OCDL methods while substantially shortening the training time. In particular, Algorithm-2 has the smallest objective among all OCDL algorithms for the *SIPI* dataset.

TABLE IV
AVERAGE TEST OBJECTIVE VALUES AND TRAINING TIMES OBTAINED USING THE METHODS COMPARED FOR DATASET *Flickr*.

	Objective	Training Time (s)
Initial dictionary	51.6432	-
FISTA [20]	31.3904	16032
OCSC [19]	35.4325	12689
Proposed-1	32.4064	1362
Proposed-2	31.6799	1102
ADMM-cns (batch) [17]	30.6657	16049

The convolutional dictionaries learned based on datasets *SIPI* and *Flickr* using the methods tested are shown in Figs. 8 and 9, respectively. As can be observed from the dictionaries displayed in Fig. 8, in the experiments on *SIPI*, the dictionary filters learned using the proposed algorithms are less noisy and random compared to those learned using FISTA and OCSC. For the experiment on the *Flicker* dataset, the dictionary filters learned using FISTA are crisper and sharper compared to other OCDL methods tested (FISTA also resulted in the smallest test objective for dataset *Flickr*).

G. Learning Large Dictionaries

In this experiment, we use the proposed algorithms to learn large dictionaries of sizes $K = 200$, $K = 300$, and $K = 400$ based on the *Flickr* dataset. Learning such large dictionaries over the images of the size of those in *Flickr* is not feasible using the OCDL methods, OCSC and FISTA. Indeed, in single precision, for $K = 200$, only the larger history array of these methods, that is of size K^2P, would require *more than* 10 *Gigabytes memory*. The learned large dictionaries are visualized in Fig. 10. It can be seen that all dictionaries learned are mostly composed of visually valid features. The obtained training times are reported in Table V and Fig. 11. As can be seen, the longest training times obtained using the proposed methods are still significantly shorter than those resulting from using other methods tested for learning smaller dictionaries (see Table IV, for example).

TABLE V
TRAINING TIMES (SECONDS) OBTAINED USING THE PROPOSED METHODS FOR DATASET *Flickr*.

	$K = 200$	$K = 300$	$K = 400$
Proposed-1	2691	3695	4574
Proposed-2	2466	3487	4258

H. CDL Over a Large Dataset

In this section, we demonstrate the scalability of the proposed algorithms using the *Flickr-large* dataset (with 1000

Fig. 5. 10 randomly selected images from dataset *SIPI*.

Fig. 6. 10 randomly selected images from dataset *Flickr*.

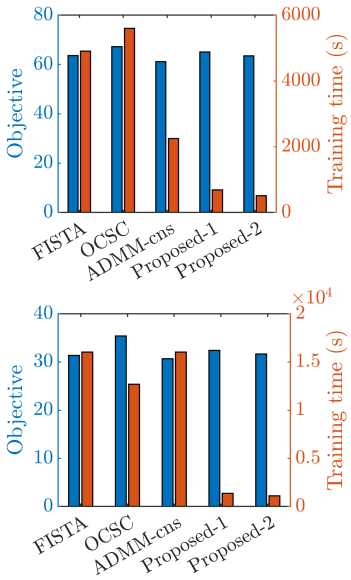


Fig. 7. Comparison of test objective values and training times obtained using all methods compared for datasets *SIPI* (top) and *Flickr* (bottom).

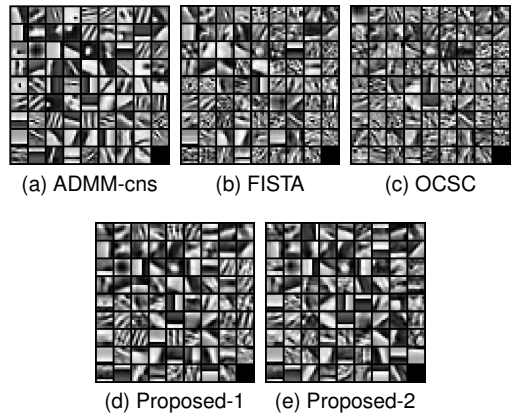


Fig. 8. Dictionaries learned ($K = 80$) using the methods compared for dataset *SIPI*.

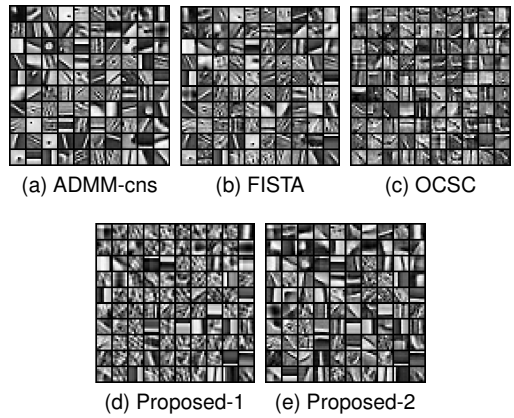


Fig. 9. Dictionaries learned ($K = 100$) using the methods compared for dataset *Flickr*.

training images). Dictionaries composed of $K = 100$ filters are used in this experiment. Fig. 12 shows the average test objective values obtained using the learned dictionaries after processing 1, 10, 100, and 1000 images. The results show that both proposed algorithms are applicable to large training datasets. However, Algorithm-2 leads to considerably lower objective values.

V. Conclusion

An efficient approximate method for CDL has been presented. The proposed method is based on a novel formulation of the CDL problem that incorporates approximate sparse decomposition of training data samples. We have developed two computationally efficient OCDL algorithms based on ADMM to address the proposed approximate CDL problem. The proposed OCDL algorithms substantially reduce the required memory and improve the computational complexities of the state-of-the-art CDL algorithms. Extensive experimental evaluations using multiple image datasets have demonstrated

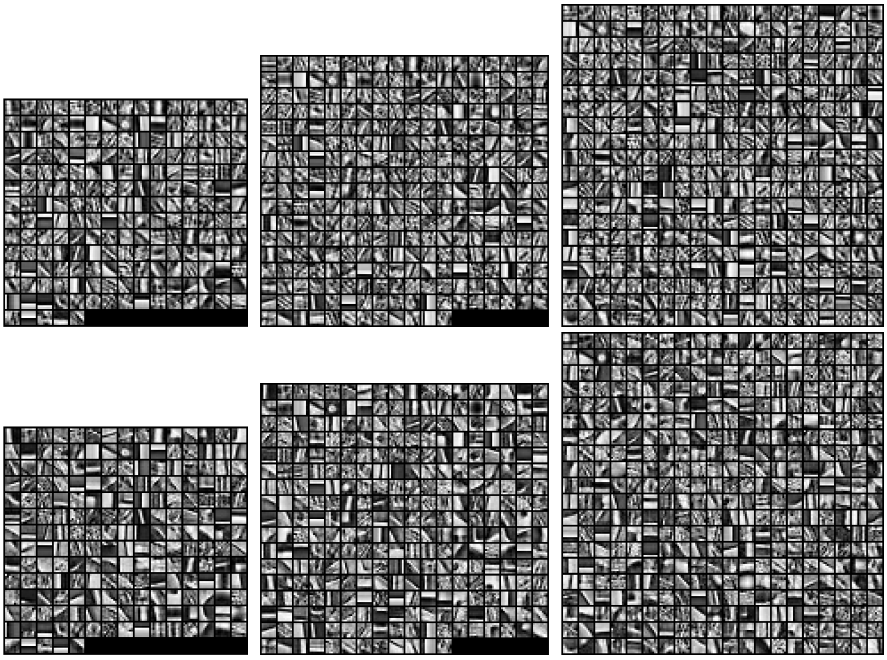


Fig. 10. Large dictionaries learned using the proposed algorithms (top: proposed-1, bottom: proposed-2) for dataset *Flickr* with $K = 200$ (left), $K = 300$ (middle), and $K = 400$ (right).

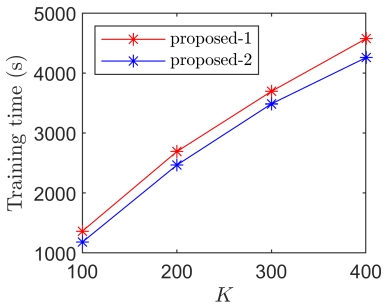


Fig. 11. Comparison of training times obtained using the proposed algorithms and dataset *Flickr* for learning dictionaries of different sizes.

the effectiveness of the proposed OCDL algorithms.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[2] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[3] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[4] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011.

[5] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Process.*, vol. 5, Phoenix, AZ, USA, Mar. 1999, pp. 2443–2446.

[6] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, 2006.

[7] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Mach. Learn.*, Montreal, Quebec, Canada, Jun. 2009, pp. 689–696.

[8] M. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in *Advances in Neural Information Processing Systems*, vol. 11, Dec. 1998, pp. 730–736.

[9] M. Mørup, M. N. Schmidt, and L. K. Hansen, "Shift invariant sparse coding of image and music data," *DTU Informatics, Tech. Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep. IMM2008-04659*, 2008.

[10] V. Papyan, Y. Romano, M. Elad, and J. Sulam, "Convolutional dictionary learning via local processing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Oct. 2017, pp. 5306–5314.

[11] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, 2016.

[12] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 5135–5143.

[13] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 391–398.

[14] B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich, "Consensus convolutional sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Venice, Italy, Oct. 2017, pp. 4290–4298.

[15] R. Chalasani, J. C. Principe, and N. Ramakrishnan, "A fast proximal method for convolutional sparse coding," in *Proc. Int. Jt. Conf. Neural Netw.*, Dallas, TX, USA, Aug. 2013, pp. 1–5.

[16] G. Peng, "Adaptive ADMM for dictionary learning in convolutional

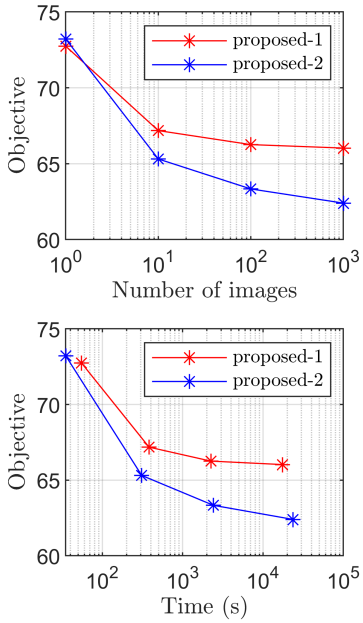


Fig. 12. Results for CDL on *Flickr-large* dataset using the proposed algorithms: average test objective values over the number of processed training images (top) and training time (bottom).

sparse representation," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3408–3422, 2019.

[17] F. G. Veshki and S. A. Vorobyov, "Efficient ADMM-based algorithms for convolutional sparse coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 389–393, 2021.

[18] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[19] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Scalable online convolutional sparse coding," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4850–4859, 2018.

[20] J. Liu, C. Garcia-Cardona, B. Wohlberg, and W. Yin, "First-and second-order methods for online convolutional dictionary learning," *SIAM J. Imaging Sci.*, vol. 11, no. 2, pp. 1589–1628, 2018.

[21] B. Wohlberg, "SParse Optimization Research COde (SPORCO)," Software library available from http://purl.org/brendt/software/sporco, 2017.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foun. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[23] E. Zisselman, J. Sulam, and M. Elad, "A local block coordinate descent algorithm for the CSC model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, Jun. 2019, pp. 8200–8209.

Aalto-DT 70/2023

ISBN 978-952-64-1266-5 (printed)
ISBN 978-952-64-1267-2 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Information and Communications Engineering
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL THESES