Structure Modeling of High Dimensional Data: New Algorithms and Applications

THÈSE N^{O} 6590 (2015)

PRÉSENTÉE LE 26 JUIN 2015

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE TRAITEMENT DES SIGNAUX 2
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mahdad HOSSEINI KAMAL

acceptée sur proposition du jury:

Prof. F. Rachidi-Haeri, président du jury
Prof. P. Vandergheynst, directeur de thèse
Prof. P. Favaro, rapporteur
Prof. S. Süsstrunk, rapporteuse
Prof. G. Wetzstein, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

You can never solve a problem on the level
on which it was created.
— Albert Einstein

Acknowledgements

I am truly honored and fortunate to have been advised by Professor Pierre Vandergheynst. Indeed, this thesis was not possible without his supervision, guidance and his attitude toward independent research. I am deeply thankful to Pierre for the freedom to explore all my ideas and work with amazing researchers.

I would like to thank my thesis committee members: Professors Farhad Rachidi-Haeri, Sabine Süsstrunk, Gordon Wetzstein, and Paolo Favaro for their helpful comments and taking the time to read my thesis. It was my great honor to have them in my thesis committee.

I was really fortunate to work with Professor Paolo Favaro. Paolo is full of innovative ideas. He has a great sense of humor and passion for science. Working with Paolo was an exceptional experience, Thank you Paolo for the flow of innovative ideas.

I profoundly thank Professor Gordon Wetzstein who I had a chance to work with during my visit to Camera Culture Group, Media Lab, MIT. His incredible hard-working attitude is always my greatest source of motivation. I am really indebted to his excellent knowledge in light field photography and his great personality. I will never forget his quote: "Always swim with big fishes and let small fishes be ignorant together".

I am deeply thankful to Professor Ramesh Raskar for accepting me at Camera Culture Group, Media Lab, MIT. He always inspired me with his exceptional team management and it has been a great pleasure to work under his supervision. I would like to thank all the members of Camera Culture specially Barmak for the inspiring discussion on optics and terahertz and I will never forget the days and nights we worked together to bring our camera to existence.

I have been very fortunate to collaborate with incredible scientists, in particular, Hossein Afshari, Alireza Ghasemi, Mohammad Golbabaee, Barmak Heshmat, Nikola Katic, and Mahsa Shoaran for the fruitful collaborations.

I would like to offer my special thanks to the past and present members of the Signal Processing Laboratories specially Ashkan, Benjamin, Christine, Dorina, Gilles, Hossein, Hamed, Jason, Kirell, Luca, Marina, Mohammad, Murat, Rafael, Simon, Tamara, Vassilis, Vijay, Xiaowen for the good memories we share at EPFL and during conferences.

Acknowledgements

I am so grateful to be surrounded by great friends who made my life full of enjoyable moments. Specially, I would like to thank: Amin and Elina, Hamed and Shirin, Reza, Farid, Vahid Majidzadeh, Hamed Amini, Hamed Izadi, Maryam and Mohsen Bahramipanah, Maryam Zargari, Maryna and Sebastian, Mostafa, Momo and Alice, Sohrab, Vahid and Azadeh for their support and friendship.

I want to express my greatest gratitude to Naghmeh for her true friendship, support and care.

Finally, I profoundly thank my mother, Maryam, my father, Hassan, and my sisters Anoosh, Farnaz, and Fereshteh for their unconditional love, endless support and guidance in all stages of my life. They always encourage me to confront challenges and dare to make impossible possible. This thesis is dedicated with love to them.

Lausanne, 11 April 2015 Mahdad Hosseini Kamal

Abstract

The digitization of our most common appliances has led to a literal data deluge, sometimes referred to as Big Data. The ever increasing volume of data we generate, coupled with our desire to exploit it ever faster, forces us to come up with innovative data processing techniques. Interestingly, the information we often look for has a very specific structure that distinguishes it from pure clutter. In this thesis, we explore the use of structured representations to propose new sensing techniques that severely limit the data throughput necessary to recover meaningful information.

In particular, we exploit the intrinsic low-dimensionality of light field videos using tensor low-rank and sparse constraints to recover light field views from a single coded image per video frame. As opposed to conventional methods, our scheme neither alters the spatial resolution for angular resolution nor requires computationally extensive learning stage but rather depends on the intrinsic structures of light fields.

In the second part of this thesis, we propose a novel algorithm to estimate depth from light fields. This method is based on representation of each patch in a light field view as a linear combination of patches from other views for a set of depth hypotheses. The structure in this representation is deployed to estimate accurate depth values.

Finally, we introduce a low-power multi-channel cortical signal acquisition based on compressive sampling theory as an alternative to Nyquist-Shannon sampling theorem. Our scheme exploits the strong correlations between cortical signals to recover neural signals from a few compressive measurements.

Key words: structured sparsity, affine rank minimization, inverse problems, compressed sensing, computation photography, light fields, depth estimation, cortical signals

چکیده

زندگی در عصر اطلاعات، بسیاری از جنبه‌های زندگی انسان را به واسطه‌ی سیل گسترده‌ی داده‌ها را تحت تاثیر قرار داده‌است. افزایش بی‌شمار داده‌ها استفاده از ابزار نو برای مقابله با چالش‌های این روزگار را برجسته می‌کند. در این ابزار ساختارهای میان متغیرهای داده‌ها نقش مهمتری را نسبت به حجم گسترده‌ی داده‌ها در طراحی الگوریتم‌های کارآمد ایفا می‌کنند. افزون بر گسترش ابزارهای کارآمد، سیل داده‌ها موجب ایجاد محدودیت در جمع‌آوری داده‌ها شده‌است که نیاز به بررسی روش‌های نوین جمع‌آوری داده را بیش از پیش پررنگ می‌کند.

این پایان‌نامه به بررسی راهکارهای نوین برای استفاده از پیوندهای موجود بین متغیرهای داده در ساختمان الگوریتم‌ها بازیابی اطلاعات و الگوهای نو برای کاهش نرخ نمونه‌برداری می‌پردازد. به ویژه این پایان‌نامه با بهره‌گیری از ذات کم‌بعد ویدیوهای میدان نوری به بازیابی تمامی زاویه‌های دید ویدیوی میدان نوری از یک تصویر کد شده در ازای هر فریم برپایه‌ی تقسیم تانسور میدان نوری به یک تانسور کم‌مرتبه و یک تانسور تنک می‌پردازد. برخلاف روش‌های مرسوم طرح ارائه شده تغییری در وضوح فضایی و زاویه‌ای میدان نوری ایجاد نمی‌کند. هم‌چنین این روش برپایه‌ی ساختار ذاتی میدان نوری بنا بوده و در نتیجه بی‌نیاز از روش‌های یادگیری ساختار تصویر براساس محاسبات پیچیده‌است.

در بخش دوم این پایان‌نامه یک الگوریتم نوین برای تخمین عمق تصویر میدان نوری ارائه شده‌است. این روش هر پاره از یک زاویه‌ی دید مشخص از تصویر میدان نوری را براساس ترکیب خطی از پاره‌های دیگر زاویه‌های دید میدان نوری برای یک مجموعه از عمق‌های فرضی مختلف بیان می‌کند. این گونه بیان هر پاره برمبنای پاره‌های دیگر، بسیار ساختاریافته است که می‌توان با استفاده از تنکی ساختارگونه عمق تصویر دقیق هر پاره را تخمین زد.

در قسمت آخر این پایان‌نامه، به معرفی یک سامانه کم‌قدرت نمونه‌برداری از سیگنال‌های چند مجرایی غشای مغز براساس حسگری فشرده به عنوان یک جانشین کارساز برای نمونه‌برداری نایکوییست‌ـشانون می‌پردازیم. در طرح معرفی شده شباهت بین سیگنال‌های مغزی توسط تنکی ساختارگونه شکل دهی می شود که نقش به سزای برای بازیابی این سیگنال‌ها از نمونه‌های فشرده شده ایفا می‌کند.

کلید واژه‌ها: تنکی ساختارگونه، کمینه‌سازی مرتبه، مسائل معکوس، حسگری فشرده، عکس‌برداری رایانشی، عکس برداری میدان نوری، سنجش عمق تصویر، سیگنال‌های غشای مغز

Résumé

La digitalisation des technologies de notre quotidien mène à un véritable déluge de données. Ainsi le volume toujours croissant de ces données et notre désir de pouvoir les exploiter toujours plus vite, nous pousse à inventer de nouvelles méthodes de traitement de l'information toujours plus efficaces. Par chance, l'information qui nous intéresse possède souvent une structure très particulière et qui la distingue du désordre. Dans cette thèse, nous explorons l'utilisation de représentations structurées et proposons de nouvelles modalités de capture de signaux limitant fortement la quantité de données nécessaires pour remonter à l'information pertinente.

En particulier, nous exploitons la faible dimension intrinsèque de vidéos du champ lumineux au moyen d'un modèle tensoriel combinant des contraintes de rang faible et de parcimonie et montrons qu'il est possible de reconstituer le champ lumineux à partir d'une seule image codée par trame. Par rapport aux méthodes conventionnelles, notre technique ne souffre pas de dégradation de la résolution spatiale ou angulaire et ne nécessite aucune étape de pré-calcul.

Nous proposons également un nouvel algorithme d'estimation de la profondeur d'une scène à partir de mesures du champ lumineux. Cette méthode se base sur une représentation très structurée du champ lumineux que nous exploitons au moyen de contraintes de parcimonie structurée.

Finalement, nous développons une méthode de mesure de signaux d'activité corticale basée sur l'échantillonnage compressif en vue de réaliser un système à basse consommation d'énergie. La technique que nous proposons utilise la structure particulière des corrélations entre canaux dans le système au moyen d'un modèle parcimonieux structuré par groupes.

Mots clefs : parcimonie structurée, minimisation de rang faible, problème inverse, échantillonnage compressif, photographie computationnelle, champ lumineux, estimation de profondeur, signaux corticaux

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Due to the technological advancements of today, there has been a rapid increase in accumulation of massive amount of data. The data flood in areas like image/signal processing, e-commerce, information retrieval and business has led to modern algorithms which deploy diverse information to address complex problems. Besides the fundamental interest in the ability to collect high-dimensional data, the insights gained from the data plays more important role. Therefore, developing efficient computational methods based on data structures is necessary to extract useful information from data.

Designing feasible algorithms to exploit data structures needs handling data with as many as billion variables. For example, the plenoptic function [2] was introduced as a ray-based model of all visual information: spatial, angular, and temporal light variation and the spectrum. The amount of data produced by plenoptic samples makes this problem extremely difficult. For instance, in a light field video recorded at 30 Hz, with 11.1 Megapixel resolution, 10×10 angular samples, and three color channels—about 100 billion light rays ($\sim$ 100 GB of raw data) have to be recorded, processed, and stored per second. In functional Magnetic Resonance Images (fMRI), each image contains more than 50000 voxels which produces thousands of terabytes of data during each acquisition. In addition to image and video processing applications, massive amount of high-dimensional data is also being gathered from social media, and web relevant data. In such application domains, data routinely lie in thousands to even billions of dimensions, with a number of samples sometimes of the same order of magnitude.

However, high-dimensional data are highly redundant. Due to redundancy, the information volume is much smaller than the data volume. Tremendous amount of work has been conducted over past decade to develop methods that exploit the extensive data redundancy. Many successful methods take advantage of the fact that most correlated structures in data have *sparse representation* in a suitable orthogonal basis or dictionary. For example, JPEG encodes few important coefficients of discrete cosine transform.

Although sparsity intends to represent data structures, this model cannot consider the relation between the sparse coefficients. For example, wavelet transform not only sparsely represents images but also the wavelet coefficients have strong relations with each other [11]. Recently, more advance sparse coding techniques known as *structured sparsity* [7, 74, 89, 172] have been introduced to leverage both sparsity and the correlations between the data variables.

Furthermore, the large number of redundant samples in high-dimensional data gives rise to *intrinsic low-dimensionality* of high-dimensional data. For example, recommender systems have huge data volume that often replete with missing entries. Many techniques such as matrix completion [25, 85] have been introduced to leverage the intrinsic low-dimensionality of the data to fill the missing ratings.

In addition to the potential that high-dimensional data bring, the huge flow of data requires more efficient sampling techniques and poses limitations on the existing technology. For example, the high data rate of wireless implantable neural recording systems results in unacceptable transmission power which can pass the limit of safety concerns. The *compressive sensing* paradigm [9, 23, 41] leverages data structures to greatly reduce the sampling rate, while preserving the overall data quality. In the case of neural implants, for instance, the compressive sensing can provide a low-power acquisition system to respect the safety conditions [31, 142].

In short, the richness of high-dimensional data offers the potential to improve the algorithms, however the new models require to exploit the data structures to effectively address the challenges in high-dimensional data.

1.0.1 Outline

In the following, we present a brief summery of each chapter.

From Sparsity to Structure Modeling

In Chapter 2, we extend data structures beyond sparsity and introduce structured sparsity norms to leverage correlations between data variables. In addition, we explain a set of tools to exploit the intrinsic low-dimensionality of data.

Computational Light Field Imaging

In Chapter 3, we introduce computational camera as a mean to exceed the limitation of analogue photography. Furthermore, we explain a set of modifications either on lens or sensor of cameras to build computational light field cameras. We observe that the available computational light field cameras either sacrifice spatial resolution, use multiple

devices/images or an extensive dictionary learning phase to acquire angular resolution.

Tensor Low-rank and Sparse Light Field Photography

High-quality light field photography has been one of the most difficult challenges in computational photography. Combining coded image acquisition and compressive reconstruction is one of the most promising directions to overcome limitations of conventional light field cameras. In Chapter 4, we present a new approach to compressive light field photography that optically codes light field views into a single camera sensor and exploits a joint tensor low-rank and sparse prior (LRSP) on natural light fields. As opposed to available light field acquisition models, our method does not require a computationally expensive learning stage but rather models the intrinsic redundancies of high dimensional visual signals using a tensor low-rank prior. This is not only computationally more efficient but also more flexible with respect to camera parameters such as aperture sizes and baseline.

A Convex Solution to Disparity Estimation from Light Fields

In Chapter 5, we present a novel convex approach to the reconstruction of depth from light fields. Our method exploits the similarity between patches from light field views to estimate disparity. The proposed scheme looks for the best representation of each patch as a linear combination of other patches for a set of depth candidates such that each depth hypothesis is either chosen or discarded. To achieve this, we model the structure raised in the patch representation via group sparsity. We keep numerical complexity at bay by restricting the space of solutions and by exploiting an efficient Primal-Dual algorithm [124]. Our formulation recovers accurate depth values even for specular surfaces and shows promising performance.

Low-Power Compressive Multi-channel Cortical Recording

We use sparse structured representations for a different problem in Chapter 6 and propose a power-efficient approach for wireless monitoring of brain activity based on compressive sampling. We show that high-dimensional multi-channel cortical signals can be efficiently sampled by a smaller number of linear measurements than dictated by the Nyquist sampling theorem. Our scheme exploits the strong correlations of cortical signals in Gabor transform to recover the multi-channel neural signals from the compressive measurements. Leveraging the group structure of the Gabor coefficients results in more accurate signal recovery in contrast to sparse recovery which does not consider the dependency between the Gabor coefficients.

1.0.2 Main Contributions

Considering the challenges and potentials in high-dimensional data, the premises of this thesis can be summarized as: 1) the algorithms concentrate on data structures. 2) the degree of freedom of data volume is much smaller than data dimension. 3) the acquisition systems should avoid sampling the flow of redundant information. The summary of the main contributions of this thesis is as given.

Tensor Low-rank and Sparse Light Field Photography

- We present a computational camera system that facilitates efficient acquisition of light field image and video.
- We introduce a mathematical framework that models intrinsic low-dimensionality of light fields using tensor low-rank and sparse priors. We show that this model captures redundancy in the high-dimensional plenoptic function well and allows for new optical setups to be derived.
- We design and implement the prototype of a compressive light field camera that is evaluated by both simulation and experiments.
- In addition to the proposed camera prototype, we show that the proposed tensor low-rank and sparse model is universal with respect to baseline and number of views.

Light Field Disparity Estimation

- We present a novel model for light field disparity estimation to represent a light field image patch as a linear combination of other light field patches. This representation satisfies a group sparse model and depends only on a group of light field patches of the same disparity.
- Occlusions are handled uniformly in our framework as a sparse component and this brings more robustness than in traditional matching methods.
- We introduce a robust and globally optimal solution for light field patch matching based on a preconditioned primal-dual algorithm [124], which allows to match a light field patch in all the views to estimate the disparity map.

Low-Power Compressive Multi-channel Cortical Recording

- We design and prototype a novel compressive iEEG recording system with sampling rate far below the Nyquist rate based on correlations between neural channels.

- Our scheme is highly power efficient and significantly reduces the area overhead resulting in an improved power-area product.

Chapter 2

From Sparsity to Structure Modeling

Sparsity is a key concept in many scientific domain. Sparse approximations have gained a lot of attention in various signal and image processing problems such as deconvolution, compression, image deblurring and denoising. Sparsity has become so appealing since for most signal classes there is a sparse representation in a suitable basis or dictionary. That is they can be represented by a small number of coefficients while the remainder is either zero or negligible.

Clearly sparsity is linked to compression, e.g. JPEG compressed images are obtained by thresholding the coefficients of Discrete Cosine Transform (DCT) of each image block of 8×8 to keep a few important coefficients. Except from a compression tool, the introduction of sparsity in signal processing [32] resulted in extensive usage of sparse linear models in various inverse problems.

In a sparse model, each coefficient of signal is processed independent of other coefficients. Although this approach results in low-complexity algorithms, the sparse representation ignores the relations between coefficients. For example, JPEG2000 exploits not only sparsity of wavelet coefficients but also the fact that the location of large wavelet coefficients have a specific structure. Therefore, coding the coefficients according to their structures allows the algorithm to achieve higher compression in compare with a naive coder that deals with each coefficients independently. In brain imaging based on functional Magnetic Resonance (fMRI) or magnetoencephalography (MEG), a set of voxels that represents a brain activity has small localized spatio-temporal activation patterns (e.g. see [62] and references therein). Similarly, the time-frequency representation of audio signals in Modulated Discrete Cosine Transform (MDCT) arranges the significant coefficients in a specific order [89].

These problems along with many others motivate the need of efficient prior constraints to incorporate signal patterns as a the prior knowledge beyond the sparse assumption. In this chapter, we introduce a family of norms that can model a large variety of signal patterns.

2.1 Mathematical Framework

2.1.1 Notation

We denote vectors with bold lower case letters, matrices with bold upper case ones and tensors with italic letters. For any integer i in the set $[[1;n]] = \{1,\dots,n\}$, the i^{th} element of a vector $\mathbf{x} \in \mathbb{R}^n$ is denoted by $\mathbf{x}_i$. Similarly, the entry of a matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ on the i-th row and j^{th} column is denoted as $\mathbf{X}_{ij}$.

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$ be two n-dimensional vectors. Then the inner product between $\mathbf{x}$ and $\mathbf{y}$ is defined as $<\mathbf{x},\mathbf{y}> \triangleq \mathbf{x}^\top\mathbf{y}$, where $\cdot^\top$ is the transpose operator. The squared ℓ_2 norm of a vector is defined as $\|\mathbf{x}\|_2^2 \triangleq <\mathbf{x},\mathbf{x}>$. More generally, the ℓ_q norm of a vector is defined

$$\begin{aligned} \|\mathbf{x}\|_q^q &\triangleq \sum_{i=1}^{n} |\mathbf{x}_i|^q, \quad \forall q \in [1,\infty), \\ \|\mathbf{x}\|_\infty &\triangleq \max_{i\in[[1;n]]} |\mathbf{w}_j|. \end{aligned} \tag{2.1}$$

The support of $\mathbf{x}$ is denoted by $\text{supp}(\mathbf{x}) \triangleq \{i : \mathbf{x}_i \neq 0, 1 \leqslant i \leqslant n\}$. Consequently, the $\#\text{supp}(\mathbf{x})$ is the number of non-zero entries of $\mathbf{x}$, which is known as ℓ_0 pseudo-norm, i.e.

$$\|\mathbf{x}\|_0 \triangleq \#\text{supp}(\mathbf{x}).$$

For matrices $\mathbf{X},\mathbf{Y} \in \mathbb{R}^{n\times m}$, the inner product is defined as $<\mathbf{X},\mathbf{Y}> \triangleq \text{Tr}(\mathbf{X}^\top\mathbf{Y}) \triangleq \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{X}_{ij}\mathbf{Y}_{ij}$. The rank of a matrix (denoted by r) is the number of non-zero singular values of the matrix. The associated norm to the inner product is called the Frobenius norm, which is defined as

$$\|\mathbf{X}\|_F^2 \triangleq \text{Tr}(\mathbf{X}^\top\mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{X}_{ij}^2 = \sum_{i=1}^{r}\sigma_i^2, \tag{2.2}$$

where σ_i is the i-th singular value of $\mathbf{X}$. The operator norm of a matrix is equal to its largest singular value [1]

$$\|\mathbf{X}\| \triangleq \sigma_1. \tag{2.3}$$

[1] σ_1 is the biggest singular value of a matrix.

The nuclear norm of a matrix is equal to the sum of its singular values

$$\|\mathbf{X}\|_* \triangleq \sum_{i=1}^{r} \sigma_i. \tag{2.4}$$

Since singular values are all positive the nuclear norm is equal to ℓ_1 norm of the vector of singular values.

2.1.2 Linear Inverse Problems

Let assume $\mathbf{y} \in \mathbb{R}^m$ is an m-dimensional signal obtained from the original signal $\mathbf{x}$ through a bounded linear operator $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ and corrupted by some additive noise $\mathbf{n}$, i.e.

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{n}. \tag{2.5}$$

This problem casts as a linear inverse problem, where the operator $\mathcal{A}$ and $\mathbf{y}$ are known and the aim is to estimate the unknown signal/image $\mathbf{x}$. For example, in deconvolution problem $\mathbf{y}$ represents the blurred image, $\mathbf{x}$, whose size is the same as $\mathbf{y}$ (i.e. $m = n$), is the sharp image, and the operator $\mathcal{A}$ is the blur kernel. The estimation of latent sharp image $\mathbf{x}$ from the blurred image $\mathbf{y}$ is known as *image deblurring*. *Compressive Sensing* (CS) acquisition recovers the signal $\mathbf{x}$ form m linear measurements with $m \ll n$ [23]. In this case the operator $\mathcal{A}$ is the sampling operator generated from a random distribution. The goal of this problem is to recover the original signal $\mathbf{x}$ by solving the equation (2.5). Unfortunately, this task seems to be impossible as the problem is ill-posed. However, given some conditions on $\mathcal{A}$, it is possible to recover $\mathbf{x}$ from $\mathbf{y}$ using some prior knowledge on $\mathbf{x}$. The goal of priors is to bias problems towards desired solutions known in advance. Thus priors are able to avoid overfitting or transfer an ill-posed problem to a well-posed problem. One well-known prior is the Tikhonov prior [152] that is to improve the conditioning of problems. Sparsity of wavelet coefficients and gradient values, known as Total Variation (TV), [104, 134] is another popular prior that promotes piecewise smooth solution. The classical regularization reads

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}}{\operatorname{argmin}} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 + \lambda f(\mathbf{x}), \tag{2.6}$$

where f(·) is some penalty term on $\mathbf{x}$ depending on the choice of structures to be imposed on $\mathbf{x}$. λ is a weight to control the influence of each term which should be set with respect to noise level. Note that depending on the statistical behavior of noise other fidelity penalties replace the ℓ_2 norm. The problem (2.6) also has equivalent constraint form as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}}{\operatorname{argmin}} \quad f(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2 \leqslant \epsilon, \tag{2.7}$$

where ϵ is a bound on noise level.

2.1.3 Short Reminder on Convex Optimization

We assume that the penalty term is a convex lower-semicontinuous (l.s.c.) and a proper function on $\mathbb{R}^n$. This condition ensures the existence of a minimizer to problem (2.6). Let us recall the proximity operator of a l.s.c. function introduced by [13, 111].

Definition 2.1.1 (Proximity operator)**.** Let $J:\mathbb{R}^n \mapsto \mathbb{R}^n$ be a l.s.c. convex function. For any $\eta \in (0,\infty)$, the proximity operator of J is denoted by $\text{prox}_{\eta J}:\mathbb{R}^n \mapsto \mathbb{R}^n$ is defined as:

$$\text{prox}_{\eta J}(\mathbf{z}) \triangleq \underset{\alpha\in\mathbb{R}^{\mathbf{n}}}{\text{argmin}} \quad \eta J(\alpha) + \frac{1}{2}\|\mathbf{z}-\alpha\|_2^2. \tag{2.8}$$

If the proximity operator to the prior constraint $f(\cdot)$ exists, then the solution to problem (2.6) can be obtained by using proximal algorithms. The most well-known example of the proximity operator is the shrinkage given by the Tikhonov regularization and the soft-thresholding prompted by the ℓ_1 norm. The interested readers can refer to [13] for more example of proximity operators. One of the most popular methods for solving problem (2.6) is the *Iterative Shrinkage/Thresholding Algorithm* [37, 49] (ISTA), where each iteration involves matrix-vector multiplication followed by a shrinkage step. Note that ISTA belongs to a general family of *forward-backward* splitting algorithms [35], which combines a gradient step followed by a proximity calculation at each step. The forward-backward algorithm is described in Algorithm 1.

Algorithm 1: Forward-backward splitting algorithm

Initialization: $\mathbf{x} \in \mathbb{R}^n, \gamma = \sigma_1(\mathbf{A})$
while *not converged* **do**
  $\mathbf{x}_{n+1} = \text{prox}_{\frac{\lambda}{\gamma}f}(\mathbf{x}_n - \frac{1}{\gamma}\mathbf{A}^\top(\mathbf{y}-\mathbf{A}\mathbf{x}_n)$

Although backward-forward algorithm has simple updates, it has slow convergence in practice. We refer to [35] for more efficient algorithms and their accelerations e.g. FISTA [14] as a fast variant of ISTA.

Generalized Soft Thresholding

The soft thresholding operator is a point-wise operator, given $\mathbf{z} \in \mathbb{R}^n$ the soft thresolding on $\mathbf{z}$ for a threshold η is defined as

$$\mathcal{S}_\eta(\mathbf{z}_i) = \text{sign}(\mathbf{z}_i)\cdot\max(|\mathbf{z}_i|-\eta,0) \triangleq \text{sign}(\mathbf{z}_i)\big(|\mathbf{z}_i|-\eta\big)_+, \tag{2.9}$$

The soft thresholding can be defined in form of generalized thresholding as follows

$$\begin{aligned} \mathcal{S}_\eta(\mathbf{z}_i) &= \operatorname{sign}(\mathbf{z}_i)\big(|\mathbf{z}_i| - \eta\big)_+ \\ &= \frac{\mathbf{x}_i}{|\mathbf{x}_i|}\big(|\mathbf{z}_i| - \eta\big)_+ \\ &= \mathbf{x}_i\left(1 - \frac{\eta}{|\mathbf{x}_i|}\right)_+ \\ \mathcal{S}_\eta(\mathbf{z}_i) &= \mathbf{z}_i\big(1 - \Omega_i(\mathbf{z}_i)\big)_+. \end{aligned} \tag{2.10}$$

$\Omega_i(\mathbf{z}_i)$ is the shrinkage coefficient and is different for each element of $\mathbf{z}$. The general soft thresholding is useful to unify different proximity operators.

2.2 Sparse Representation

Sparsity looks for the smallest number of coefficients to represent a signal in a basis or dictionary and do not consider the possible relations between data variables. Sparse representation in the context of optimization is modeled by the ℓ_0 pseudo-norm. Let assume the signal $\mathbf{x} \in \mathbb{R}^n$ is expressed in some domain $\Phi \in \mathbb{R}^p$, i.e. $\mathbf{x} = \Phi\alpha$, where $\alpha \in \mathbb{R}^p$ is the decomposition of $\mathbf{x}$ in Φ. Then the linear inverse problem (2.5) is defined as

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{n} = \mathcal{A}(\Phi\alpha) + \mathbf{n} = \Psi\alpha + \mathbf{n}. \tag{2.11}$$

Sparse regularization in form of (2.6) follows

$$\hat{\mathbf{x}} = \Psi \underset{\alpha \in \mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}} \quad \|\mathbf{y} - \Psi\alpha\|_2^2 + \lambda\|\alpha\|_0. \tag{2.12}$$

However, inducing sparsity with ℓ_0 pseudo-norm is NP-hard. In practice to circumvent this problem either greedy methods and relaxation of the ℓ_0 pseudo-norm is used. Greedy methods such as Iterative Hard Thresholding (IHT) [17] are relatively fast algorithms, however their performance are not guaranteed and only under strict conditions can solve sparse regularization [153]. The relaxations replace the ℓ_0 pseudo-norm by convex surrogate such as the ℓ_1 norm [32] and these methods have guarantees but in exchange for slow convergence.

Iterative Hard Thresholding

Iterative Hard Thresholding (IHT) retrieves sparse solutions by fixing the maximum number of non-zero coefficients to some constant S, i.e. IHT solves the S-sparse problem

$$\hat{\mathbf{x}} = \Psi \underset{\alpha \in \mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}} \quad \|\mathbf{y} - \Psi\alpha\|_2^2 \quad \text{s.t.} \quad \|\alpha\|_0 \leq S. \tag{2.13}$$

IHT updates involve multiplications by Ψ and $\Psi^\top$ as well as two vector additions. The IHT at each iteration solves

$$\alpha_{n+1} = \mathcal{H}_S\big(\alpha_n + \Psi^\top(\mathbf{y} - \Psi\alpha_n)\big), \tag{2.14}$$

where $\mathcal{H}_S$ is the hard thresholding operator that keeps the S largest coefficients and ignores the remaining ones.

Convex Relaxation

The ℓ_1 norm relaxation of the sparse recovery problem was introduced in the context of sparse representation by Chen et al. [32] and in sparse regression by Tibshirani [151]. Thereafter, sparse regularization has found different applications notably in compressive sensing [23, 41].

Within the context of least-square regression, sparse regularization is known as Least Absolute Shrinkage and Selection Operator (LASSO) [151] and in signal processing is known as Basis Pursuit (BP). Sparse representation of signal $\mathbf{x}$ in Φ is

$$\underset{\alpha\in\mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}} \quad \|\mathbf{x} - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1. \tag{2.15}$$

In statistics, the LASSO formulate the sparse recovery problem as

$$\underset{\beta\in\mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}} \quad \|\mathbf{z} - \Gamma\beta\|_2^2 + \lambda\|\beta\|_1, \tag{2.16}$$

where $\Gamma \in \mathbb{R}^{n\times p}$ is the observations described by p variables and $\mathbf{z} \in \mathbb{R}^n$ denotes the desired solution. The LASSO formulation can be employed in the context of classification where $\mathbf{z}$ would be the discrete entries to be classified.

The ℓ_1 norm can be employed to solve linear inverse problem. Using BP formulation as

$$\hat{\mathbf{x}} = \Psi \underset{\alpha\in\mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}} \quad \|\mathbf{y} - \Psi\alpha\|_2^2 + \lambda\|\alpha\|_1. \tag{2.17}$$

The sparse regularization problem (2.15) can be solved by the forward-backward algorithm 1. The proximity operator of ℓ_1 norm is coefficient-wise soft thresholding

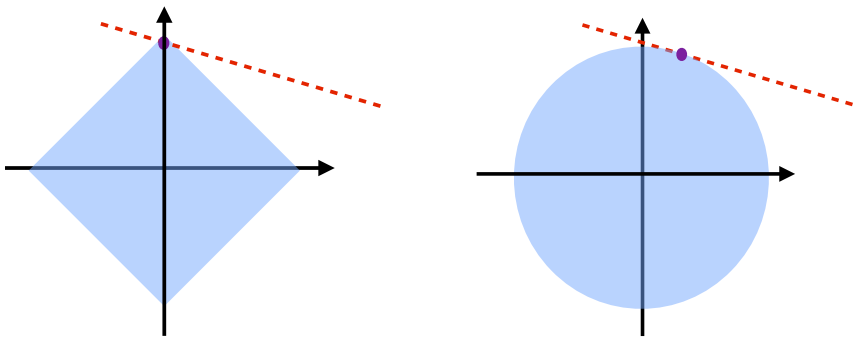
$$\operatorname{prox}_{\eta\|\cdot\|_1}(\alpha_i) = \operatorname{sign}(\alpha_i)\big(|\alpha_i| - \eta\big)_+. \tag{2.18}$$

Geometric Intuition of ℓ_1-norm Ball

We consider sparse recovery in regularization form (BP), however the regularization problem is equivalent to the constraint problem for some $\mu \geq 0$

$$\hat{\mathbf{x}} = \Psi \underset{\alpha \in \mathbb{R}^p}{\operatorname{argmin}} \ \|\mathbf{y} - \Psi\alpha\|_2^2 \quad \text{s.t.} \quad \|\alpha\|_1 \leq \mu, \tag{2.19}$$

which indicates the solution to sparse linear regression depends on the geometry of the ℓ_1 ball. Figure 2.1 compares when the ℓ_1 or ℓ_2 norms are used as the constraints. Since the ℓ_2 ball is isotropic, the regularization does not favoring any particular direction. However, the ℓ_1 ball is anisotropic and the singular points of ℓ_1 ball are located on axis-aligned linear subspaces in $\mathbb{R}^p$. if the hyperplane corresponding to the data constraint is tangent to the ball at any of those points, ℓ_1 ball promotes sparsity.

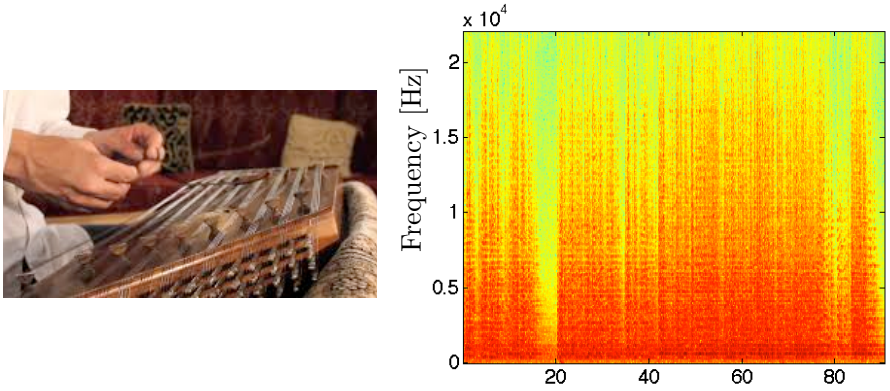


(a) $\|\alpha\|_1$ (b) $\|\alpha\|_2^2$

Figure 2.1: Comparison between ℓ_1 ball (on the right) and ℓ_2 ball (on the left). The ℓ_1 ball has singular points on axes, the red line depicts the data constraint. One can observe that the ℓ_1 norm unlike the ℓ_2 norm results in a sparse solution, i.e. the constraint line intersects the ℓ_1 ball on axis which leads to thresholding some coefficients to zero. While, the isotropic shape of the ℓ_2 norm cannot promote sparsity.

2.3 Structured Sparsity Inducing Norms

Relation between data variables in real word problems motivates constraints that raise structures. For instance, in Figure 2.2 significant data coefficients are organized along horizontal or vertical lines. As explained, the ℓ_1 norm only looks for sparse solution and does not encode information about the structures of coefficients. In this section, we define a set of constraints known as structured sparsity inducing norms where all sparse patterns do not have equal probability. For example, group sparsity leverages a group structure between data variables such that all coefficients inside a group are either accepted or rejected together.



Santour (Dulcimer)

Figure 2.2: Sparsity cannot model relations between data coefficients. For instance, the time-frequency representation of a piece of Santur shows that the coefficients are structured into groups.

2.3.1 Structured Sparsity Norms with Disjoint Grouping

We observed in Figure 2.2 that one way to incorporate signal patterns in the sparse model is to group the coefficients such that the coefficients within a group are compared together. In order to index the coefficients and their corresponding groups, each coefficient is indexed by a pair of (g,k) where g is the group index and k is the index of the coefficients in group g.

Definition 2.3.1 (Mixed $\ell_{p,q}$-norm)**.** Let $\mathbf{x}\in\mathbb{R}^n$ be a vector indexed by $(g,k)\in\mathbb{N}^2$ and w_g be a positive weight of group g, then the mixed $\ell_{p,q}$-norm is defined as

$$\|\mathbf{x}\|_{p,q}=\left(\sum_{g=1}^{G} w_g\left(\sum_{k=1}^{K}|\mathbf{x}_{g,m}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}. \tag{2.20}$$

The appropriate choice of group weights w_g is important when the groups have different size. The sparse recovery problem in (2.17) is redefined as

$$\hat{\mathbf{x}}=\Psi\underset{\alpha\in\mathbb{R}^{\mathbf{p}}}{\operatorname{argmin}}\ \|\mathbf{y}-\Psi\alpha\|_2^2+\lambda\|\alpha\|_{p,q}. \tag{2.21}$$

Group Lasso

A practical choice of $(p,q)=(2,1)$, this choice of grouping promotes variables in the same group to be jointly selected or set to zero. In the context of least-square regularization, this choice is known as the group Lasso [155, 172] and in signal processing community as joint sparsity [51]. The group Lasso has shown to have applications in various fields such as in Machine learning [117, 172], image/signal processing [62, 88] and to improve

the learned model for a block structured data [133, 146].

The proximity operator of the group Lasso is a group version of the generalized thresholding [89]. The proximity operator of the group Lasso for each coordinate reads

$$\text{prox}_{\eta\|\cdot\|_{w,2,1}}(\mathbf{x}_{g,k}) = \mathbf{x}_{g,k}\left(1 - \frac{\eta\sqrt{w_g}}{\|\mathbf{x}_g\|_2}\right)_+, \tag{2.22}$$

where $\|\mathbf{x}_g\|_2$ is the ℓ_2 norm of groups. In form of the generalized thresholding the shrinkage factor is defined as

$$\Omega_{g,k} = \frac{\eta\sqrt{w_g}}{\|\mathbf{x}_g\|_2}. \tag{2.23}$$

The interpretation of (2.22) is straightforward. If a group ℓ_2 norm is less than the shrinkage coefficient ($\eta\sqrt{w_g}$), the whole group is ignored, otherwise the non-zero coefficients in the group are kept and shrunk.

Elitist Lasso

We change the order of norms in the group Lasso, i.e. $(p, q) = (1, 2)$ and the mixed norm is $\ell_{w,1,2}$. This norm is called the Elitist Lasso [89] in signal processing and Exclusive Lasso in Machine learning society [176]. The Elitist Lasso promotes sparsity in group members while imposing equal importance to each group. Recall that the group Lasso looks for a sparse set of active groups and favors dense coefficients within groups. However, the Elitist Lasso keeps all groups and shrunk entries within each groups. That is the Elitist Lasso promotes the most dominant elements of each group and ignores the others, which explains why it is called the Elitist Lasso.

To compute the proximity operator of Elitist Lasso, we define $\mathbf{d}_{g,k} = |\mathbf{x}_{g,k}|/w_{g,k}$ and sort them in descending order for each group to obtain a new order $\tilde{\mathbf{d}}_g$ [89]

$$\tilde{\mathbf{d}}_{g,1} \geq \tilde{\mathbf{d}}_{g,2} \geq, \ldots \geq \tilde{\mathbf{d}}_{g,K} \quad \forall g. \tag{2.24}$$

We define K_g as

$$\begin{cases} \tilde{\mathbf{d}}_{g,K_g} & > \eta \sum_{k=1}^{K_g} w_g^2 (\tilde{\mathbf{d}}_{g,k} - \tilde{\mathbf{d}}_{g,K_g}), \\ \tilde{\mathbf{d}}_{g,K_g+1} & \leq \eta \sum_{k=1}^{K_g+1} w_g^2 (\tilde{\mathbf{d}}_{g,k} - \tilde{\mathbf{d}}_{g,K_g}). \end{cases} \tag{2.25}$$

K_g corresponds to the largest member of $\tilde{\mathbf{d}}_g$ and to the last element when all members are equal. Finally, the proximity operator of $\ell_{w,1,2}$ is given

$$\text{prox}_{\eta\|\cdot\|_{w,1,2}}(\mathbf{x}_{g,k}) = \text{sgn}(\mathbf{x}_{g,k})\left(|\mathbf{x}_{g,k}| - \frac{\eta}{1 + \eta\Xi_{w_g}} \sum_{k=1}^{K_g} |\mathbf{x}_{g,k}|\right)_+, \tag{2.26}$$

where $\Xi_{w_g} = \sum_{k=1}^{K_g} w_g^2$. In generalized thresholding form, the shrinkage factor is

$$\Omega_{g,k} = \frac{\eta}{1+\eta\Xi_{w_g}} \frac{\sum_{k=1}^{K_g} |\mathbf{x}_{g,k}|}{|\mathbf{x}_{g,k}|}. \tag{2.27}$$

The shrinkage of the Elitist Lasso unlike the group Lasso is not proportional to the energy of groups, instead it is proportional to the cumulative norm of the greatest member of groups. In addition, the shrinkage is not fixed for group members but it varies among the coefficients. In a given group, a coefficient is set to zero if its value is a fraction of the cumulative norm of the dominant coefficients in that group.

2.3.2 Structured Sparsity Norms with Overlapping Groups

As explained the group Lasso discards all coefficients inside a group. Therefore, the groups are independent and a coefficient cannot belong to different groups. Though this structure can be of interest for some applications, there are tasks that require dependency among the groups (e.g. background subtraction [11, 71], dictionary learning [84], and wavelet based denoising [127]).

When there is no overlap between groups, discarding a group sets all its element to zero, therefore the group Lasso selects a small number of dense groups. However, in the overlapping case, if a group is ignored its entries are set to zero, though they belong to other groups which are not shrunk to zero. That is in the overlapping case the groups may not be dense [74]. Figure 2.3 illustrates three overlapping groups. If the penalty sets the first and third group to zero, what remains is the second group with non-zero members that do not belong to either the first or third groups (for more information see [74]). However, there is no closed form solution for the proximity operator of this penalty, authors in [74] suggested an iterative scheme to solve (2.21) using the overlapping group Lasso.

Hierarchical Structure

One of the interesting overlapping group structure is the Hierarchical structure. More precisely, an element of the tree structured vector $\mathbf{x}$ may be selected if all its ancestors in the tree are also selected. Figure 2.4 displays an example of groups for this structured sparsity model. The hierarchical structure has various applications, for example in wavelet-based denoising [11, 71, 76], prediction of cognitive tasks using fMRI [75], and hierarchical dictionary learning [76].

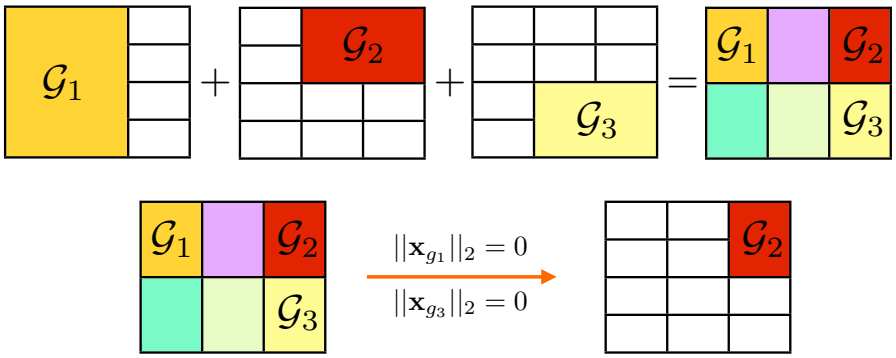


Figure 2.3: Overlapping group Lasso. Top: Decomposition of $\mathbf{x}$ into groups. Bottom: Shrinking any group to zero removes its member from the result. In this example the first and third groups are set to zero, the solution contains those members of second group that neither belong to the first nor to the third group.

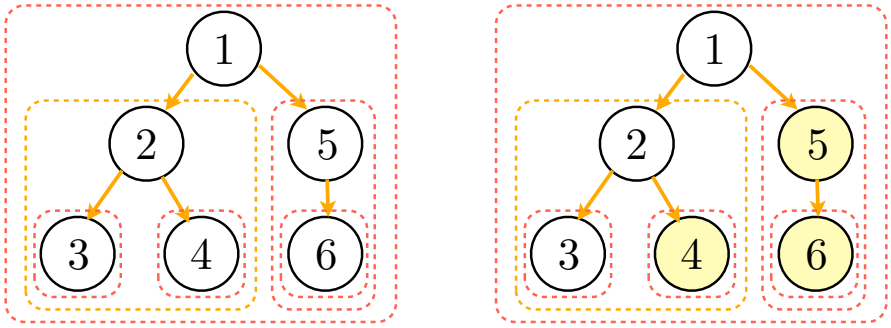


Figure 2.4: Hierarchical structured sparsity. Left: example of a tree structured vector and corresponding groups. Right: example of a sparsity pattern induced by the hierarchical structured norm. The groups $\{5,6\}$, $\{5\}$, and $\{4\}$ is set to zero (colored). The non-zero entries $\{1,2,3\}$ are connected. In hierarchical form a node is selected if its ancestors are also selected and if a entries is discarded all its descended will be ignored.

Latent Group Lasso

The group Lasso with overlapping results in groups which are not dense. For instance, Figure 2.3 illustrates three overlapping groups where the group Lasso penalty leads to the shrinkage of the first and third groups. The non-zeros coefficients are not the entire second group but the members of the second group which do not belong to either the first or the third group. Obozinski et al. [116] defined the Latent Group Lasso to model this structure. The latent group Lasso keeps all elements of a group, despite they also belong to another discarded group. Let define a set of latent variable $\mathbf{z}_g$ such that $\mathbf{z}_{g,i} = 0$ for all elements that do not belong to g. The latent group Lasso is represented by the

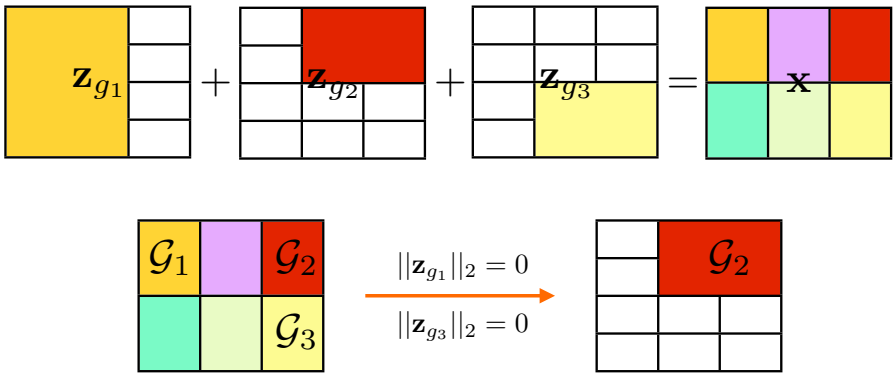


Figure 2.5: Latent group Lasso. Top: Decomposition of $\mathbf{x}$ into latent vectors. Bottom: applying latent group Lasso to the decomposition removes those latent variables that do not belong to any selected groups. In this example, the $\mathbf{z}_{g_1}$ and $\mathbf{z}_{g_3}$ are shrunk to zero, unlike to group Lasso with overlapping, all variables in the second group are kept.

penalty

$$\underset{\mathbf{z}\in\mathbb{R}^{\mathbf{n}\times\mathcal{G}}}{\operatorname{argmin}} \sum w_g \|\mathbf{z}_g\|_2 \quad \text{s.t.} \quad \begin{cases} \sum_{g\in\mathcal{G}} \mathbf{z}_g = \mathbf{x}, \\ \mathbf{z}_{g,i} = 0, \qquad \forall g \in \mathcal{G}, \quad i \notin g, \end{cases} \tag{2.28}$$

where w_g is a positive weight associated to each group. Intuitively, $\mathbf{x}$ is expressed as the sum of latent variables (see Figure 2.5). Applying group sparsity to the latent vectors shrinks some $\mathbf{z}_g$ to zero, since we impose $\sum_{g\in\mathcal{G}} \mathbf{z}_g = \mathbf{x}$, $\mathbf{x}_i$ is non-zero if it belongs to a group that is not shrunk to zero. Therefore, in contrast to the group Lasso with overlapping that promotes non-dense groups, the latent group Lasso leads to dense latent groups. The linear inverse problem (2.21) using latent group Lasso solves the following

$$\hat{\mathbf{x}} = \Psi \underset{\alpha\in\mathbb{R}^{\mathbf{P}},\ \beta\in\mathbb{R}^{\mathbf{p}\times\mathcal{G}}}{\operatorname{argmin}} \|\mathbf{y} - \Psi\alpha\|_2^2 + \lambda \sum_{g\in\mathcal{G}} w_g \|\beta_g\|_2 \quad \text{s.t.} \quad \begin{cases} \sum_{g\in\mathcal{G}} \beta_g = \alpha, \\ \beta_{g,i} = 0, \qquad \forall g \in \mathcal{G}, \quad i \notin g. \end{cases} \tag{2.29}$$

2.3.3 Three-level Structured Sparsity

We assumed that coefficients are divided into groups, however one can cluster the groups and add a third layer of grouping. Therefore, a coefficient is indexed by a triplet (c, g, k) where c represents clusters. The three-level mixed norm has applications, for example, in audio processing [89] and image denoising [36].

Definition 2.3.2 (Three-level mixed norm: $\ell_{p,q,r}$-norm)**.** Let $\mathbf{x} \in \mathbb{R}^n$ be a vector indexed

by $(c, g, k) \in \mathbb{N}^3$ and w be a positive weight, then the mixed $\ell_{p,q,r}$-norm is defined as

$$\|\mathbf{x}\|_{w,p,q,r} = \left(\sum_{c=1}^{C} \left(\sum_{g=1}^{G} \left(\sum_{k=1}^{K} w_{c,g,k} |\mathbf{x}_{c,g,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{r}{q}} \right)^{\frac{1}{r}}. \tag{2.30}$$

We restrict our choice of the three-level mixed norm to $\ell_{2,1,2}$. This norm promotes sparse groups similar to the Elitist Lasso when the third layer is not considered. The last layer of ℓ_2 norm favors coefficients that have important contribution in the cluster and discards other coefficients. To define the proximity operator of the $\ell_{w,2,1,2}$, we assume that the weight is constant for members of a group, i.e. $\forall\ w_{c,g,k} = w_{c,g}$. Similar to the $\ell_{1,2}$ norm, we define an intermediate variable

$$\mathbf{d}_{c,g} = \frac{\|\mathbf{x}\|_2}{\sqrt{w_{c,g}}}. \tag{2.31}$$

Then for each c, we obtain a new variable $\tilde{\mathbf{d}}_{c,g}$ by sorting $\mathbf{d}_{c,g}$, i.e. $\forall\ g_c,\ \mathbf{d}_{c,g_c+1} \leq \mathbf{d}_{c,g_c}$. Likewise to (2.25) for $\ell_{1,2}$ norm a new indexing G_c is obtained. Finally the proximity operator of the $\ell_{w,2,1,2}$ norm is given by [89]

$$\text{prox}_{\eta\|\cdot\|^2_{w,2,1,2}}(\mathbf{x}_{c,g,k}) = \mathbf{x}_{c,g,k} \left(1 - \frac{\eta\sqrt{w_{c,g}}}{1 + \eta\Xi_{w_c}} \frac{\sum_{g=1}^{G_c} \sqrt{w_{c,g}} \|\mathbf{x}_{c,g}\|_2}{\|\mathbf{x}_{c,g}\|_2} \right)_+, \tag{2.32}$$

where $\Xi_{w_c} = \sum_{g=1}^{G_c} w_{c,g}$.

2.4 Simultaneous Structure Modeling with Sum of Constraints

In some applications, the aforementioned structured sparsity norms cannot model multiple structural information in data. A popular methodology to leverage simultaneous patterns is to combine multiple structure promoting constraints.

2.4.1 Sum of the Group Lasso and Sparsity

Gramfort et al. [63] defined a linear inverse problem for M/EEG sensors that deploys the group Lasso to insure a few active sources in a given time window. However, the group Lasso favors dense groups and does not promote sparse source estimation. In order to recover a set of groups with sparse coefficients, the authors proposed to simultaneously impose sparsity and group Lasso on the coefficients as

$$f(\mathbf{x}) = \eta\|\mathbf{x}\|_{2,1} + \mu\|\mathbf{x}\|_1, \tag{2.33}$$

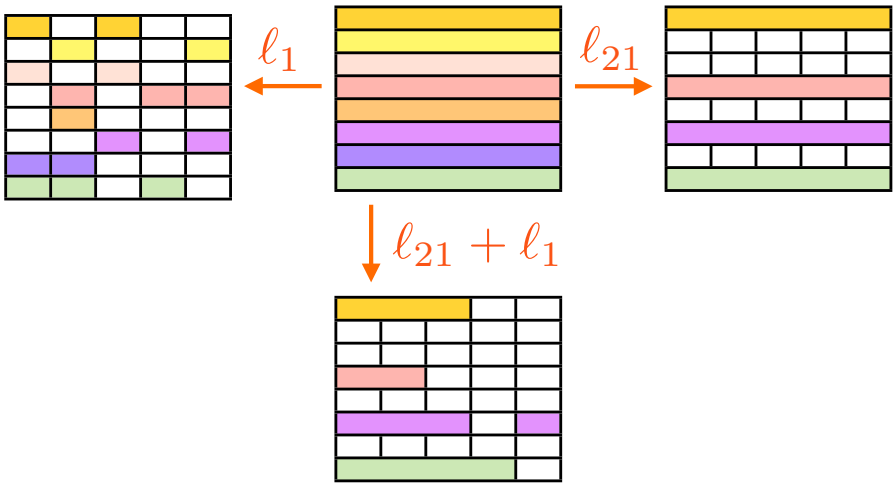


Figure 2.6: Comparison between the sparsity pattern promoted by ℓ_1 norm, ℓ_{21} mixed norm and $\ell_{21} + \ell_1$ penalties. ℓ_1 norm does not promote any structure and the non-zero coefficients are scattered. ℓ_{21} mixed norm favors for dense group structure. $\ell_{21} + \ell_1$ promotes structured group with intra-sparsity.

where the regularization parameters η and μ control the density of the groups. Figure 2.6 visualizes the ℓ_1 norm, $\ell_{2,1}$ and some of the norms. We observe that $\ell_{2,1} + \ell_1$ promotes a sparser solution in comparison to the group Lasso.

Let $\mathbf{x} \in \mathbb{R}^n$ be double indexed (g, k). The proximity operator of the $\ell_{2,1} + \ell_1$ is given by [63]

$$\text{prox}_{\eta\|\cdot\|_{2,1}+\mu\|\cdot\|_1}(\mathbf{x}_{g,k}) = \frac{\mathbf{x}_{g,k}}{|\mathbf{x}_{g,k}|}\left(|\mathbf{x}_{g,k}| - \mu\right)_+ \left(1 - \frac{\eta}{\sqrt{\sum_{k=1}^{K}\left(|\mathbf{x}_{g,k}| - \eta\right)_+^2}}\right)_+ . \tag{2.34}$$

2.5 Non-convex Structured Sparsity

We mainly focused on the convex penalties to impose structures of the underlying signals, however there are many non-convex approaches introduced to address signals structures. For instance, Baraniuk et al. [11] proposed a framework to model the inter-dependencies in data coefficients. This framework models signal structures using a union-of-subspaces to decrease the degree of freedom of a signal by permitting specific supports of coefficients.

An S-sparse signal $\mathbf{x} \in \mathbb{R}^n$ lives in $\mathbb{K}_S \subset \mathbb{R}^n$, a union of $\binom{n}{S}$ S-dimensional subspaces. This structured sparsity scheme favors certain configuration of the coefficients, i.e. certain subspaces in $\mathbb{K}_S$ are allowed. To formally state the structure model, let Δ represents a possible configuration of entries of $\mathbf{x}$ and Δ^C represents its complement.

Definition 2.5.1 (see [11])**.** A structured sparsity model Λ_S is defined as the union of λ_S canonical subspaces of S-dimension

$$\Lambda_S = \bigcup_{i=1}^{\lambda_S} \{\mathbf{x} | \Delta_i \in \mathbb{R}^S, \Delta_i^C = 0\}, \tag{2.35}$$

where $\{\Delta_1, \ldots, \Delta_{\lambda_S}\}$ is the set of all possible indices of $\mathbf{x}$, with $|\Delta_\cdot| = S$.

Signals from Λ_S are called S-model sparse. To recover the underlying signal from the linear inverse problem (2.5), the model based scheme modified CoSaMP algorithm [114] merely by replacing the best S-term sparse approximation step by the best S-structured sparse approximation. It is clear that $\Lambda_s \subset \mathbb{K}_s$, therefore the model based algorithm at each step requires to search over λ_s subspaces of Λ_s rather than $\mathbb{K}_s$.

2.6 Intrinsic Low-dimensionality

The Big Bang-like explosion of information about everything from the World Wide Web to science and engineering is being propelled by massive amounts of high-dimensional data continuously produced and stored at decreasing cost. The quickening pace of data collection presents a challenge as well as an opportunity, as a result scientific advances are becoming more and more data-driven.

To address the curse of dimensionality, we rely on the fact that though such data lie on high-dimensional space, their intrinsic dimensionality is low, i.e. they lie on some low-dimensional subspace [44] or some low-dimensional manifolds [150].

One can stack the data points into columns of a matrix. Since the high dimensional data is supposed to have intrinsic low-dimensionality, the underlying matrix would be a (approximately) low-rank matrix. Low-rank matrices has been utilized in many applications, for instance, low-rank matrices play a central role in large-scale data analysis and dimensionality reduction, including system identification [109], collaborative filtering [144] and Principal Component Analysis (PCA) [80].

2.6.1 Matrix Completion

In many practical problems, one would like to recover the data matrix from a set of know entries. For example, in recommender systems such as Netflix, users provide ratings on a subset of entries in a database, and the vendors would like to know if based on the available ratings they can estimate ratings of missing entries.

However, in many problems, we know the data matrix is structured such that it has intrinsic low-dimensionality, i.e. it is low-rank or approximately low-rank. Let assume the data matrix $\mathbf{X}$ is a square $n \times n$ matrix of rank r. The matrix $\mathbf{X}$ has n^2 entries but it

has $(2n-r)r$ degree of freedom[1]. When the rank is small the degree of freedom is smaller than the number of entries. Specially, for high dimensional data, the information (degree of freedom) is much less than the data dimension. Then the problem is to recover a low-rank matrix from a small set of observations. In signal and image processing the low-rank structure is used to fill in the missing entries of a large low-rank matrix, this problem is known as matrix completion [25, 85, 130]. Filling the Netflix rating database [25], or denoising the corrupted entries of a video sequence [77] are examples of the matrix completion.

Clearly, one cannot recover all forms of low-rank matrices. For instance, if the underlying matrix has one non-zero entries. Clearly, one cannot guess the entries of the matrix till almost all elements of the matrix are observed. In addition, it is impossible to recover a low-rank matrix, even a rank-1 matrix, from a sampling set which avoids any column or row of the matrix. For instance, if the sampling set avoids any entry of the first row, no method can estimate the unobserved row.

Let assume the set Ω corresponds to the location of observed entries ($(i,j) \in \Omega$ if $\mathbf{X}_{ij}$ is observed), the projection operator $\mathcal{P}_\Omega$ is the orthogonal projection onto the matrices supported on Ω

$$\mathcal{P}_\Omega(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij}, & (i,j) \in \Omega, \\ 0, & (i,j) \notin \Omega. \end{cases} \tag{2.36}$$

The observed entries of $\mathbf{X}$ is defined as

$$\mathbf{Y} = \mathcal{P}_\Omega(\mathbf{X}). \tag{2.37}$$

If the number of measurements is sufficiently large, and the observed entries are uniformly distributed, one might hope that there is only one low-rank matrix related to these entries. The intrinsic low-dimensionality assumption on the data can recover the underlying low-rank matrix from the observed entries through

$$\hat{\mathbf{X}} = \underset{\mathbf{X} \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \ \operatorname{rank}(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{Y} = \mathcal{P}_\Omega(\mathbf{X}). \tag{2.38}$$

Similar to ℓ_0 minimization, the rank minimization problem is an NP-hard [47]. In practice, there are two approximation techniques to address this problem, greedy algorithms such as ADMiRA [94] or convex relaxation using the tightest convex envelope for rank

[1] The degree of freedom of a $n \times n$ matrix of rank r interprets as follows: one selects the first r columns of the matrix with n degree of freedom. Then the remaining columns is the linear combination of the first r columns which gives r degree of freedom for the remaining columns, namely the r coefficients of the linear combination. Thus the degree of freedom of a rank r matrix is

$$rn + (n-r)r = (2n-r)r$$

constrain [25, 28]. Let $\mathbb{C}$ be a given set, the convex envelope of a function $f : \mathbb{C} \to \mathbb{R}$ is defined as the largest convex function b such that $b(x) \leq f(x)\ \forall x \in \mathbb{C}$. That is among all convex functions that lower bound f, b is the best approximation. Therefore, b can be used to approximate f for a convex minimization [131].

Definition 2.6.1 (see [47])**.** For a given matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$, $\|\mathbf{X}\| \leq 1$, we have $\text{rank}(\mathbf{X}) \geq \|\mathbf{X}\|_*$, the nuclear norm is the tightest convex lower bound of the matrix rank.

Finally, we can relax the matrix rank minimization by the nuclear norm minimization. Therefore, the low-dimensional data recovery can be relaxed to nuclear norm minimization [25, 131] as follows

$$\hat{\mathbf{X}} = \underset{\mathbf{X} \in \mathbb{R}^{\mathbf{n} \times \mathbf{m}}}{\text{argmin}} \ \|\mathbf{X}\|_* \quad \text{s.t.} \quad \mathbf{Y} = \mathcal{P}_\Omega(\mathbf{X}). \tag{2.39}$$

To calculate the proximity operator of the nuclear norm of matrix $\mathbf{X}$ the Singular Value Decomposition (SVD) is computed as $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top$. The proximity operator of nuclear norm is defined as

$$\text{prox}_{\eta\|\cdot\|_*}(\mathbf{X}) = \mathbf{U}\big(\Sigma - \eta\big)_+\mathbf{V}^\top. \tag{2.40}$$

2.6.2 Low-rank Matrix Recovery

We extend the linear inverse problem (2.5) to the case where underlying data lie on some low-dimensional subspace. Let assume that the original data points are stacked into columns of the matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$, and the observed m-dimensional vector $\mathbf{y} \in \mathbb{R}^m$ is obtained through the linear operator $\mathcal{A} : \mathbb{R}^{n \times k} \to \mathbb{R}^m$. We can exploit the intrinsic low-dimensionality assumption on data to recover it from the linear measurements. Thus, the *affine rank minimization* reads

$$\hat{\mathbf{X}} = \underset{\mathbf{X} \in \mathbb{R}^{\mathbf{n} \times \mathbf{m}}}{\text{argmin}} \ \text{rank}(\mathbf{X}) \quad \text{s.t.} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2 \leq \epsilon, \tag{2.41}$$

where ϵ is a bound on the measurement noise. Similar to matrix completion problem, we can relax the affine rank minimization to nuclear norm minimization as follows

$$\hat{\mathbf{X}} = \underset{\mathbf{X} \in \mathbb{R}^{\mathbf{n} \times \mathbf{m}}}{\text{argmin}} \ \|\mathbf{X}\|_* \quad \text{s.t.} \quad \|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2 \leq \epsilon. \tag{2.42}$$

2.6.3 Robust Principal Component Analysis

PCA is one of the mostly used statistical tool for dimensionality reduction. However, PCA performance is limited when the observation is corrupted. In many applications such as image/video processing and web data analysis, it is impossible to have perfect noiseless data. However, the corrupted data values are uncorrelated to the low-

dimensionality of data. Candès et al. [29] proposed *Robust PCA* to recover a low-rank matrix $\mathbf{L} \in \mathbb{R}^{n\times k}$ from noisy observation $\mathbf{X} \in \mathbb{R}^{n\times k}$ which reads

$$\mathbf{X} = \mathbf{L} + \mathbf{S}. \tag{2.43}$$

The matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$ separates the noisy observation from the low-rank matrix which can have arbitrarily large entries but it is assumed to be sparse with unknown support. To recover the low-rank matrix and the sparse corrupted values from the noisy observation, the Robust PCA solves

$$\underset{\mathbf{L},\mathbf{S}\in\mathbb{R}^{n\times k}}{\operatorname{argmin}} \ \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathbf{L}+\mathbf{S}=\mathbf{X}, \tag{2.44}$$

the parameter λ controls the balance between between the two terms.

We define $\mathbb{R}_0 \triangleq \mathbb{R}^{n\times k} \times \mathbb{R}^{n\times k}$ by the Cartesian product of $\mathbb{R}^{n\times k}$ and a point in $\mathbb{R}_0$ is defined as $\mathbf{X} \triangleq (\mathbf{X}_1, \mathbf{X}_2) \in \mathbb{R}_0$. The inner product in $\mathbb{R}_0$ is defined as

$$\begin{aligned}\langle \mathbf{X}, \mathbf{Y} \rangle &\triangleq \langle \mathbf{X}_1, \mathbf{Y}_1 \rangle + \langle \mathbf{X}_2, \mathbf{Y}_2 \rangle \\ &= \operatorname{trace}(\mathbf{X}_1^\top \mathbf{Y}_1) + \operatorname{trace}(\mathbf{X}_2^\top \mathbf{Y}_2).\end{aligned} \tag{2.45}$$

The norm on $\mathbb{R}_0$ induced by inner product is

$$\|\mathbf{X}\|_{\mathbb{R}_0} = \langle \mathbf{X}, \mathbf{X} \rangle = \|\mathbf{X}_1\|_{\mathrm{F}} + \|\mathbf{X}_2\|_{\mathrm{F}}, \tag{2.46}$$

where $\|\cdot\|_{\mathrm{F}}$ is the matrix Frobenius norm.

Proposition 2.6.1. *For any point $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) \in \mathbb{R}_0$ and a function $f(\mathbf{X}) = f_1(\mathbf{X}_1) + f_2(\mathbf{X}_2)$, the proximity operator of f is defined as*

$$prox_f(\mathbf{X}) = \Big(prox_{\eta f_1}(\mathbf{X}_1),\ prox_{\mu f_2}(\mathbf{X}_2) \Big). \tag{2.47}$$

Proof. Since $\mathbf{X}$ is defined in $\mathbb{R}_0$, then (4.17) yields

$$\begin{aligned}\operatorname{prox}_{\eta f}(\mathbf{X}) &= \underset{\mathbf{Y}\in\mathbb{R}_0}{\operatorname{argmin}} \ f(\mathbf{X}) + \frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_{\mathbb{R}_0}^2 \\ &= \underset{\mathbf{Y}\in\mathbb{R}_0}{\operatorname{argmin}} \ \eta f_1(\mathbf{X}_1) + \mu f_2(\mathbf{X}_2) + \frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_{\mathbb{R}_0}^2 \\ &= \underset{\mathbf{Y}\in\mathbb{R}_0}{\operatorname{argmin}} \ \eta f_1(\mathbf{X}_1) + \frac{1}{2}\|\mathbf{X}_1-\mathbf{Y}_1\|_{\mathrm{F}}^2 + \mu f_2(\mathbf{X}_2) + \frac{1}{2}\|\mathbf{X}_2-\mathbf{Y}_2\|_{\mathrm{F}}^2 \\ &= \Big(\operatorname{prox}_{\eta f_1}(\mathbf{X}_1),\ \operatorname{prox}_{\mu f_2}(\mathbf{X}_2) \Big).\end{aligned}$$

□

If we define $\mathbf{M} = (\mathbf{L}, \mathbf{S})$, the proximity operator for the sparse and low-rank decomposition is calculated by

$$\text{prox}_{\|\cdot\|_*+\lambda\|\cdot\|_1}(\mathbf{M}) = \Big(\text{prox}_{\|\cdot\|_*}(\mathbf{L}),\ \text{prox}_{\mu\|\cdot\|_1}(\mathbf{S})\Big). \tag{2.48}$$

It is obvious that not any arbitrary matrix can be composed into the low-rank and sparse components. For instance, if the data matrix $\mathbf{X}$ has one non-zero value, then since $\mathbf{X}$ is both low-rank and sparse, one cannot differentiate between the low-rank and the sparse components. Therefore, the low-rank component should not be sparse.

Another important condition arises when the sparse component is low-rank. For example, this can occurs when all non-zero values of the sparse component lie on a few columns. Then, it is clear that one cannot differentiate the sparse component from the low-rank component. To avoid such situations, one can assume that the sparsity pattern of the sparse component is selected from a uniform distribution.

2.6.4 Matrix Completion from Corrupted Data

This problem seeks for a low-rank matrix $\mathbf{L}$ from a few observations where some of them are corrupted. Similar to the matrix completion, let the set Ω represents the observed entries and $\mathcal{P}_\Omega$ is the orthogonal projection onto Ω. The matrix completion from corrupted data reads

$$\underset{\mathbf{L},\mathbf{S}\in\mathbb{R}^{\mathbf{n}\times\mathbf{k}}}{\text{argmin}}\ \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \text{s.t.} \quad \mathcal{P}_\Omega(\mathbf{L}+\mathbf{S}) = \mathbf{Y}. \tag{2.49}$$

In words, PCA seeks among all possible solution the one that matches the observed corrupted data and also minimizes the weighted sum of the nuclear norm and ℓ_1 norm.

2.7 Conclusion

In this chapter, we have explored several approaches for structural modeling based on convex optimization where prior knowledge allows to favor certain patterns. Traditional sparse constraint can easily be extended to these priors to express more complex structures which makes these constraints a powerful tool to promote prior knowledge on high-dimensional data. In the following chapters, we will present several applications where we are going to benefit from these tools to improve the performance of our algorithms.

Chapter 3

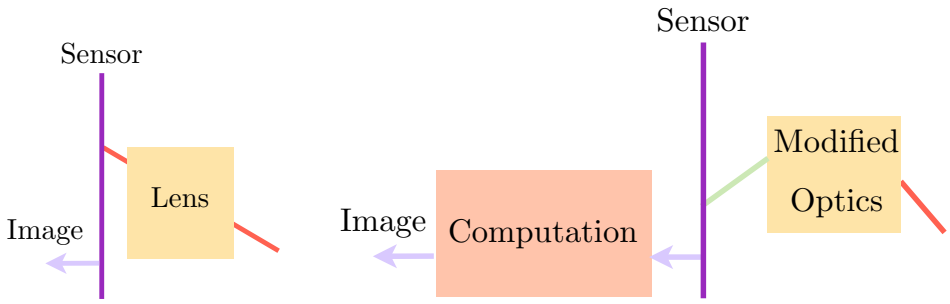
Computational Light Field Imaging

Traditional cameras are designed to model what a single human eye can observe: a two-dimensional colored image. The advent of image processing and digital camera technology has provided the capacity to exceed the limitation of analogue photography and introduce computational photography. One of the main goals of computational photography is to design camera systems which record visual information that cannot be acquired by traditional cameras. The computational camera would capture the visual information that allows new imaging modalities such as motion deblurring, hyper-spectral imaging and light field imaging. In this chapter, we provided a review on research that has been conducted on light field imaging using different computational techniques. This review will be served in the following chapter to introduce a new light field acquisition system which addresses the limitation of available light field camera designs.

3.1 What is a Computational Camera?

Traditional cameras consist of a sensor and a standard lens (see Figure 3.1(a)) which projects the rays passing through the lens onto the sensor. In other words, traditional cameras sample the complete set of rays emitted from a scene.

A computational camera is inspired by the diversity of perceptual system and improved by the advances in camera technology, image processing and optical fabrication. A computational camera is the combination of modified optics and a computation unit to acquire highly detailed visual information of a scene [174]. Figure 3.1(b) demonstrate a schematic of a computational camera where, in contrast to traditional cameras, the scene rays are coded and deviated by the optics to a different pixel location.



(a) Traditional Camera (b) Computational Camera

Figure 3.1: Comparison of traditional cameras and computational cameras. (a) Traditional cameras linearly project those rays passing through the lens to produce image. (b) Computational cameras capture coded rays through modified optics and the computation unit decodes the captured rays to produce the final image [174].

Images acquired by computational cameras are optically coded and their raw format may not be even visually informative. However, the information can be retrieved by using computation. The combination of computation and the specific optical design leads to a special type of imaging systems that can potentially enhance spatial, temporal, spectral and directional resolution of the imaging system. The enhancement achieved by the computerized acquisition promotes diverse applications in medical imaging, remote sensing, surveillance and automated fabrication.

3.2 Light Field Fundamentals

light fields describe the amount of light traveling in every direction through every point in space, time and wavelengths[1]. Light field was introduced to computer graphics in the 1990s [2, 59, 97]. A complete light field is described by 7-dimensional plenoptic function $L(x, y, z, \phi, \theta, \lambda, t)$, where (x, y, z) is spatial coordinates, (ϕ, θ) is the direction of the ray, λ is light wavelength, and t is time. For given time and wavelength, the light field is limited to 5-dimensional space. However, the 5-dimensional representation can further be reduced to 4-dimensional in free space (regions free of occluders) because the radiance[2] does not change along its propagation line [59, 97].

The 4-dimensional light field can be represented in several ways, as demonstrated in Figure 3.2. Though all representation are essentially the same, among all possible representation we use position (u, v) and direction (s, t) parameterization (Figure 3.2(d)). Since in this parameterization the light field transform is simple linear operations.

[1] We describe light in ray optics and the wave nature of light, i.e. polarization, diffraction and interference are ignored.

[2] Radiance is the measure of the amount of radiation per surface per steradian (steradian is the SI unit of solid angle). The SI unit of radiance is $\frac{W}{sr \cdot m^2}$

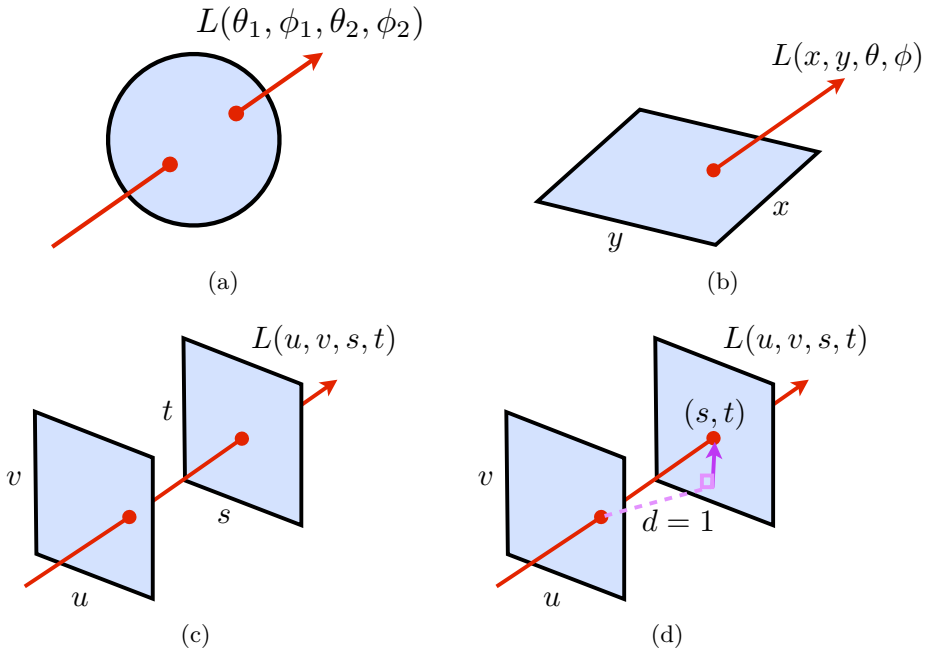


Figure 3.2: Light field representation. (a) Two-spherical point parameterization (b) A point on a surface and its direction (c) Two-plane parameterization(the light slab representation). (d) A point on a plane and tangent direction.

In Figure 3.3, we represent 2-dimensional light field using the position and direction parameterization. We observe that a ray passing through a scene point will form a line in the position and direction space. The slope of the line is the inverse of the distance of the point to the reference plane.

3.2.1 Basics in Light Field Operations

A camera consists of a number of optical elements. Mathematically speaking, a camera projects high-dimensional light fields onto a 2-dimensional image. The input light passes a number of optical elements to reach the camera sensor. Most optical devices apply a linear operation on light fields. Therefore, a camera can be defined as a set of linear transforms in the light field space. In this section, we provide an insight into various optical devices usually employed for design of a computational camera [53, 174].

1) **Space:** when the light field propagates from one plane to another parallel plane, it will shear in dimension. The shear slope is equal to the inverse distance between the two parallel planes.

2) **Lens:** focuses the rays and shears the input light fields. The amount of shear is of slope $1/f$, where f is the focal length of the lens.

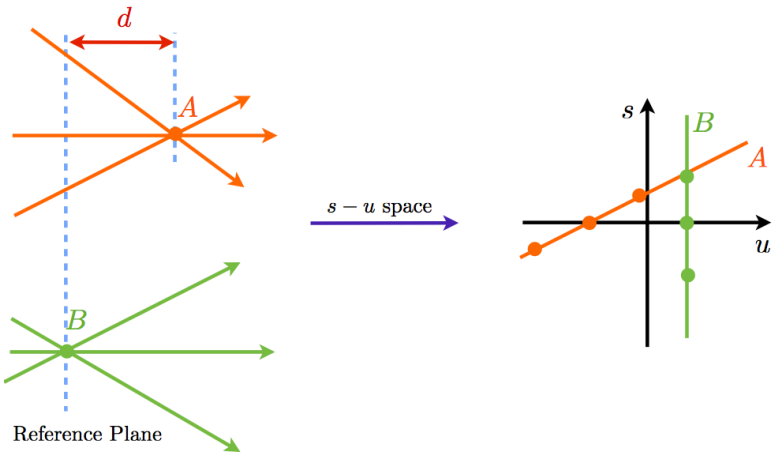


Figure 3.3: Light field representation in s-u plane. All the rays passing through a scene point becomes a line in s-u space with slope equal to the inverse distance of the point from the reference plane.

3) **Prism:** deviates the incoming rays by angle $\theta = \alpha \cdot (n-1)$, where α is the angle of the wedge of prism and n is the refractive index of the glass. In the $s-u$ space a prism translates light fields along the s dimension.

4) **Diffuser:** scatters light rays. Diffusers of various scattering patterns can be produced by changing their holographic profile. In $s-u$ space, a diffuser acts as a convolution in the s dimension and the convolution kernel depends on the scattering pattern of the diffuser [175].

5) **Intensity Modulator:** attenuates the intensity of light rays. Intensity modulators can be made from many materials such as photomasks [95], liquid crystal display (LCD) [98], liquid crystal on silicon (LCOS) [113] and digital micromirror devices (DMD) [43]. The color filter is a type of intensity modulator which attenuates wavelengths. An intensity modulator in $s-u$ space performs dot product in the u dimension.

3.3 Coding Strategies for Computational Light Field Cameras

The design space of computational cameras is large and their design criterion includes performance and complexity of cameras. There is no unique design criterion for computational cameras. The optical design of computational cameras is classified into different approaches. We are going to explain each design strategy for computational light field cameras.

3.3.1 Sensor Side Coding

In sensor side coding, an optical element is placed between the sensor and lens, therefore the sensor observes modulated light fields in both dimensions. One advantageous of the sensor side coding is that the acquisition system is compact and everything is placed inside the camera.

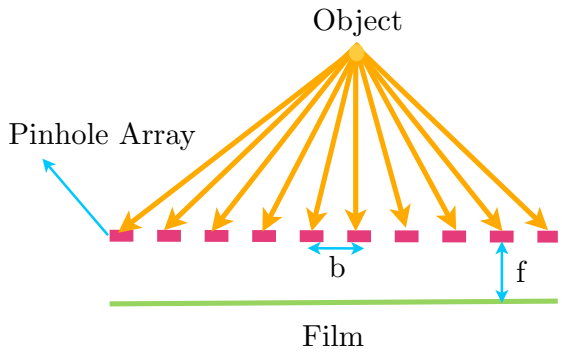


Figure 3.4: The Ives's light field camera. 1−dimensional representation of the Ives's light field camera [73]. This camera consists of an array of pinholes in front of a conventional camera to capture different light rays.

The first light field camera known as Process of Making Parallax Stereograms was proposed by Ives [73] in 1903. This light field camera consists of an array of pinhole cameras with the same focal distance which is placed at the focal plane of a conventional camera. Figure 3.4 represents the schematic of the Ives's light field camera. In the Ives' light field camera, the light rays are project through the pinhole array to different position on the film, therefore the pinhole array is an ideal ray separator. Let $L(x,\theta)$ denotes a 2-dimensional light field represented by the two-plane parametrization. The aperture of a camera allows only those rays that pass through the aperture to enter the camera. The light field after the aperture is given by

$$L_A(x,\theta) = L(x,\theta)\,A(x,\theta), \tag{3.1}$$

where $A(x,\theta)$ is the optical transfer function of the aperture. For the Ives' camera the aperture is an array of pinholes, therefore $A(x,\theta) = \sum_{n=-\infty}^{\infty}\delta(x-nb)$ where b is the distance between pinholes. The light field sampled after the pinhole array reads

$$L_A(x,\theta) = L(x,\theta)\sum_{n=-\infty}^{\infty}\delta(x-nb). \tag{3.2}$$

We assume that the light field is band limited, i.e. $\hat{L}(x,\theta) = 0\ \forall|\omega_x| \geq \omega_{x_0},\ |\omega_\theta| \geq \omega_{\theta_0}$. The Fourier transform of the light field after passing through the pinhole array reads (more

detail in [55])

$$\hat{L}_A(\omega_x, \omega_\theta) = \frac{1}{b} \sum_{n=-\infty}^{\infty} \hat{L}(\omega_x + n\frac{2\pi}{b}, \omega_\theta). \tag{3.3}$$

Therefore the light field after the pinhole array is the replications of the original light field shifted by $n\frac{2\pi}{b}$. These rays arrive at the film after traveling the focal length

$$\hat{L}_A(\omega_x, \omega_\theta) = \frac{1}{b} \sum_{n=-\infty}^{\infty} \hat{L}(\omega_x + n\frac{2\pi}{b}, \omega_\theta + f\omega_x). \tag{3.4}$$

Therefore the light field is sheared along angular frequency. This process is similar to the principle of modulation in telecommunications where a base band signal is modulated to transmit over a communication channel. The receiver demodulate the signals to recover the base band signal. In essence, the Ives' camera follows the same principle. Since the film only response to zero angular frequency, it captures only the thin slice along the intersection of the light field spectrum with the ω_x axis. Therefore, the process of the demodulation is to rearrange the frequency response of the sensor to reconstruct the original light field. Finally, the recovered 1-dimensional Fourier coefficients of the sensor is reshaped to the 2-dimensional Fourier coefficients and an inverse Fourier transform is applied to recover the light fields. The light field acquisition by the Ive's light field camera is shown in Figure 3.5.

Lippmann introduced a light field camera by replacing the pinhole array of the Ive's camera with a lens array [99]. In contrast to a pinhole a lens captures more light, therefore the Lippmann camera, called Integral Camera, provides higher quality images. Figure 3.6 show the diagram of an integral camera. In this diagram, L_1 is the distance between the sensor and the microlens array and L_2 is the distance between the microlens array and the main lens. The light rays incident on the image sensor based on their incident angle. The position, size and focal length of the microlens array controls the angular and spatial resolution of acquired light fields.

Ng et al. [115] introduced a specific design for Lippmann camera known as Lytro (see Figure 3.7) where they chose the distance between the microlens array and the sensor (L_1) to be equal to the microlens focal length f. To minimize the pixel waste the f-number of the main lens is equal to the f-number of the microlenses. In Lytro, there is a trade-off between the angular and spatial resolution and the spatial resolution is sacrificed for angular resolution.

The Raytrix camera [1] has successfully implemented a plenoptic camera that achieves higher resolution than Lytro. The microlens array has hexagonal pattern to increase the density of microlenses and decrease the pixel waste. Furthermore they use microlenses of three distinct focal lengths to improve the spatial sampling of the scene, therefore the Raytrix camera achieves higher spatial resolution comparing to Lytro.

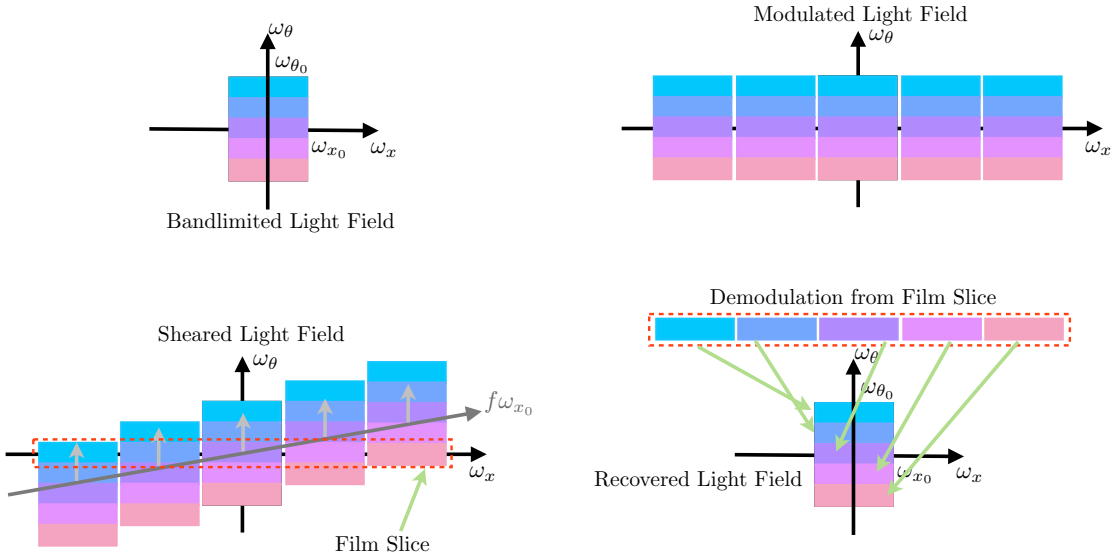


Figure 3.5: Spectral Reconstruction in Ives' light field camera. Top right: bandlimited light field of the scene. Top left: spectral modulation of the light field after the pinhole array. Bottom right: light field spectal at the film, the distance f between the pinhole array and the film shears the spectral by $f\omega_x$. Bottom left: spectral reconstruction of the light field from the film observation. The reconstruction consists of the re-assembling the light field spectrum and implying inverse Fourier transform to recover the original light field.

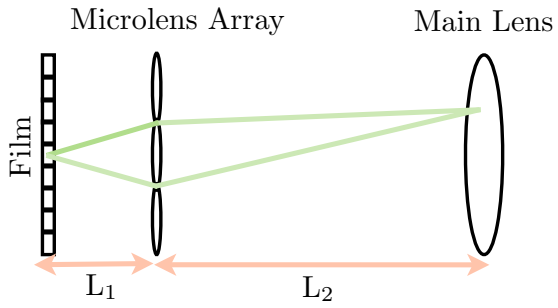


Figure 3.6: Integral Camera. The Lippmann camera consists of a main lens and an array of microlenses in front of the sensor to project different images corresponding to different angular views on the film.

Another extension of the Ives' camera is to replace the pinhole array by a mask in front of the sensor. Then light fields are modulated by the mask and sheared to arrive at the sensor due to the distance between the mask and sensor. Veeraraghavan et al. [158] proposed to use a mask which is sum of several cosine functions. Similar to the analysis of the Ives' camera, the mask modulates light fields into several identical copies of light fields in front of the sensor, i.e. the mask behaves similar to the pinhole array of the Ives' camera.

Raytrix Lytro

Figure 3.7: Comercial light field cameras. Left: Raytrix camera achieves higher spatial resolution by using three layers of microlenses. Right: Lytro sacrifises the spatial resolution to increase the angular resolution.

Marwah et al. [107] proposed a light field camera design similar to Lippmann's camera where they replace the microlens array with a mask. The mask weights each light field view and the weighted views are averaged by the camera sensor. The light field views are recovered from the coded sensor image using a light field dictionary learned on a set of known light fields. Marwah's light field camera is the first hand-held light field camera that employs the full sensor resolution and requires a single shot to capture a light field of 5×5 views. However, their model is limited to small baseline and number of views since it is practically impossible to learn a light field dictionary when the baseline is large or light fields have many views.

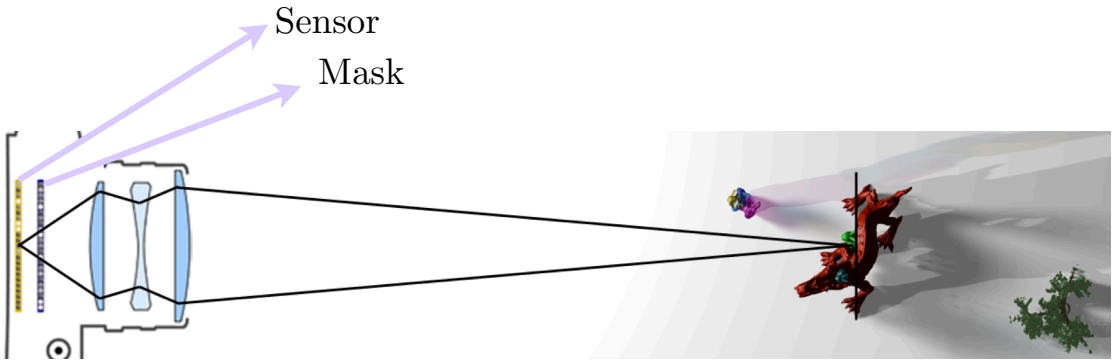


Figure 3.8: Light field camera using single coded projection. Marwah et al. [107] introduced a light field camera that captures light field from single coded projection of the scene. Unlike the Lippmann's camera, this light field camera does not trade spatial resolution for angular resolution

3.3.2 Coded Aperture

An optical element is placed close to pupil plane of a camera to code its aperture. Therefore the coded aperture modulates the Point Spread Function (PSF) of imaging systems. The PSF of pupil coding scheme using Fourier optics (incoherent light) is

described by Fresnel transform [58, 174] as follows

$$k(x) = |\mathscr{F}(W(x)Q(x))|^2, \tag{3.5}$$

where $\mathscr{F}(\cdot)$ is the Fourier transform, $W(\cdot)$ is the aperture coding mask, $Q(\cdot)$ is a focus dependent term, and $k(x)$ is the PSF function. The captured image (blurred image) is the convolution between the sharp image and the PSF. In coded aperture, since the PSF is known one can use demodulating techniques such as sparse assumption on the gradient of the sharp image to recover the latent sharp image [96].

Traditional cameras project scenes onto a 2−dimensional sensor that integrates angular information. However, if one could modify the aperture such that it blocks all rays except those from a specific region of the aperture, then the angular information is preserved. Liang et al. [98] proposed a light field camera that blocks undesired light rays and captures rays from specific regions of the aperture at each time (see Figure 3.9). Therefore, unlike the light field camera based on Lippmann model, this camera does not trade spatial resolution for angular resolution. However, the scene needs to be static while the camera captures rays passing through the different areas of the aperture. To acquire a light field with n views, the camera requires n exposures using a programmable aperture. This camera multiplexes light fields at each exposure to increase the light exposure and decrease the acquisition time, then the light field views are recovered from the linearly multiplexed light rays.

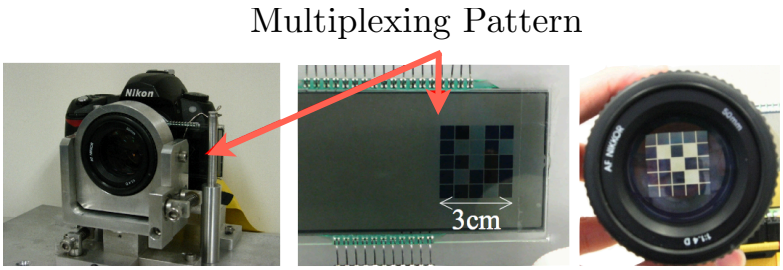


Figure 3.9: Coded aperture light field acquisition [98]. Using a programable mask (LCD or a scroll of paper patterns) behind the aperture to block specific region of the aperture to avoid angular integration by the sensor.

3.3.3 Object Side Coding

In the object side coding approach, some optical elements are attached to a traditional camera. Since the optical surface is not homogeneous, the coding strategy results in spatial modulation of light fields. The combination of traditional cameras and object side coding approach provides extra visual information of the scene using multiple observations.

Light field cameras based on the integral camera require the arrangement of microlenses

in front of cameras sensor. Georgeiv et al. [54] introduced a light field camera using an array of prisms in front of camera main lens as shown in Figure 3.10. Each prism has different angle of deviation, therefore the prisms divide FOV into multiple regions and the camera sensor observes an array of virtual images each corresponding to different view points projected by the prisms. This light field camera similar to Lippmann's camera scarifies the spatial resolution to increase the angular resolution. We should note that in practice a negative lens in front of each prism is used to increase the FOV.

Figure 3.10: Light field camera using an array of prisms and negative lenses [54]. The prisms have different deviation angle to project an array of virtual images corresponding to different angular views on the sensor.

3.3.4 Camera Arrays

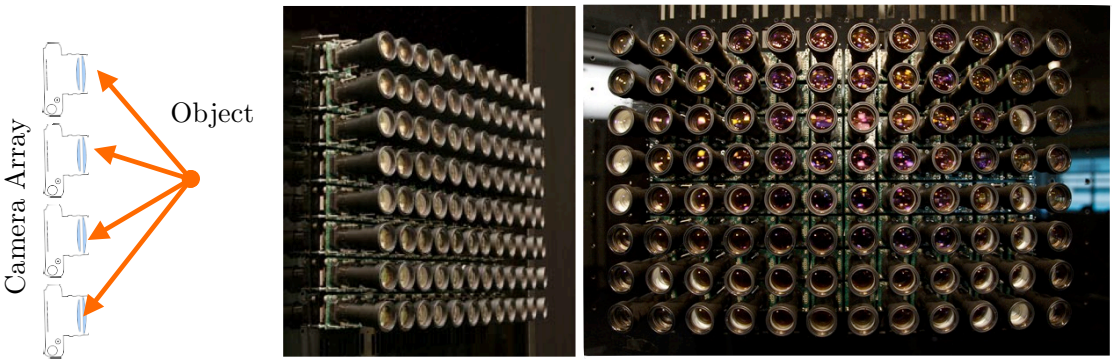


Figure 3.11: Camera array. Camera array with overlapping FOV for light field acquisition [167].

A single camera is used to capture the input light field from a fixed point-of-view. Therefore, it is intuitive to capture the light fields by moving a single camera through a frozen scene. However, the advent of inexpensive image sensors has allowed to propose light

field acquisition systems that use a number of low-cost small cameras to capture more visual information [143, 167].

Wilburn et al. [167] have design a high performance camera array where all cameras have overlapping Field-of-View (FOV) and each of them capture an slice of 4-dimensional light fields (Figure 3.11). Moreover, the camera array is used for High Dynamic Range (HDR) panoramic video, synthetic aperture photography, high-frame-rate video capturing. In contrast to previous hand-held light field cameras, the camera array provides higher FOV, however due to the cost and size of a camera array, it impossible to use it for many practical applications.

3.4 Discussion

We have explained different approaches for light field acquisition. These approaches are either sacrifice the spatial resolution to acquire angular resolution or they assume the scene is static for sequential acquisition. Marwah's light field camera [107] was the first camera that does not trade spatial resolution for angular resolution. However, the model is limited to small baselines and require a heavy dictionary learning stage.

There are strong correlations between light fields dimensions, for example, the correlations in angular and time have not been deployed by any of the discussed acquisition schemes. Therefore, instead of learning a dictionary for light fields one can use these correlations to develop efficient computational light field acquisition systems.

Chapter 4

Tensor Low-rank and Sparse Light Field Photography

4.1 Introduction

The plenoptic function [2] was introduced as a ray-based model for light that encompasses all visual information: spatial, angular, and temporal light variation as well as the color spectrum. So what makes it hard to design a camera that captures the plenoptic function in a single image? The shear amount of required plenoptic samples makes this a "big data" problem with extreme challenges for camera optics, sensor electronics, and computation.

However, high-dimensional visual signals are highly redundant. Compression algorithms, for instance, exploit this fact to minimize the memory footprint of images and videos. Moreover, recent proposals have shown that images [42], videos [68], and light fields [107] can be recovered from only a few measurements using sparsity-constrained optimization. In this paper, we present a new mathematical framework for efficient high-dimensional visual signal processing, acquisition, and storage. We demonstrate that there is a large amount of correlation between the dimensions of time-varying light fields, which can be exploited by a low-rank prior applied to the five-dimensional tensor space containing spatial, temporal, and angular light variation. This prior is a good model for exploiting view-independent and slow-moving scene parts whereas an additional sparse term captures view-dependent effects and fast motions.

We also propose a light field camera design that is well-suited for capturing coded projections of the plenoptic function that can be reconstructed by the proposed algorithms. In contrast to existing, dictionary-based light field capture systems [107], our tensor low-rank and sparse light field recovery does not require a learning phase, which is computationally expensive (tens to hundreds of hours). Further, the LRSP prior is flexible

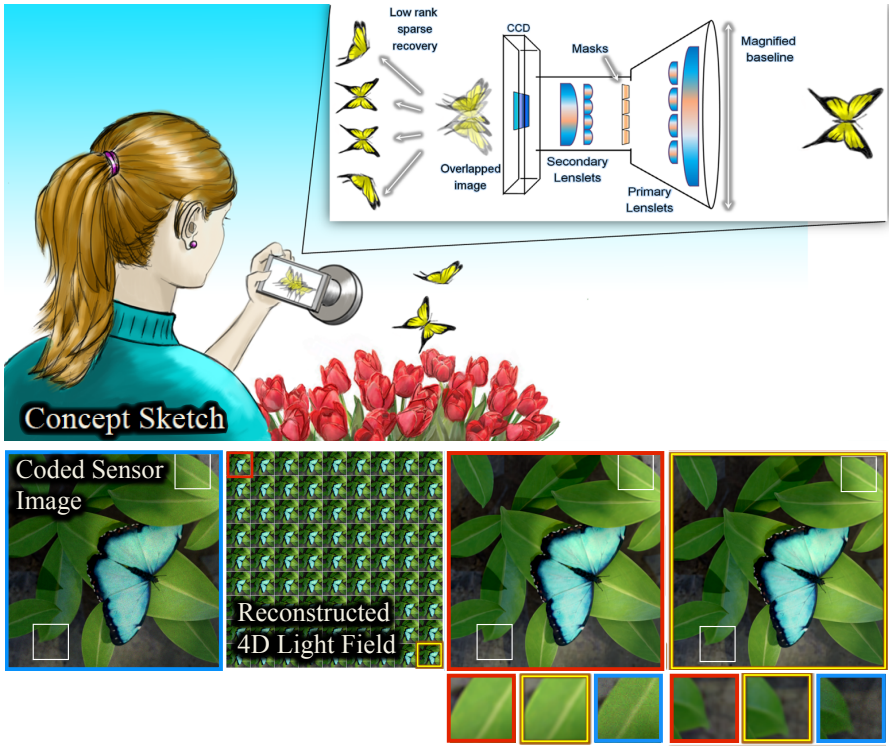


Figure 4.1: Light fields exhibit a tremendous amount of correlation within their spatial, angular, and temporal dimensionality. We propose a new tensor low-rank and sparse (LRSP) prior to model these correlations. The proposed LRSP prior is especially useful for compressive light field photography, where a high-dimensional signal is recovered from only a few lower-dimensional measurements. In addition, we propose a new camera setup that is well suited for compressive light field photography. Together, the optical and mathematical methods are much more flexible than previous techniques and allow for light fields to be recovered from a single or a few shots.

enough to be applied to a wide range of scenes and optical systems. In particular, we make the following contributions:

4.2 Related Work

Compressive Computational Photography Compressive sensing has been employed in image and video acquisition [10, 106, 135, 136, 159, 161], but these approaches are often depth-dependent. The compressive rendering scheme [138, 139] exploits sparsity of the entire light field in Fourier transform to address an inpainting problem. Compressive sensing for camera arrays has been addressed, but either requires the knowledge of disparity [69] or the method relies on image alignment [120]. Measurements taken with most light field cameras contain aliasing; a variety of approaches has recently been proposed to exploit this optical effect using advanced reconstruction algorithms to im-

prove image resolution [16, 21, 103, 108, 123, 140, 149, 160, 163]. All of these approaches use some type of super-resolution algorithm that uses linear optimization to reconstruct a higher-resolution light field and cannot be used for compressive light field acquisition. Nonlinear, sparsity-constrained approaches have also been proposed [4, 6, 171]. Most recently, these have been shown to recover high-resolution light fields from a single coded image [107]. This method however is based on a computationally expensive 4D light field dictionary learning stage which is extremely slow and memory intensive; extending this dictionary learning to 5D space-time-angle light field patches is currently intractable.

We introduce tensor low-rank and sparse light fields as a computational photography architecture that generalizes compressive light field photography. The proposed techniques are more flexible than previously-proposed dictionaries [107], we show that temporal variation and other plenoptic dimensions [2] can easily be integrated into the proposed framework, and we demonstrate our techniques with a prototype compressive light field video camera.

Multilinear Methods in Computer Graphics The global structure of multilinear datasets is exploited either by modeling it in matrix format, for instance in image alignment [122], video denoising and background subtraction [29, 78] or using tensor algebra in multilinear image-based rendering [157], BRDF [93] representation, multispectral reflectance field acquisition with a light stage [3] and subsurface scattering [121] acquisition as well as 3D display [166]. The method most closely related to ours is [3], where low-rank and sparse priors are employed for efficient capture and recovery of lighting- and wavelength-dependent material reflectance properties with a multi-spectral light stage. While similar in spirit, we address a completely different application—light field video capture—which uses a vastly different optical setup and we also employ a different formulation for low-rank and sparse tensor factorization. The proposed methods facilitate novel applications in computational optics and photography.

4.3 Background on Tensor Algebra

Tensors naturally arise in many multi-dimensional problems where data are indexed by several variables, for example in a hyperspectral cube is index by three variables; a video sequence is indexed by two spatial variables and one temporal variables; an in-depth survey on tensor related applications can be found in [87]. This section briefly reviews some concepts of tensor algebra more details are included in Appendix A. A *tensor* is a multi-dimensional array of data which is the generalization of matrices to higher dimensions. The number of dimensions of tensors is called mode or order. A mode N tensor is denoted as $\mathscr{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$. A vector is a mode-1 tensor and a matrix is a mode 2 tensor.

4.3.1 Tensor Low-rank Approximation

In literature different notion of rank for tensors have been introduced. One analogous of tensor decomposition to matrix decomposition is CANDECOMP (CP) [34, 87]decomposition which factorizes the tensor into sum of rank one tensors and the rank is equal to the minimum required rank one tensors to form the factorization. For a mode-N tensor $\mathscr{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$ the CP decomposition is defined as

$$\mathscr{X} = \sum_{i=1}^{R} \mathbf{u}_i^{(1)} \circ \mathbf{u}_i^{(2)} \circ \cdots \circ \mathbf{u}_i^{(N)}, \tag{4.1}$$

$\circ$ represents the vector outer product. The CP decomposition is NP-hard, therefore the nuclear norm of a tensor is not tractable. However for a fixed rank, we can factorize a tensor into CP components by *Alternating Least Square* (ALS) [87].

The other tensor decomposition is called Tucker decomposition [87] which factorizes a tensor into a core tensor $\mathscr{C}$ and a set of factor matrices $\mathbf{U}^{(i)}$

$$\mathscr{X} = \mathscr{C} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}. \tag{4.2}$$

Here, $\times_i$ is *tensor-matrix product*. The Tucker decomposition is also known as *higher-order SVD* [92] (HOSVD). The HOSVD is a generalization of matrix SVD to higher order tensor decomposition. The HOSVD decomposition applies matrix SVD to tensor unfolding along each mode. The HOSVD-rank of a mode-N tensor $\mathscr{X}$ is a N-dimensional vector whose i-th entry is the matrix rank of mode-i unfolding of the tensor

$$\text{rank}_{(N)} = (\text{rank}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \ldots, \mathbf{X}_{(N)}). \tag{4.3}$$

The mode-i tensor unfolding metricize a tensor into a matrix along mode-i. Unlike CP decomposition, the Tucker decomposition is easy to calculate but one needs to define a rank for each tensor mode. However, the CP decomposition using ALS approximates the tensor with a certain number of rank one tensors independent of the structures of the different tensor unfoldings.

The definition of n-rank motivates a convex model for low-rank tensor approximation based on the minimization of the sum of nuclear norms obtained from tensor unfoldings [52, 102]. However, Mu et al. [112] show that the sum of nuclear norms for tensor unfolding does not represent the tensor structure and this model is similar to nuclear norm minimization of the matrix shaped from the tensor unfolded along just one mode.

Another convex surrogate for tensor approximation is the *square norm*, which is matrix reshaping of a tensor unfolding [112]. However, unlike nuclear norm which is the tightest convex envelop to matrix rank [131], the square norm for tensor rank is not the tightest convex envelop to tensor rank.

The square norm reshapes a tensor into a matrix such that the produced matrix is more balanced (square) and also preserved the low-rank property of the tensor. Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$ and $j \in [N] \triangleq \{1, 2, \ldots N$. Then the matrix $\mathcal{X}_{[j]}$ is defined as

$$\mathcal{X}_{[j]} = \text{reshape}\Big(\mathcal{X}_{(1)}, \prod_{i=1}^{j} n_i, \prod_{i=j+!}^{N} n_i\Big). \tag{4.4}$$

$\mathcal{X}_{[j]}$ is generalization of the tensor unfolding. When $j = 1$, $\mathcal{X}_{[j]}$ is nothing but $\mathcal{X}_{(1)}$. However, for $j > 1$, $\mathcal{X}_{[j]}$ becomes a more balanced matrix. Therefore, $\mathcal{X}_{[j]}$ is more balanced matrix and also preserves the low-rank property of the tensor [112]. We assume $\mathcal{X}$ has the same length along all modes, and we define $\mathcal{X}_{\square} = \mathcal{X}_{[\lfloor \frac{N}{2} \rfloor]}$ then $\|\mathcal{X}\|_{\square} \triangleq \|\mathcal{X}_{\square}\|_*$ is called the *square norm* of tensor $\mathcal{X}$.

4.3.2 Low-rank Tensor Recovery

Similar to matrix completion explained in Section 2.6.2, we can define tensor recovery from a set of linear measurements. Given a linear map $\mathcal{A} : \mathbb{R}^{n_1 \times \cdots \times n_N} \to \mathbb{R}^m$ and the measurement vector $\mathbf{b} \in \mathbb{R}^m$, we look for a low-rank tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N} \to \mathbb{R}^m$ that fulfills the linear measurements $\mathbf{b} = \mathcal{A}(\mathcal{X}) + n$. As explained, the notion of tensor rank is not unique.

The definition of $\text{rank}_{(N)}$ motivates to recover a low-rank tensor using tensor $\text{rank}_{(N)}$ as follows

$$\underset{\mathcal{X}}{\text{argmin}} \ \text{rank}(\mathbf{X}_{(i)}) \quad \text{s.t.} \quad \|\mathbf{b} - \mathcal{A}(\mathcal{X})\|_2 \leq \epsilon, \tag{4.5}$$

where $\text{rank}(\mathbf{X}_{(i)})$ is the rank of different tensor unfoldings. Similar to low-rank matrix recovery, the matrix rank is replaced by nuclear norm as the tightest convex envelop of rank. Therefore, Eq.(4.5) is redefined as

$$\underset{\mathcal{X}}{\text{argmin}} \ \sum_{i=1}^{N} \|\mathbf{X}_{(i)}\|_* \quad \text{s.t.} \quad \|\mathbf{b} - \mathcal{A}(\mathcal{X})\|_2 \leq \epsilon. \tag{4.6}$$

The low-rank tensor recovery from sum of the nuclear norms of tensor unfoldings has widely used in [3, 52, 92, 102].

Oymak et al. [119] has shown that thought it seems trivial to recover the simultaneous structures of an object by combining the convex relaxations of each structure, the recovery is not more successful than the best single regularizer. Mu et al. [112] used this proof to show that the sum of nuclear norms for tensor unfolding does not represent the tensor structure and the number of required measurements to recover the low-rank tensor is the same as number of measurements required to the tensor unfolded along

just one mode.

One could use the square norm of the tensor and recover the low-rank tensor from the linear measurements as follows

$$\underset{\mathscr{X}}{\text{argmin}} \quad \|\mathscr{X}\|_{\square} \quad \text{s.t.} \quad \|\mathbf{b} - \mathcal{A}(\mathscr{X})\|_2 \leq \epsilon. \tag{4.7}$$

4.4 Motivation

Our choice of a low-rank prior for light field, or the plenoptic function in general, is motivated by a simple insight. Light fields have smooth variation between views and frames. Thus static light fields are highly redundant in angular dimension in addition to the spatial redundancy of individual views. To further clarify the smooth behavior of light field in angular direction, we plot a 2D light field in Figure 4.2 which shows the direct link between parallax and rank. We observe that when the objects are in the focal plane (disparity equal to zero), there is no parallax so the structures in the light field are constant along the angular dimension, i.e. the rank of the 2D light field matrix is equal to 1. For objects out of focal plane, the amount of disparity is increased, however the variation between views is still smooth. Thus light fields are highly redundant along the angular direction. This means light field angular rank is small in comparison to the maximum possible rank: the number of views.

When objects move in time, the captured light field frames changes smoothly. Thus similar to the angular direction, the light fields represent highly correlated structures in time. Intuitively, the intrinsic dimensionality of light fields is significantly lower that the size of light fields and the actual information is contained within some lower-dimensional manifold. We exploit the redundancy in spatial, angular and motion of 5D light fields using a low-rank prior. The low-rank structure of static light fields is also discussed in [65, 108].

Similar to Heber et al. [65], one way to exploit the light field low-rank structure is to reshape views into column vectors and concatenate them in the columns of a matrix. However, this model is sub-optimal, since the low-rank structure on matrix assembled from the light field views promotes a global pattern and cannot consider different degree of freedom of individual dimensions of the light field. We preserve the original structure of light fields by representing 5D light fields with 5D tensors which independently models the correlated structure of each dimension.

In practice, we reduce the computational cost of our light field structure modeling scheme by working in parallel on independent 5D light field patches. The size of light field patch depends on the amount of parallax and motion in the scene. For a fixed patch size, increase in the amount of parallax and motion decrease the similarity between

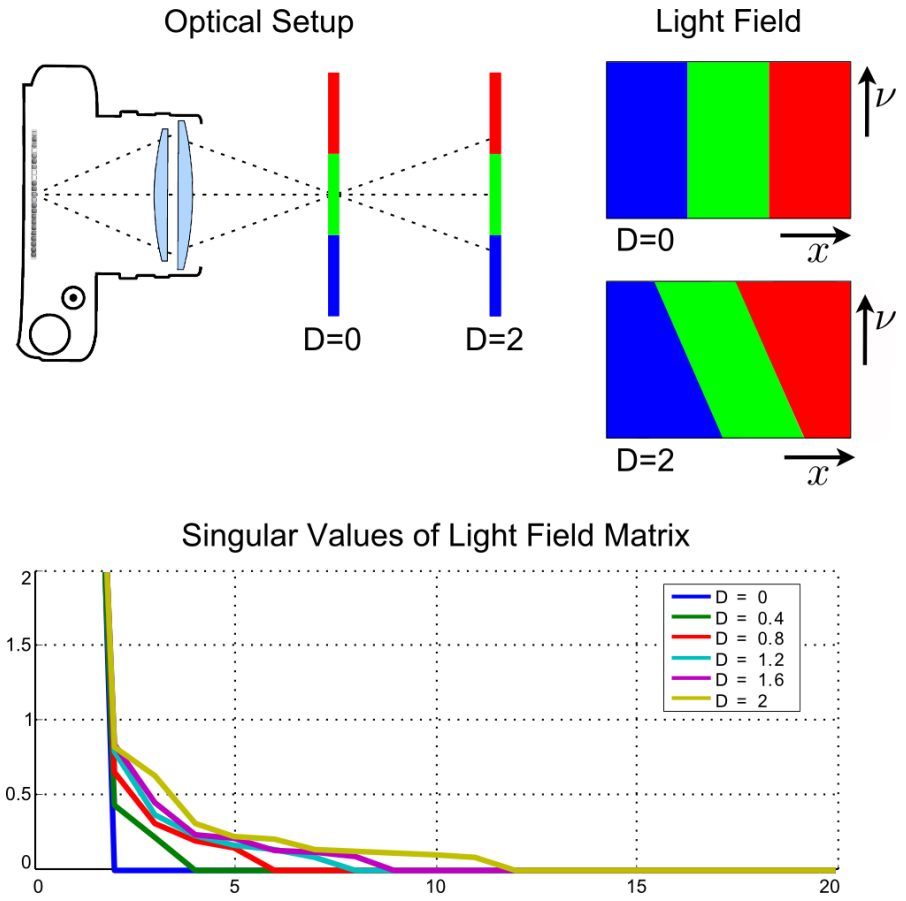


Figure 4.2: The amount of parallax in a light field (top) is directly related to its rank (bottom). The same is true for motion in the temporal domain (not shown); the proposed tensor low-rank prior models this in a unified manner for high-dimensional visual signals.

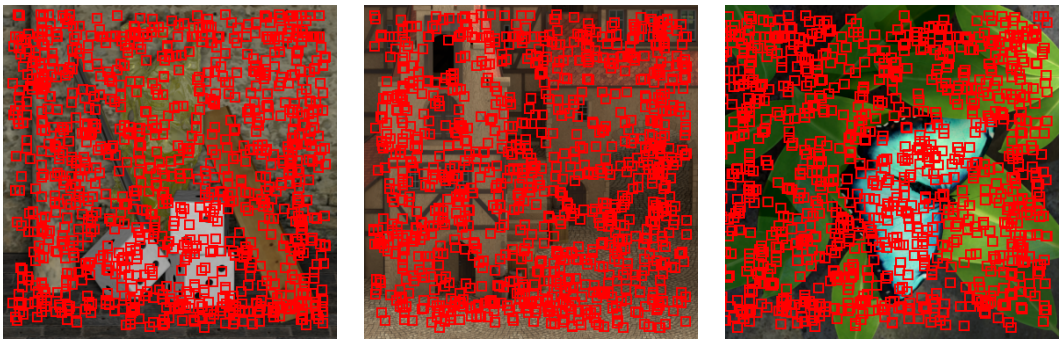
views. Thus, one needs to adapt the spatial size and the number of frames grouped in light field patches with respect to the amount of parallax and motion so that the correlations in light field patches are preserved. Therefore, the patch size can influence on the maximum tolerated parallax and motion modeled by our proposed scheme.

4.5 Low-rank and Sparse Light Field Tensors

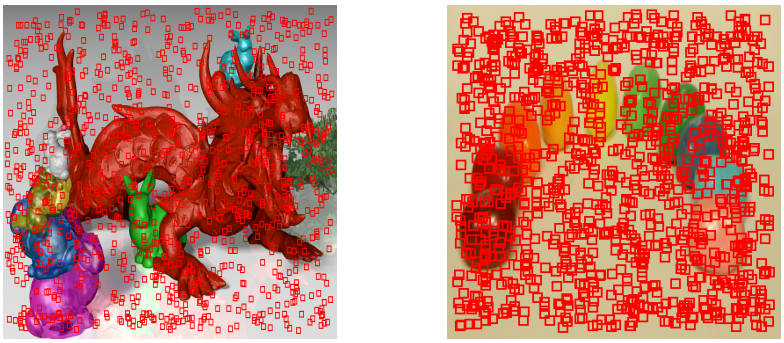
4.5.1 Which Tensor Low-rank Model

Unfortunately, as explained, the notion of a high-dimensional singular value decomposition (SVD) is not clearly defined. In order to choose, the best tensor low-rank approximation, we compared all mentioned tensor low-rank approximations w.r.t. speed and quality. For the comparison, we randomly select 1000 patches from each light field, then we compare different tensor low-rank approximations on a tensor completion scenario

where a random subset of 10% of pixels from each patch is selected. Figure 4.3 shows the employed light field datasets and the corresponding selected patches. Table 4.1 demonstrates performance of different tensor low-rank approximations.



(a) Buddha [164] (b) Medival [164] (c) Butterfly [164]



(d) Dragon [107] (e) Jelly Beans [167]

Figure 4.3: Light field datasets employed for comparison of different low-rank tensor model. A sub set of 1000 light field patches from each dataset is selected to choose the best tensor low-rank model.

Low-rank Model	Performance [dB]	Speed [sec]
	LR	LR
CP	33.83	18.80
HOSVD	31.61	2.34
Square norm	25.40	1.53

Table 4.1: Average performance of different low-rank tensor models. The comparison is based on inpainting with a compression ratio of 10 applied to a random selection of 1000 light field patches. We observe that CP outperforms other tensor low-rank schemes, but is also slower.

We choose to work with CP as a general low-rank tensor model. This not only provides the best quality but is also more flexible than HOSVD, because CP is oblivious to the actual dimension where the signal is low rank. HOSVD on the other hand requires a specific rank to be assigned to each dimension a prior.

4.5.2 Why Low-rank and Sparse Decomposition?

Light fields cannot always be perfectly represented as low-rank tensors. For example, a large amount of parallax or motion, specularity, reflections, and the noise of acquisition devices can distort the similarity between views. However, the distortion has sparse structure. This argument is supported by Figures 4.4 and 4.5, where the remainders between target light fields and rank-6 CP decompositions are shown. Inspired by robust Principle Analysis (RPCA) [29], we decompose the light fields into the low-rank and sparse components to improve the reconstruction performance. RPCA with exactly the same principle is employed in various computer vision problems (e.g., [78, 122]).

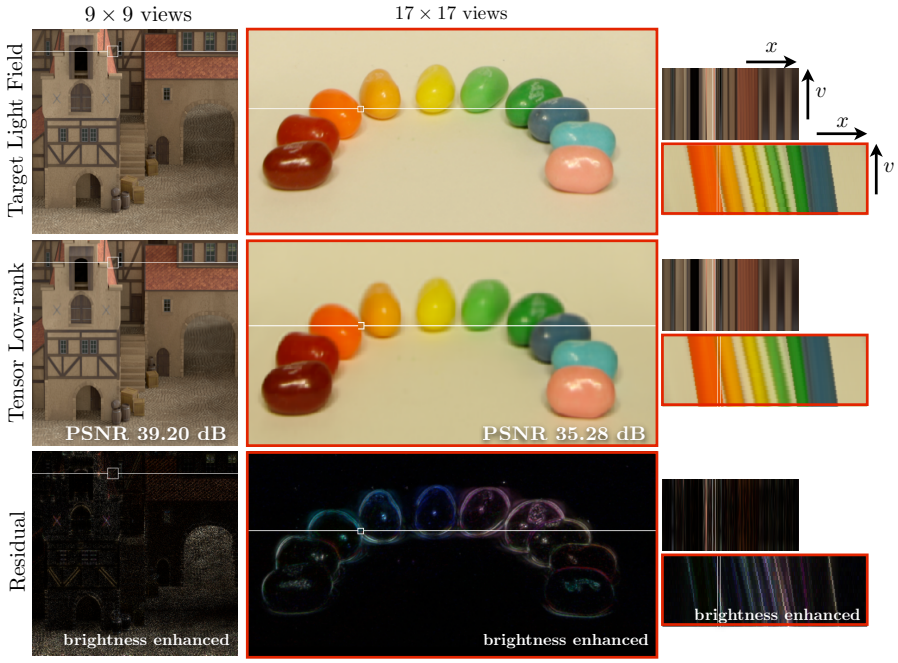


Figure 4.4: Tensor low-rank approximations of two datasets. The proposed low-rank model is, as opposed to previously-employed dictionaries, flexible enough to apply to datasets with a different number of views (here 9×9 and 17×17 views) or other parameters. The coherence within the light fields is exploited by the low-rank prior—even for a relatively large amount of parallax, reconstructions are faithful. The remaining error is concentrated around depth discontinuities and therefore sparse. This observation motivates our choice of a combined tensor low-rank and sparse framework for light field photography.

We compare a variety of possible choices for low rank and sparse priors in Figure 4.6 and conclude that CP as a tensor low-rank model combined with a discrete cosine transform-based sparsity prior is the best choice among the ones tested. So why not simply use the dictionaries of light field atoms for 5D light field videos? Dictionaries for high-dimensional visual data have a lot of advantages, but also two major disadvantages: a) they require a dictionary learning phase and b) the dictionary only models structures

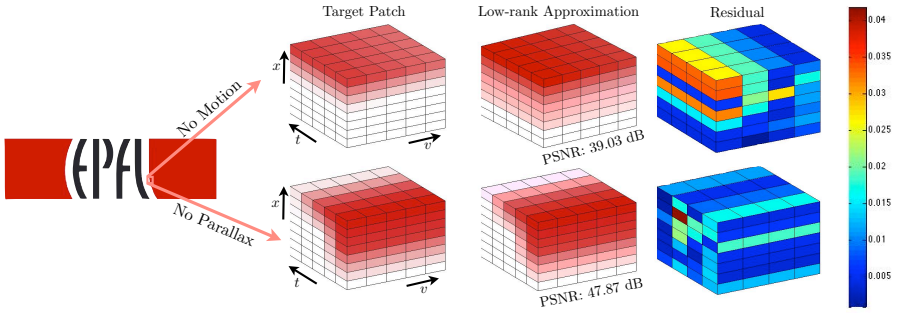


Figure 4.5: Low-rank approximation of a single light field video patch. If there is no motion and little parallax (top), the patch can be well represented by a low-rank tensor. The same is true for an in-focus object that has no parallax and some motion (bottom).

well that were also part of the training data. For a), compute times can be extremely long (tens of hours or days for a 4D light field) or even prohibitive (we were not able to learn dictionaries on 5D light field videos). Whereas a large set of training light fields should be sufficient to learn a good dictionary, oftentimes not all possible configurations of f-number settings, scene types, depth ranges, or other scene properties are available to learn from. Hence, the proposed model is more tractable and also more flexible.

4.6 Light Field Acquisition and Synthesis

4.6.1 Coded Light Field Acquisition

A video $y(x,t)$ for a conventional camera sensor is formed by integrating the incident, time-varying light field $l(x,v,t)$ over its angular domain Ω_v as

$$y(x,t) = \int_{\Omega_v} l(x,v,t)\, dv. \tag{4.8}$$

In this formulation, vignetting and other angle-dependent effects are absorbed by the light field. Whereas the recorded video y varies over the sensor surface[1] x and over time t, all angular variation of the light field is irreversibly lost.

Light field photography is concerned with the design of camera systems that preserve the desired angular information optically, such that it can be recovered using computation. In the most general sense, the optical image formation can be expressed as a convolution along the plenoptic dimensions

$$y(x',t') = \int_{\Omega_x}\int_{\Omega_v}\int_{\Omega_t} l(x,v,t)\, m(x-x', v, t-t')\, dx\, dv\, dt. \tag{4.9}$$

[1] We consider a single dimension x and v in both space and time, respectively. Extensions to the full 4D case are straightforward.

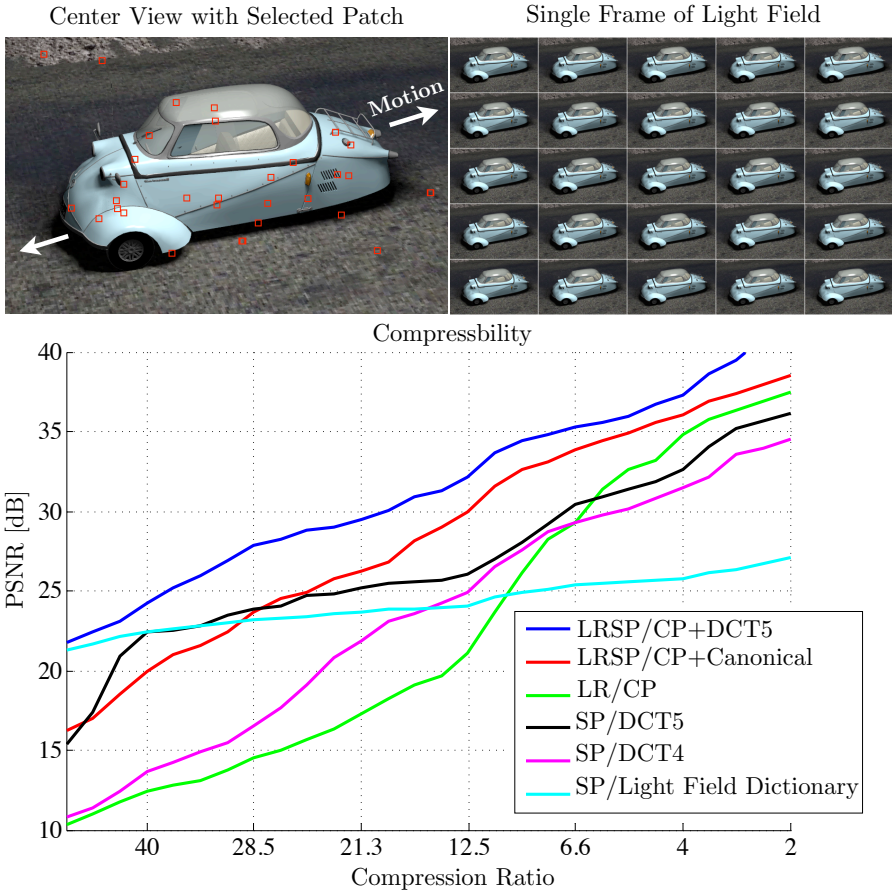


Figure 4.6: Compressibility of light field video. We randomly sample a set of 5D light field patches of size $9 \times 9 \times 5 \times 5 \times 5$ (top left) and solve an in-painting problem to compare how well different priors perform. The simulation is run 20 times and the resulting peak signal-to-noise ratios (PSNR) averaged. For 4D DCT (magenta plot) and light field dictionaries (cyan plot), we apply the 4D prior to each temporal slice separately. Dictionaries of light field atoms by Marwah et al. [107] are optimized for high compression ratios, so they perform best in that area but quickly fall below other priors when the conditions are different from those of the training phase. All low-rank priors and also DCT5 take advantage of correlations in space, time, and angle. Low-rank modeled by CP combined with a 5D DCT (blue plot) always performs better than any alternative approaches.

As discussed in Chapter 3, a variety of different light field acquisition schemes have been proposed, each one resulting in different tradeoffs. For the purpose of this paper, we only consider a subset of all possible convolution kernels m: the ones allowing for coded optical attenuation, but no ray mixing or refraction. This simplifies the image formation to

$$y(x,t) = \int_{\Omega_v} l(x,v,t)\, m(x,v,t)\, dv. \tag{4.10}$$

In discretized form, this is a coded projection of an order-3 light field tensor $\mathscr{L} \in \mathbb{R}^{n \times m \times p}$

to an order-2 video tensor $\mathscr{Y} \in \mathbb{R}^{n \times p}$:

$$\mathscr{Y} = \mathscr{P}(\mathscr{L}), \tag{4.11}$$

where $\mathscr{P} : \mathbb{R}^{n \times m \times p} \to \mathbb{R}^{n \times p}$ is the linear projection operator incorporating the effect of the modulation kernel m. The number of spatial, angular, and temporal samples is n, m, and p, respectively. Although $\mathscr{Y}$ is technically a matrix in this intuitive "flatland" model, in practice we work with order-5 light field tensors that are coded in order-3 video tensors. Hence, we use tensor notation for both quantities.

The specific choice of the kernel in (4.10) is unique in that it allows the measured video tensor $\mathscr{Y}$ to be subdivided into small spatio-temporal windows—neighboring windows in the light field tensor space are not linked by their angles, which could be the case for general convolution kernels (4.9).

4.6.2 Low-rank and Sparse Light Field Tensors

As discussed, light fields containing a moderate amount of parallax and scene motion can be well approximated by a low-rank prior. View-dependent effects and larger amounts of parallax or motion result in sparse remainders. Robust principal component analysis (RPCA), as introduced by Candés et al. [29], models exactly this problem and is employed in various computer vision problems (e.g., [78, 122]). We follow RPCA and represent the light field tensor as the sum of a low-rank and a sparse tensor $\mathscr{L} = \mathscr{R} + \mathscr{S}$. Then (4.11) becomes

$$\mathscr{Y} = \mathscr{P}(\mathscr{R} + \mathscr{S}), \tag{4.12}$$

where $\mathscr{R}, \mathscr{S} \in \mathbb{R}^{n \times m \times p}$ are low-rank and sparse, respectively. We allow $\mathscr{S}$ to be sparse in some transform domain $\Psi(\mathscr{S})$, which is expressed by the operator $\Psi : \mathbb{R}^{n \times m \times p} \to \mathbb{R}^{n \times m \times p}$. Motivated by the experiment shown in Figure 4.6, we use a high-dimensional discrete cosine transform (DCT) for Ψ and the CANDECOMP (CP) decomposition for $\mathscr{R}$ throughout this paper. The CP decomposition [34, 87] of an order-3 tensor $\mathscr{R}$ is the sum of vectors aligned with the dimensions of the tensor:

$$\mathscr{R} = \sum_{i=1}^{r} \mathbf{u}_i^{(x)} \circ \mathbf{u}_i^{(v)} \circ \mathbf{u}_i^{(t)}, \tag{4.13}$$

where r is the rank of $\mathscr{R}$ and $\mathbf{u}_i^{(x)} \in \mathbb{R}^n$, $\mathbf{u}_i^{(v)} \in \mathbb{R}^m$, $\mathbf{u}_i^{(t)} \in \mathbb{R}^p$ for $i = 1, \ldots, r$. Figure 4.7 schematically demonstrate our optical setup and light field tensor decomposition.

We follow general convex formulations for robust PCA and define the light field low-rank

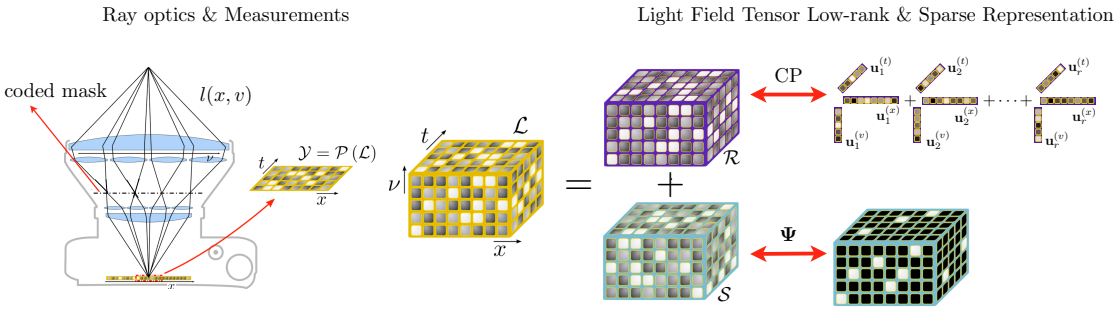


Figure 4.7: Visualization of optical setup and light field tensor decomposition. The proposed camera design optically overlays multiple images onto a single sensor (left). A patterned attenuation mask optically codes the images before they are integrated by the sensor. We represent light field videos as high-dimensional tensors (center) and formulate their reconstruction from coded sensor images as a compressive tensor low-rank and sparse recovery problem.

and sparse tensor decomposition by solving the following objective function

$$\begin{aligned} &\underset{\{\mathscr{R},\mathscr{S}\}}{\arg\min} \quad \|\Psi(\mathscr{S})\|_1 \\ &\text{to subject} \quad \|\mathscr{Y} - \mathscr{P}(\mathscr{R}+\mathscr{S})\|_2^2 \leq \epsilon, \end{aligned} \tag{4.14}$$

where ϵ is the sensor noise level and ℓ_1 norm of a tensor is defined as absolute sum of its elements.

4.6.3 Efficient Light Field Synthesis

To solve (4.14) we resort to the *parallel proximal algorithm* (PPXA) described by Combettes et al. [35]. For this purpose, the Lagrangian form of the objective is solved as

$$\underset{\{\mathscr{R},\mathscr{S}\}}{\arg\min}\, \lambda\|\Psi(\mathscr{S})\|_1 + \|\mathscr{Y} - \mathscr{P}(\mathscr{R}+\mathscr{S})\|_2^2, \tag{4.15}$$

where the inequality constraints are incorporated into the objective function. Although there exists a weight λ that corresponds to the sensor noise level ϵ, in practice neither is known exactly and needs to be approximated by a user-defined value. We list all algorithmic parameters in Section 4.8.

Parallel Proximal Algorithm

In this section, we explain more detail on PPXA and how it scales with different constraints on data. PPXA is derived from the Douglas-Rachford algorithm [35] and looks for a minimizer of sum of multiple functions. Each function f_i can be a prior constraint

on the solution or on the data acquisition scheme. This form of optimization looks for the solution of

$$\underset{\mathbf{x}\in\mathbb{R}^{\mathbf{N}}}{\operatorname{argmin}}\ f_1(\mathbf{x})+\cdots+f_n(\mathbf{x}). \tag{4.16}$$

PPXA is an iterative method at each iteration the proximity operator of all functions are calculated, which can be computed in parallel, and their results are averaged; the process is continued until convergence to a point. Algorithm 3 describes PPXA to solve (4.16).

Algorithm 2: Parallel Proximal Algorithm

Initialize:
$\mathbf{z}_1 \in \mathbb{R}^n, \ldots, \mathbf{z}_n \in \mathbb{R}^n,\ \mathbf{x} \in \mathbb{R}^n,\ \gamma > 0;$
while *not converged* **do**
  for $i = 1, \ldots, n$ **do**
    $\mathbf{p}_i \leftarrow \operatorname{prox}_{n\gamma f_i}(\mathbf{z}_i);$
  $\mathbf{p} \leftarrow (\mathbf{p}_1 + \cdots + \mathbf{p}_n)/n;$
  for $i = 1, \ldots, n$ **do**
    $\mathbf{z}_i \leftarrow \mathbf{z}_i + 2\mathbf{p} - \mathbf{x} - \mathbf{p}_i;$
  $\mathbf{x} = \mathbf{p}.$

Extension of PPXA to Tensor Low-rank and Sparse Light Field Decomposition

The PPXA algorithm solves (4.15) by iteratively computing the proximity operators for each objective term and averaging them. Similar to robust PCA (see 2.6.3), we define $\mathbb{R}_0 \triangleq \mathbb{R}^{n\times m\times p} \times \mathbb{R}^{n\times m\times p}$ by the Cartesian product of $\mathbb{R}^{n\times m\times p}$ and a point in $\mathbb{R}_0$ is defined as $\mathscr{X} \triangleq (\mathscr{X}_1, \mathscr{X}_2) \in \mathbb{R}_0$. The inner product in $\mathbb{R}_0$ is defined as

$$\langle \mathscr{X}, \mathscr{Y} \rangle \triangleq \langle \mathscr{X}_1, \mathscr{Y}_1 \rangle + \langle \mathscr{X}_2, \mathscr{Y}_2 \rangle. \tag{4.17}$$

The norm on $\mathbb{R}_0$ induced by this inner product is

$$\|\mathscr{X}\|_{\mathbb{R}_0} = \langle \mathscr{X}, \mathscr{X} \rangle = \|\mathscr{X}_1\|_{\mathrm{F}} + \|\mathscr{X}_2\|_{\mathrm{F}}, \tag{4.18}$$

where $\|\cdot\|_{\mathrm{F}}$ is the matrix Frobenius norm.

Proposition 4.6.1. *For any point $\mathscr{X} = (\mathscr{X}_1, \mathscr{X}_2) \in \mathbb{R}_0$ and a function $f(\mathscr{X}) = f_1(\mathscr{X}_1) + f_2(\mathscr{X}_2)$, the proximity operator of f is defined as*

$$prox_{\eta f}(\mathscr{X}) = (prox_{\eta f_1}(\mathscr{X}_1), prox_{\eta f_2}(\mathscr{X}_2)). \tag{4.19}$$

Proof. Proof is similar to proposition 2.6.1 in Section 2.6.3. □

Pseudo-code for solving (4.15) using PPXA is outlined in Algorithm 3. All intermediate variables $\mathscr{X}_i, \mathscr{Z}_{ij}$ are tensors of the same size as the light field and $\mathscr{X}, \mathscr{Z}_{(1,2)}, \mathscr{A}, \mathscr{A}_{(1,2)}$ represent concatenations of two such tensors. The proximity operators of our problem are

$$\operatorname{prox}_{\mathrm{CP}}(\mathscr{X}) = \underset{\left\{\mathbf{u}_i^{(x,v,t)}\right\}}{\arg\min} \left\| \mathscr{X} - \sum_{i=1}^{r} \mathbf{u}_i^{(x)} \circ \mathbf{u}_i^{(v)} \circ \mathbf{u}_i^{(t)} \right\|_2^2,$$

$$\operatorname{prox}_{\eta\|\cdot\|_1}(\mathscr{X}) = \Psi^* \left(\mathcal{S}_\eta (\Psi(\mathscr{X})) \right), \tag{4.20}$$

$$\tag{4.21}$$

where $\eta \in (0, \infty)$ is a scalar, $\mathscr{Z}$ is an intermediate slack variable, $\mathcal{S}_\eta(\mathscr{X}) = \operatorname{sign}(\mathscr{X}) (|\mathscr{X}| - \eta)_+$ is a soft-thresholding operator as explained in Section 2.1.3, and $\operatorname{prox}_{\mathrm{CP}}(\mathscr{X})$ computes the CP decomposition of $\mathscr{X}$ using alternating least squares.

To compute the proximity operator of the data term $\|\mathscr{Y} - \mathscr{P}(\mathscr{R} + \mathscr{S})\|_2^2$, we vectorize $\mathscr{Z}$ and matrices the operator $\mathscr{P}$ to matrix $\mathbf{P}$ such that $\mathscr{Y}_{vec} = \mathbf{P}\mathscr{Z}_{vec}$. The operator $\mathbf{P}$ will have block diagonal structure and each blocks of $\mathbf{P}$ contains the modulation values of a pixel in the light field patch. The proximity operator of the data term reads

$$\begin{aligned} \operatorname{prox}_{\eta\|\cdot\|_2^2}(\mathscr{X}) &= \underset{\mathscr{Z}}{\operatorname{argmin}} \ \|\mathscr{Y} - \mathscr{P}(\mathscr{Z})\|_2^2 + \frac{1}{2\eta} \|\mathscr{Z} - \mathscr{X}\|_2^2 \\ &= (\mathbf{I} + \eta \mathbf{P}^T \mathbf{P})^{-1} (\mathscr{X}_{vec} + \eta \mathbf{P}^T \mathscr{Y}_{vec}) \end{aligned} \tag{4.22}$$

As discussed in more detail by Combettes et al. [35], proximal splitting methods basically split up an objective function, such as (4.15), into a set of sub-problems, each of which can be solved conveniently by applying a simple proximity operator. Convergence of this iterative scheme is only guaranteed for a sum of convex sub-problems. Although the CP decomposition is not convex, in practice we observe quick convergence.

Algorithm 3: Low-rank and Sparse Light Field Tensor Decomposition via Parallel Proximal Algorithm

initialize $\mathscr{X} = (\mathscr{X}_1, \mathscr{X}_2)$, $\mathscr{Z}_1 = (\mathscr{Z}_{11}, \mathscr{Z}_{12})$,
$\mathscr{Z}_2 = (\mathscr{Z}_{21}, \mathscr{Z}_{22})$, $\mathscr{X}_i, \mathscr{Z}_{ij} \in \mathbb{R}^{n \times m \times p}$
while *not converged* **do**
  $\mathscr{A}_1 \leftarrow \left(\operatorname{prox}_{\mathrm{CP}}(\mathscr{Z}_{11}), \operatorname{prox}_{\eta\|\cdot\|_1}(\mathscr{Z}_{12}) \right)$
  $\mathscr{A}_2 \leftarrow \left(\operatorname{prox}_{\eta\|\cdot\|^2}(\mathscr{Z}_{21} + \mathscr{Z}_{22}), \operatorname{prox}_{\eta\|\cdot\|^2}(\mathscr{Z}_{21} + \mathscr{Z}_{22}) \right)$
  $\mathscr{A} \leftarrow (\mathscr{A}_1 + \mathscr{A}_2)/2$
  $\mathscr{Z}_{(1,2)} \leftarrow \mathscr{Z}_{(1,2)} + 2\mathscr{A} - \mathscr{X} - \mathscr{A}_{(1,2)}$
  $\mathscr{X} = \mathscr{A}$
$\mathscr{R} = \mathscr{X}_1$, $\mathscr{S} = \mathscr{X}_2$

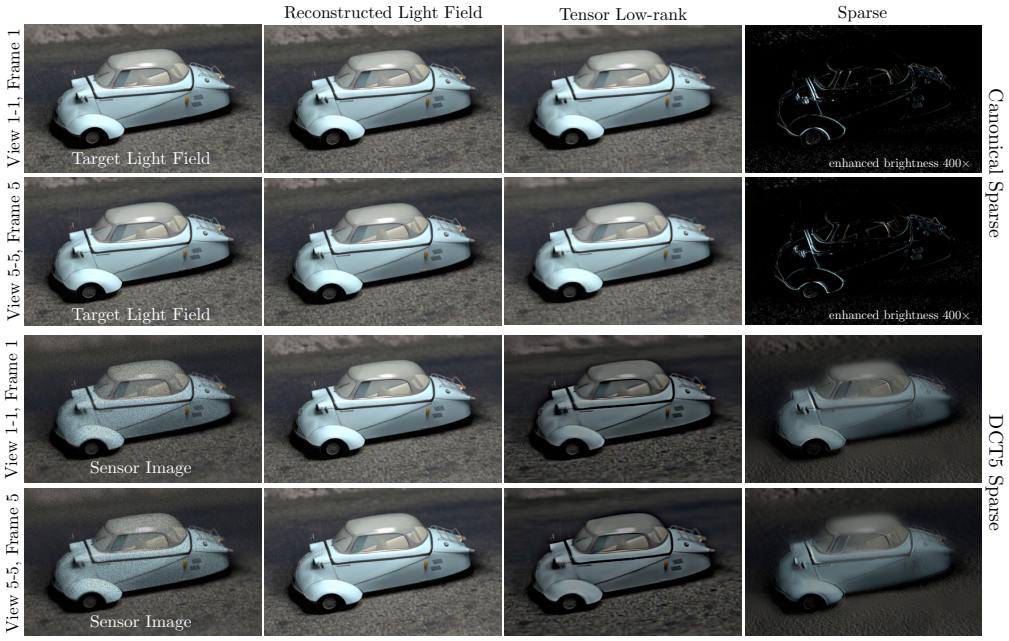


Figure 4.8: Light field decompositions. We recover a light field video with 5×5 views and 5 frames (top left, 2 frames are shown) from a coded video consisting of 5 coded sensor images (bottom left). The reconstruction algorithm operates on a patch-by-patch basis and splits the signal into low-rank (top, bottom, center right) and canonical sparse (top right) or DCT5 sparse (bottom right) components, exploiting correlations in space, time, and angle. The reconstructed light field for the first and last frames are shown (top, bottom, center left). The low-rank part contains most of the information whereas the sparse captures high-frequency details and other view-dependent effects. Decomposition using DCT5 sparse improves the performance and allows to recover areas with higher amount of parallax. In contrast to canonical sparse for DCT5 the low-rank and sparse component are not well separated since the smooth areas also have sparse representation in DCT5.

4.7 Analysis

4.7.1 Interpreting Light Field Decompositions

We show two views of a short light field video sequence and corresponding views of the low-rank and sparse components in Figure 4.8. Most parts of this scene are well-approximated by a tensor low-rank representation. Yet, the 5D discrete cosine transform helps to recover high-frequency edges and areas with larger amounts of parallax. We also illustrate the low-rank component as well as the sparse component of two selected patches. In the first patch (red), the out-of-focus high frequency structures are mostly blurred out in the low-rank component while the sparse component contains the edges. The second patch (blue) is almost constant in both angle and time, thus it is well-represented by the low-rank part.

4.7.2 What are good optical setups?

In all experiments in this paper, we use masks that contain random Gaussian patterns as modulation codes. Throughout the capture process, the codes are changed for successive frames of the video. To satisfy the restricted isometry property, employed codes should be mutually incoherent in angle and time. This property is derived for sparse priors [22] but also exists for low-rank priors [25, 129]. Basically, the optical codes should be as random as possible w.r.t. each other such that the diversity of the sampling process is maximized. In Section 4.8, we present a new compressive light field camera prototype that allows us to code all light fields views independently and which could scale to large baselines.

4.7.3 How many measurements are necessary?

It is important to understand the conditions under which low-rank and sparse components can actually be recovered. Usually this is given as the order of the number of required measurements given some properties of the projection operator $\mathscr{P}$ ("the measurement matrix"). Oftentimes, the degree of freedom of a tensor is used to derive an expression for the required number of measurements in the literature. The degree of freedom of a generic order-k tensor with rank r is $r^k + knr$, where $n = \max\{n_i; i \in [k]\}$ is the largest dimension of the tensor. The degree of freedom for a combined low-rank and sparse tensor is $d_t = r^k + knr + \|\Psi(\mathscr{S})\|_0$. Recall that $\|\cdot\|_0$ denotes the number of non-zeros elements entries in a tensor. Wright et al. [170], for instance, showed when the measurements are taken from a Gaussian random distribution, the minimum number of required measurements to decompose a matrix into low-rank and sparse components is $O\big(\log^2(n)((2nr - r^2) + \|\mathbf{S}\|_0)\big)$, i.e. $O\big(\log^2(n)d_m\big)$, where d_m is the degree of freedom of a low-rank matrix. Rauhut et al. [129] give an expression for the minimum number of required measurements to recover a low-rank tensor from linear measurements as $O\big(\log(k)d_t\big)$. Disregarding what the actual number is, it is proportional to the degree of freedom of the tensor and therefore to its rank. We experimentally verify this in the following subsection by varying parallax and motion of the tensor, hence its rank.

4.7.4 How does the Algorithm Degrade?

We evaluate reconstruction quality w.r.t. varying amounts of parallax and scene motion in Figure 4.9. All target light fields contain a resolution chart with some amount of parallax over 5×5 simulated views and motion over 5 frames. The plot shows that it is easier to recover motion than parallax (top right), which is intuitive because parallax is integrated by the sensor. Overall, relatively about (≈5 pixels) of both parallax and motion can be recovered well with the simulated setup. Nevertheless, the quality of the proposed method is still better than that achieved with light field dictionaries (lower

right). High-frequency details are successfully recovered with the proposed algorithm. For this experiment, we used the dictionary provided by Marwah et al. [107] on their project website.

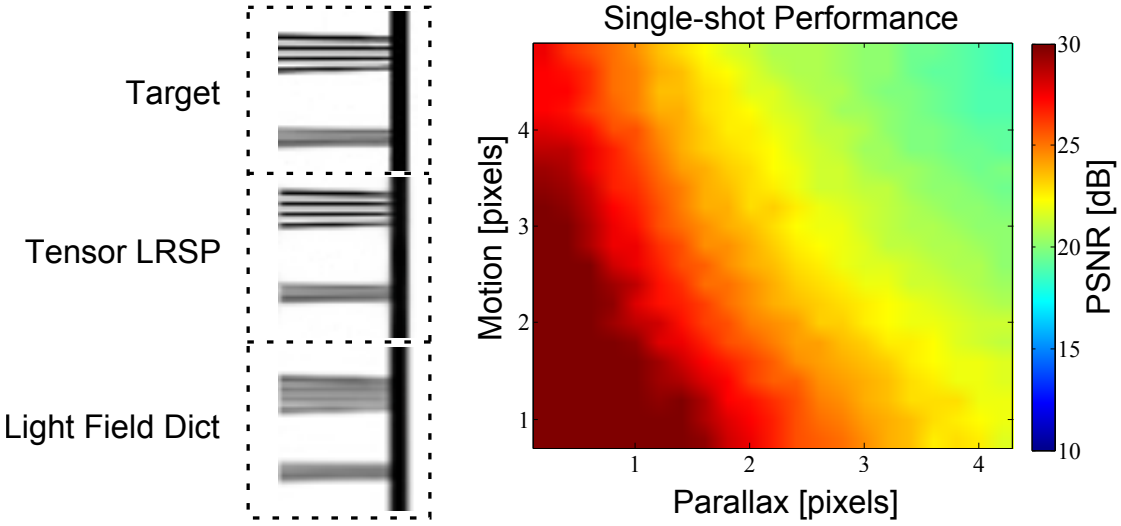


Figure 4.9: Performance analysis for a single-shot compressive light field camera. The intrinsic dimensionality of the light field depends on the amount of motion and parallax in the scene. We evaluate reconstruction quality of the proposed algorithm w.r.t. these parameters (right). As opposed to a sparse-only reconstruction with a light field dictionary (left, bottom) by Marwah et al. [107], our approach is capable of recovering high-frequency details (left, center) for a high-contrast resolution chart with observed parallax that is as much as the patch size (≈5 pixels).

4.8 Implementation

Hardware We built a proof-of-concept prototype light field camera that is optimized for the proposed algorithms (Figure 4.10). The special property of this device is that intermediate images, showing the individual views of the light field, are generated in mid air (yellow boxes). These can be independently modulated before they are optically combined on the sensor. Through the system is designed to capture light fields with 2×2 views through a lens array with a single sensor, it can be extended to more views by some modification to the design, such as use of prism sets or Fresnel plates. The array has a baseline of 6.5 cm and the optical setup consists of an entrance lens (L1, f=25 cm, D=15 cm) that feeds converging light into a set of four smaller lenslets (LL1, f=10 cm, D=2.54 cm). The image of each lenslet is projected on the mask plane (see Figure 4.10). As discussed in Section 4.7.2, the mask pattern is a random distribution and printed on a transparency with 50800 DPI by http://www.fineline-imaging.com/. In light field video acquisition, in order to increase the incoherency between consequent acquisition, we slightly displace the mask. However, one could programmed a motor

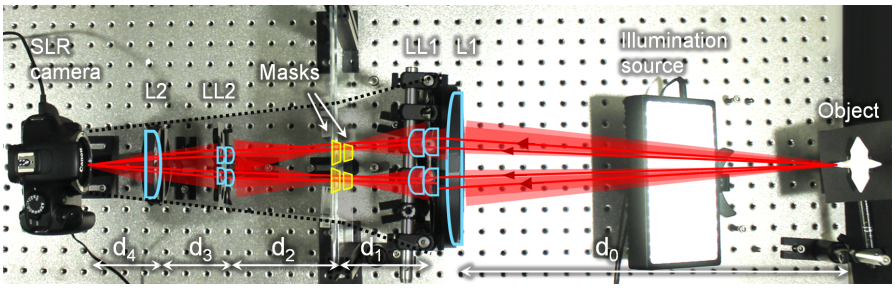


Figure 4.10: Prototype compressive camera. The device uses a relay lens system that allows for multiple images of the same scene to be multiplexed on a single sensor. The primary images are formed on the mask by L1+LL1 and these images are optically coded and re-imaged by L2+LL2. The images show the same scene over a range of viewpoints and each one is independently modulated by an optical code on a printed transparency. The parameters are: $d_0 = 45$ cm, $d_1 = 10$ cm, $d_2 = 14$ cm, $d_3 = 9$ cm, $d_4 = 8$ cm.

for the displacement. In general, LCDs could be replaced the printed mask but the resolution of LCDs are limited and the resolution of the mask imposes a limitation on the resolution of reconstructed light field. Thus, we preferred a printed mask to the available LCDs.

Each image under the lenslets is optically multiplied by a different random pattern. The masked images are then re-imaged by a set of secondary lenslets (LL2, f=10 cm, D= 1.27 cm) and overlaid via L2 (f=20 cm, D=7.5 cm) on the camera sensor to form the light field projection operator outlined by Equation 4.11. More compact setups could be realized using aberration-corrected optical elements. The monolithic entrance lens is not necessary and could be replaced by smaller elements resembling the curvature of corresponding pieces of a large lens (similar to fresnel lenses but with high imaging quality). This would remove the need for a large front lens and make the system more scalable for wider baselines. The baseline currently achieved is limited by the size of that entrance lens. However unlike the simple lenslet array our design does not suffer from dividing the sensor area and unlike previous single-sensor coded aperture it doesn't have major aperture overlap between adjacent views. We emphasize that the proposed mathematical technique is more resilient w.r.t larger baselines than previous methods, as evaluated in Figure 4.6.

Optics Implementation

The purpose of the prototype was to practically demonstrate the concept, and therefore, we did not use advanced optical elements as in professional photography camera. For alignment we used a single point source at the depth of field that was aligned with the center of the large entrance lens. This is also the symmetry axis of the system. Next the LL1 lenses are aligned along with LL2 lenses so that all the images of the point source

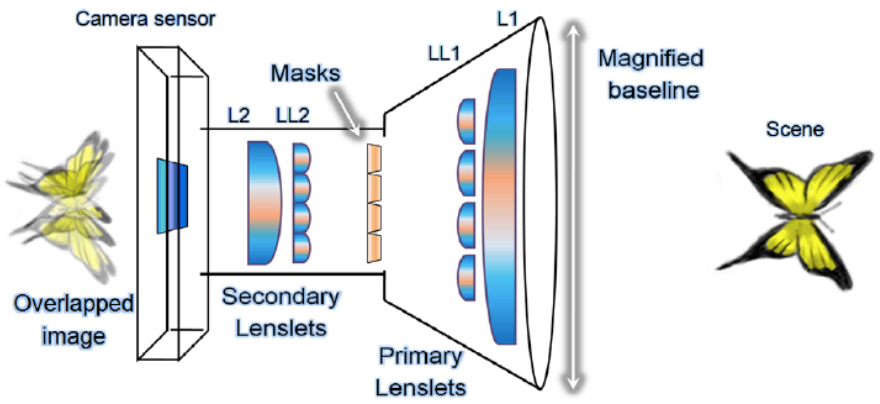


Figure 4.11: Different views of the same object are overlapped on the camera sensor.

created by lenslets are fully overlapped and are all in focus (4.11). We kept the L1 fixed and did the final fine alignments with slight adjustments of L2 and camera position.

During the acquisition we had to cover the setup so that the ambient light does not interfere with the measurements. Also to avoid cross talks due to adjacent lens glares, we used a plate with circular apertures to cover the rest of the large lens.

Mask Implementation and Calibration

We used a Gaussian mask as explained in the main text. The Gaussian pattern was quantized to 64 levels and dithered and then printed on a single transparent polymer substrate with 50800 DPI. It is important to note that this is not the resolution of the mask since an array of 10 by 10 pixels were used to make the dithered pattern of one pixel.

Since it was printed as a single piece when we moved the mask in the mask slot all the four figures were affected and therefore we did not have to change the pattern for individual views. To calibrate the mask accurately one should have independent x-y alignment control on each mask with accuracy level down to resolution of the mask. The masks should then be aligned in a way so that the grids of all masks from all the views are overlapping on the camera sensor. This can be a time consuming alignment in research level prototypes – specially when the mask should be moved or changed for every frame. However, for an industrial setup with prefabricated frames and mask holders or with spatial light modulators such repetitive alignment is not necessary. We believe that part of our artifact in recovery is because of such misalignments in the masks. On the camera side we disabled the white balance function and recorded the images in RAW format to avoid compression loss.

4.8.1 Alignment and Specifications

The optical system resembles a macro light field photography system or a single lens imaging system (with L1-L2 compound lens) combined with a 4f system between the (LL1 and LL2) lenslets (Figure 4.10). Depending on the required field of view, resolution of the mask, and the distance of the object to the lens (d_0) the rest of the geometry (d_1, d_2, d_3, d_4) can be aligned or predicted based on ray transfer matrix analysis. We experimentally aligned the system so that the plane of focus is also the parallax-free plane. Larger d_0 would form smaller images and force denser masks which would be ultimately limited by the diffraction limit and sensitivity of the camera. LL1 lenslets are flush to the L1 lens to allow the largest possible baseline. Stretching the LL1 array all the way to the peripheral of L1 can induce undesired relative aberration in the images on the side and thus some compensation elements would be required for aberration correction. The current prototype has f/7.5 and each view has approximately f/18. Unlike the case of coded aperture where adjacent views share the same aperture here the different lenslets have non-overlapping views but at the same time each view is imaged by the entire sensor. The system is extendable to larger baselines if the large entrance lens is replaced with prism pieces, however the main drawback is that a higher optical complexity is required compared to a conventional camera lens.

4.8.2 Software

For the physical experiments, light fields with 2×2 views are reconstructed from a single sensor image with a resolution of 400×400 pixels. The resolution of the prototype is currently limited by the printed mask resolution re-imaged on the sensor, and the accuracy of the calibration. Reconstruction is performed independently for each light field patch using a sliding window reconstruction. The patch size is chosen to contain the maximum amount of observed parallax in the scene. We used window sizes of 20×20 pixels with a varying number of time frames, as indicated for a specific dataset. A single 4D or 5D light field patch is recovered for each window location. Overlapping patch reconstructions are averaged for the final result. Processing time is about 8 hours for the sliding patch reconstructions on a computer with 4 nodes each with 4 cores at 2.2 GHz and 8 GB RAM. The rank threshold is set to $r = 6$ and the sparsity penalizing parameter is $\lambda = 0.05$. The solver usually converges in average within 300 iterations.

4.9 Results

Figure 4.1 shows a light field comprising 81 views recovered with the proposed algorithmic framework. We simulate a coded sensor image by multiplying each view by a Gaussian random code and summing over the modulated images. A reconstruction can faithfully estimate the target light field from two captured images.

As explained in Section 4.7.3, the minimum number of measurements to perfectly recover light fields depends on the degree of freedom of the light field tensor which is a function of number of views and parallax (rank). For light fields with high number of views such as Figure 4.1 (81 views), one cannot recover the static light fields with a single shot. However, when light fields video is captured the correlation along the motion can also be exploited. Therefore, light fields video can be recovered from single measurement from each frame. Figure 4.12 represents a static light field captured by camera array with 17×17 views. We recovered the light field with 5 measurements. Figure 4.13 shows a static 2×2 light field similar to the prototype recovered from a single measurement using the proposed tensor low-rank and sparse model.

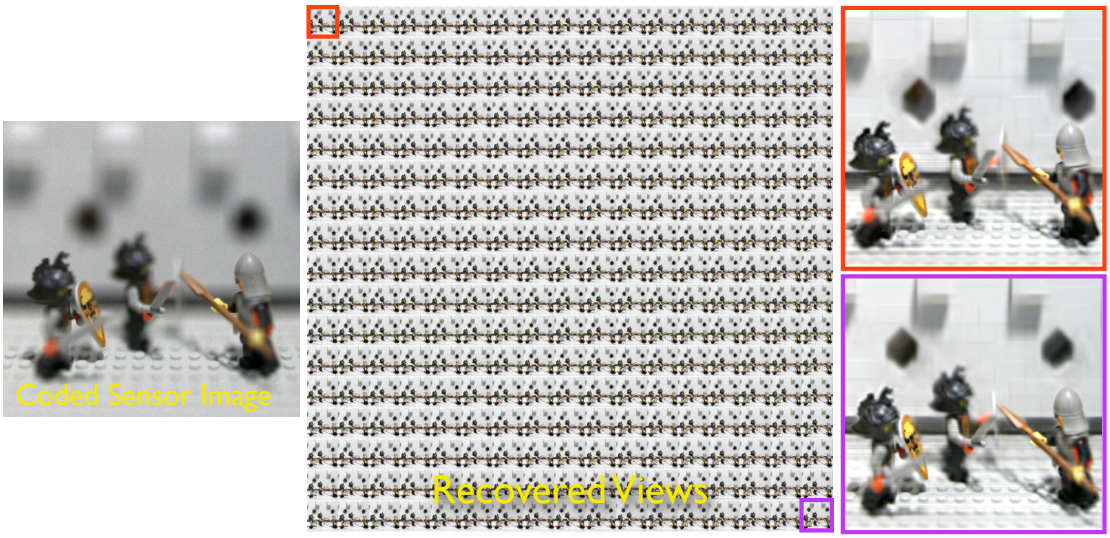


Figure 4.12: A 17×17-view light field is reconstructed by 5 coded measurements using the low-rank and sparse model. The number of shots required to recover light fields from coded measurements depends on the degree of freedom of the light field tensor which is a function of parallax and number of views.

High-quality light field recovery from a single measurement is possible with the prototype configuration when the modulation codes introduce a high degree of randomness in the measurements Figure 4.14 demonstrates a set of scenes captured with the prototype compressive light field camera and reconstructed with the proposed low-rank and sparse light field recovery. Finally, we also show reconstructions of a light field video in Figure 4.15. For this experiment, we manually move the object for 7 frames of a 9-frame sequence. A five-dimensional tensor low-rank and sparse reconstruction benefits from the coherence in time and angle of the light fields to recover the views from coded measurements in time. The former benefit is available when the optical codes change for each captured frame. Observed reconstruction quality is best when the motion between

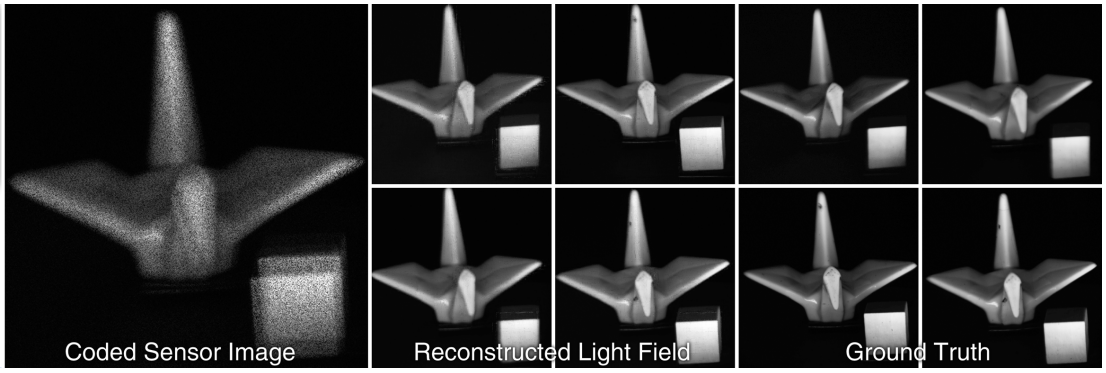


Figure 4.13: A single, coded sensor image containing all images of a light field at one frame is shown in the center. The proposed low-rank and sparsity-constrained optimization techniques allow for the full light field to be recovered from a single shot.

frames is small, though for scenes with a large amount of motion one could employ motion estimation and compensation techniques, such as [120].

4.9.1 Low-resolution Mask

The reconstruction resolution is bound by the resolution of the mask. In the proposed prototype, we benefit from high resolution printed mask to generate images with high resolution. Though we can improve the mask by replacing the printed mask by higher-resolution and dynamic spatial light modulators to improve image resolution, we use low-resolution printed mask to study the reconstruction quality of the prototype as a function of mask resolution.

We capture a scene with a low-resolution printed transparency that contains a random code. The printer is an Epson Stylus Photo 2200 and provides high-contrast and resolutions up to 1440 dpi. Compare to the coded image sensor captured with high resolution mask in the main paper, the transparency feature sizes of the low-resolution mask are relatively large in the sensor image (Figure 4.16, left). Unfortunately, large mask features partially destroy local randomness by blurring out high frequencies. To recover the light field, we need to downsample the captured image to the mask resolution, in which case the proposed reconstruction is successful (Figure 4.16, right).

Finally, we also show reconstructions of a light field video captured with low-resolution mask in Figure 4.17. For this experiment, we manually move the object for the last 5 frames of the sequence. A five-dimensional tensor low-rank and sparse reconstruction benefits from static parts of the scene (first few frames) by effectively having more measurements and also by the signal being low-rank in time.

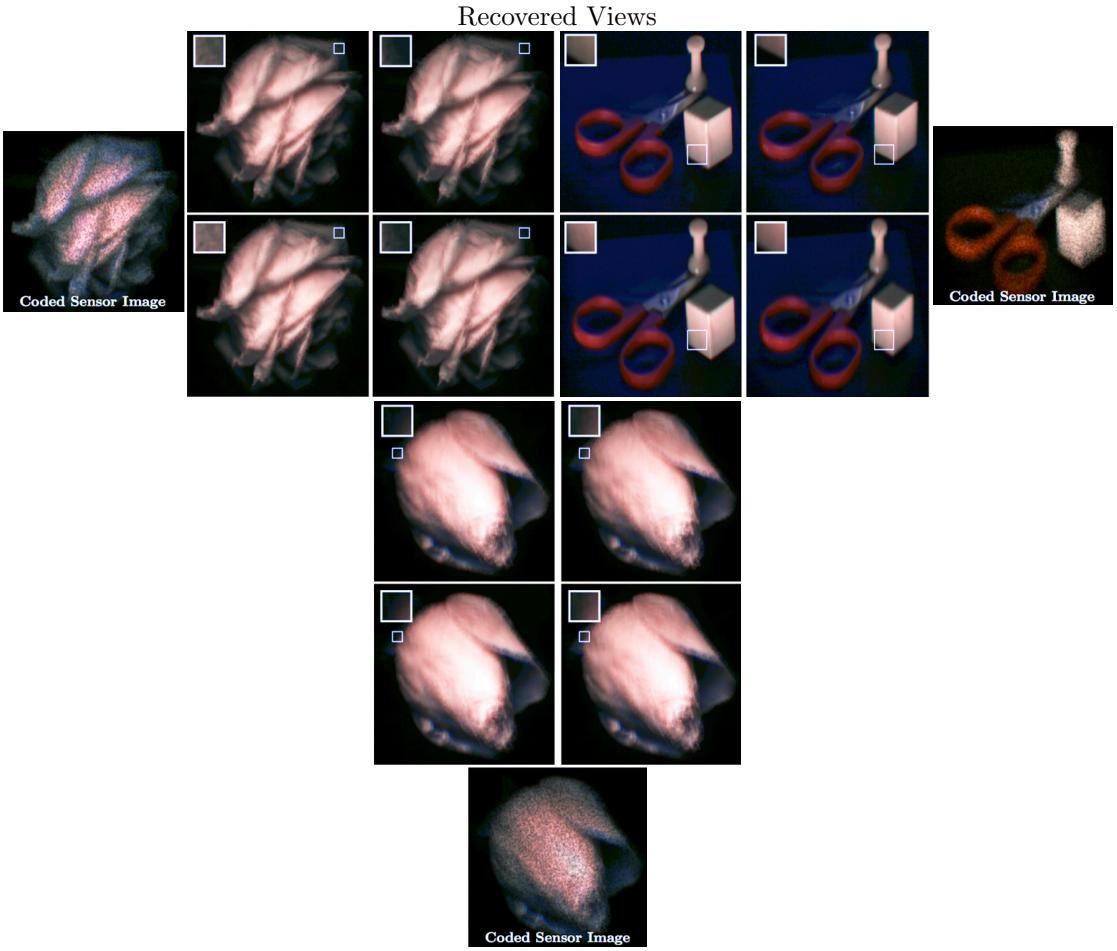


Figure 4.14: Light field reconstruction from prototype camera. These scenes are captured in a single, coded sensor image (bottom) each. The tensor low-rank and sparse recovery exploits light field correlation in space and angle to recover all 4 views from the coded image (top). Parallax is observed in the reconstructions.

4.10 Discussion

In this paper, we present a new approach to compressive light field photography. By combining tensor low-rank and sparse priors on the high-dimensional signal, we are able to efficiently model light field images and videos. We propose an efficient solver for the recovery problem and also a prototype device that allows for high-resolution light fields to be captured that have a wider baseline than previously-described single-device solutions.

Benefits and Limitations The proposed compressive camera offers a higher resolution than conventional, microlens-based light field cameras. Yet, these come at the

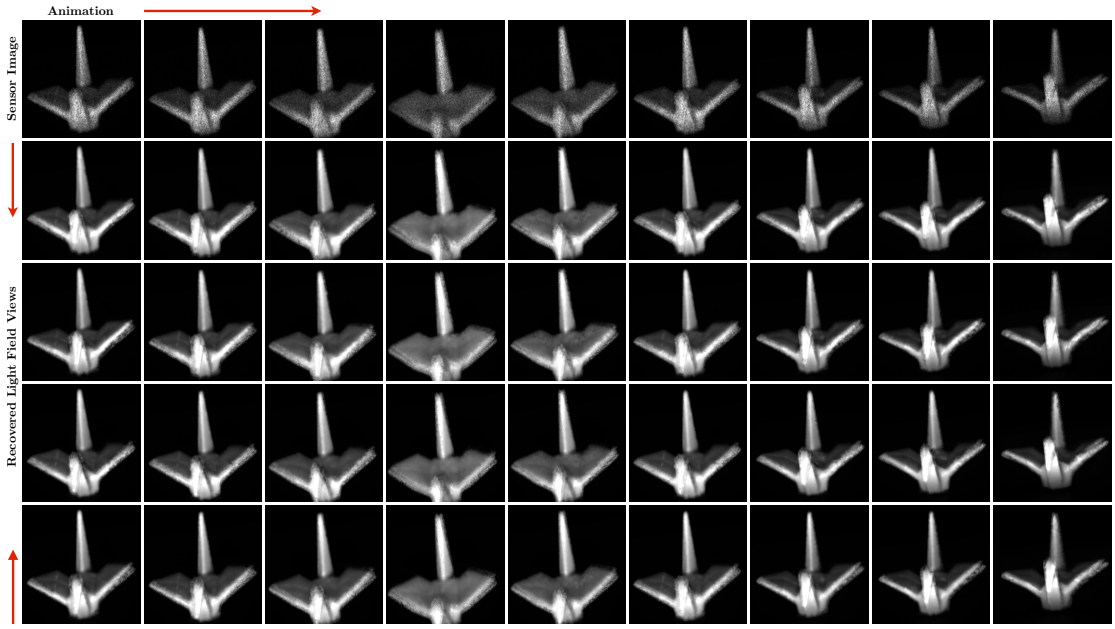


Figure 4.15: Reconstruction of light field video. We capture 9 frames of a dynamic target. For each frame, a coded sensor image is recorded—the codes vary in each frame. In this example, high-dimensional signal recovery exploits correlations in space, time, and angle to recover the light field motion and parallax. Effectively, capturing a video with dynamic optical codes increases the available measurements for static or moving objects.

cost of increased significantly reconstruction times. Compared with learned light field dictionaries, the proposed framework does not require a learning phase and is therefore also not bound to the capture parameters of the training scenes. This saves a significant amount of compute time and also makes the proposed framework more flexible and widely applicable. We believe that distributed compute infrastructures, such as cloud computing, have the potential to significantly reduce the reconstruction times.

Although the proposed framework is developed for light field videos with an arbitrary number of views and frames, the prototype we built is currently limited to four views and a few frames. The number of views is restricted by the size and cost of employed optical elements and the number of frames is limited by the fact that we record all animations in a stop-motion fashion by manually moving scene and mask while capturing coded light field projections. The prototype was designed only to experimentally verify the proposed framework. Optical aberrations and a low resolution of the printed masks place a limit on the resolution we currently achieve.

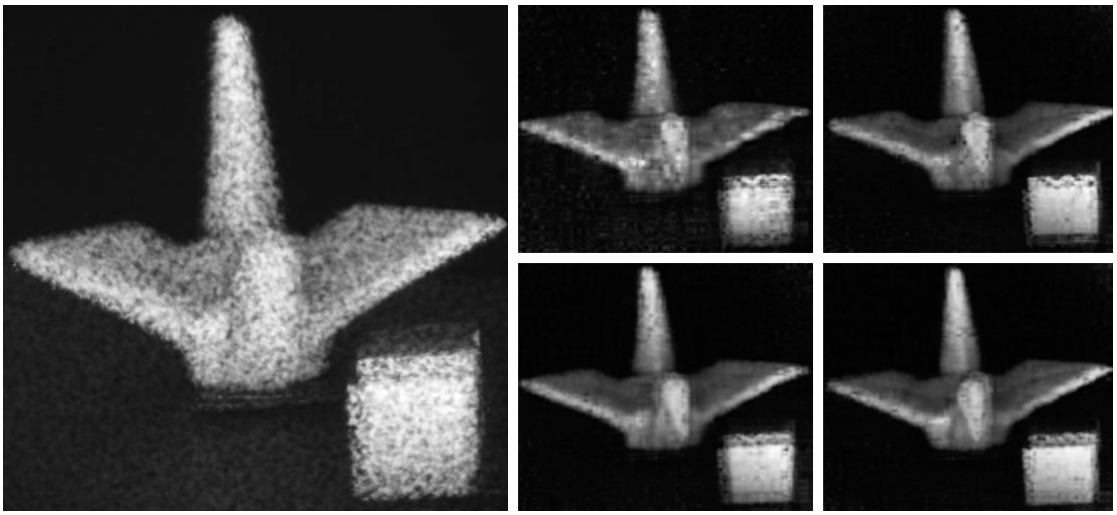


Figure 4.16: Light field reconstruction from prototype with low-resolution mask. The scene is captured in a single, coded sensor image (left) and subsequently recovered via tensor low-rank and sparse recovery. Parallax is observed in the reconstructions (right). Compare with high resolution masked employed in the main paper we observe that the mask feature sizes on the coded sensor image (left) are relatively large. Therefore, the local randomness of the mask is destroyed. We address this issue by downsampling the captured sensor image to the resolution of mask.

Future Work We would like to speed up processing times using cloud computing and improve the quality of the prototype design. Chromatic aberrations can be reduced using compound lens systems. Higher-resolution and dynamic spatial light modulators, instead of printed masks, will significantly improve image resolution and automate the capture process. Instead of using a single large imaging lens in the prototype, we would like to experiment with custom elements that do not require a monolithic lens but instead are a collection of independent compound lenses. Finally, we would like to further increase the baseline of the prototype and add more cameras to the array. The proposed mathematical framework also has applications in reducing the number of devices in camera arrays. We evaluate this application extensively in the supplement using simulations. Finally, we would like to incorporate the color spectrum, polarization, and other properties of light into our framework. Although a few approaches have been proposed to use related techniques for multi-spectral and lighting-dependent reflectance acquisition [3], a comprehensive and unified framework for compressive plenoptic or reflectance acquisition would be very interesting.

Conclusion Analyzing and exploiting redundancies of high-dimensional visual signals is the key to future camera designs. The proposed algorithmic framework is a step towards a new generation of computational imaging systems that follow this paradigm.

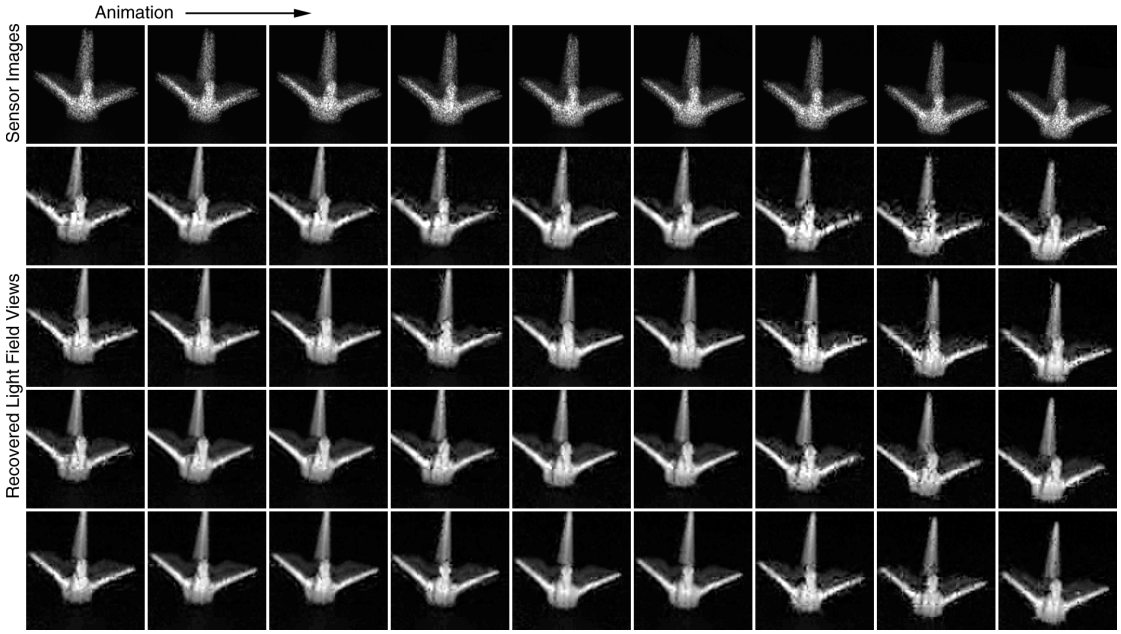


Figure 4.17: Reconstruction of light field video with low-resolution printed modulation mask. As in the static light field, the lower resolution mask decreases the resolution of recovered light field video. In this example, we capture nine frames of a dynamic target. For each frame a coded sensor image is recorded – the codes vary in each frame. In this example, high-dimensional signal recovery exploits correlations in space, time, and angle to recover the light field video. When the object is mostly static (first five frames), reconstruction quality is higher than for fast motion, because the signal is well-represented by the 5D low-rank prior. Effectively, capturing a video with dynamic optical codes increases the available measurements for static or slow-moving objects.

System designs at the intersection of optics and compressive computation, such as the proposed, facilitates higher resolutions, lower cost, new form factors, wider baselines, and many other benefits for computational photography.

Chapter 5

A Convex Solution to Disparity Estimation from Light Fields

The estimation of a disparity map from multiple images is one of the very well studied problems in computer vision. Some of the most dramatic improvements in this field occurred with the introduction of novel numerical frameworks and their corresponding theory. A non-exhaustive list of such breakthroughs are the early work on space carving [90], the level set formulation and the variational framework [46], the Markov random field framework with polynomial-complexity solvers [19], the L_1-Total Variation optimization framework [173] and, more recently, convex formulations that aim for global optimality [126]. In this work, we look at a novel approach based on recent primal-dual optimization techniques. Our approach is also convex as in the most recent developments, but we work with discrete labels (the possible disparity values).

Our formulation is based on a linear model of the data where a patch in an image is written as a linear combination of patches in other views. The key idea is that ideal Lambertian objects generate views that look alike (modulo foreshortening) and therefore corresponding patches live approximately on a 1D manifold. When objects are not Lambertian, they generate effects, such as specularities, that change with the pose of the camera. One can notice, however, that these effects are typically rare (i.e. , they happen only on some of the views) and spatially local. Hence, a natural way to model image patches of non Lambertian objects is by using an additive model where one of the two factors is sparse and the other is low-rank. If a finite set of possible depth candidates for a patch is available, one can then verify which hypothesis best fits the low-rank + sparse model. Our strategy is therefore a competition between the different disparity hypotheses. We essentially allow the data to be explained by a simultaneous linear combination of all low-rank + sparse models. However, we force coefficients to focus on only a few of the models (where each model corresponds to a single disparity

hypothesis) via group-sparsity penalty terms. We expect that coefficients be mostly non-zero at the true disparity as this is the case that gives the fit with the sparsest set of outliers. Notice that the individual coefficients of each linear combination are not important, and indeed, typically, infinite solutions might be possible especially at the correct disparity. However, as long as coefficients have most non-zero values at only one group, we can still correctly identify the disparity.

While this approach seems straightforward, in practice it faces considerable dimensionality challenges because data is replicated several times due to the patch-based model and the number of disparity hypotheses. This makes operations such as matrix inversion, often encountered in optimization schemes, impossible to carry out. To address these challenges we propose a primal-dual approach that results in simple element-wise thresholding operations and 2 (global) matrix multiplications at each step.

5.1 Related work

Light field disparity estimation: One of the first approaches to compute light field depth exploits linear structures in light fields through a line fitting algorithm [18]. Other methods use more traditional stereo reconstruction techniques to match the corresponding pixels in light field images, such as block-matching techniques [15] or clustering methods to identify similar pixel matches [12, 50]. Ziegler et al. [177] proposed a Fourier-based technique to compute depth values. To achieve higher global coherence, light field depth estimation methods employ a global cost function to impose smoothness on the estimated depth values [40, 86, 162]. A limitation common to all these methods is that they optimize a global cost function that is not convex. Therefore, the estimated depth map depends on the initial input. Moreover, fine details are lost because a coarse-to-fine multi-resolution technique is often used to avoid ending in weak local minima. Our approach overcomes these limitations by introducing a convex formulation.

Multiview stereo methods: Multiview techniques require detecting and handling outliers [5, 72]. The difficulty of outlier modeling is due to the unstructured nature of errors produced by outliers. However, these errors can only influence a small part of the image and are therefore sparse in a canonical basis [5, 169]. An alternative to explicit occlusion modeling is to match only reliable pixels and fill the unmatched correspondences via regularization [83, 147]. However, as explained in [148], these methods are prone to artifacts. Multiview stereo methods employ a large number of images [57, 82] to compute the full geometry of a scene and often yield a smooth geometry. Our light field disparity estimation yields a representation that falls in the middle: it is more complete than in stereo techniques, but less than in multiview stereo.

Sparse representation: The similarity of image structures in a dataset is used in data clustering [45, 101] to determine the low-dimensional subspace of high dimensional

data. Many schemes exploit data similarity to represent image correspondences in a dataset [100, 169]. In contrast to these clustering techniques, our proposed disparity estimation scheme looks for the best representation of each patch within a set of clusters. The clusters are generated from a number of disparity hypotheses, such that the members of a cluster are either chosen or discarded together. To achieve this we introduce a coupling term between the coefficients via group sparsity.

In this work, we estimate disparity from light fields by representing patches of a desired light field view with an overcomplete dictionary. The elements of the dictionary are patches of other views reprojected back onto a reference view for a given set of disparity candidates. If sufficiently many patch samples are available, patches of the reference view can be written as a linear combination of patches from the correct disparity hypothesis. This representation is naturally group sparse, since only a single disparity candidate of the dictionary can be assigned to a given patch. This representation can be recovered efficiently via group sparsity minimization [172].

5.2 Multiple views and light fields

We consider capturing several images of the same static scene by translating a camera on the $x-y$ plane, where z is aligned to the camera optical axis, or, equivalently, by employing a camera array, or a plenoptic camera, where all the camera sensors lie on the same plane. More in general, we can describe the captured data as a 4D light field $L: \Omega \times \Theta \mapsto [0, +\infty)$ where $\Omega \equiv \mathbb{R}^{N \times M}$ denotes the spatial domain (the pixel coordinates within each image) and Θ the angular domain (the camera center coordinates). We consider cameras arranged in a regular lattice and denote with $\Delta = [\Delta_x \ \Delta_y]^T \in \mathbb{R}^2$ the displacement between a camera and its north-west neighbor. Then, we define $\Theta = \{[\Delta_x i \ \ \Delta_y j]^T | i = 1 \dots n, j = 1 \dots m\}$ as the 3D camera center of the (i, j)-th camera is located at $[\Delta_x i \ \Delta_y j \ 0]^T$. For simplicity, we use the notation $L_{i,j}(x, y)$ to denote $L(x, y, i, j)$.

A visible plane in the scene, parallel to the images planes of the cameras, will generate images in the light field L that are related to each other by a shift or *disparity* $\rho : \Omega \mapsto [0, +\infty)$, for simplicity we denote $\rho(x, y)$ by ρ. In formulas, this can be written as

$$L_{i,j}(x, y) = L_{p,q}(x - \rho\Delta_x(p - i), y - \rho\Delta_y(q - j)) \tag{5.1}$$

for all (x, y) that fall within the spatial domain of both light field views and for all (i, j) and (p, q) camera pairs.

A common approach to estimating the disparity ρ is then to pose a variational problem

of the form

$$\min_{\rho} \sum_{\substack{i,j,p>i \\ q>j,x,y}} \Phi(L_{i,j}(x,y) - L_{p,q}(x-\rho(p-i)\Delta_x, y-\rho(q-j)\Delta_y)) + \Gamma(\rho), \tag{5.2}$$

where Φ is some robust penalty term for departures from zero and Γ is a regularization term for the unknown disparity ρ such as total variation. This problem is non-convex and therefore finding the global optimum is a very challenging task. While good solutions have been obtained for the above problem, recent efforts have produced convex variational formulations [56, 126] with high-quality disparity reconstructions. Both of these methods work with continuous representations. However, one of the key differences between these two methods is that, while [126] achieves convexity by increasing the problem dimensionality, [56] achieves convexity by fixing the structure tensor with some initial approximate disparity estimate. Our method follows the strategy of the first approach and also results in a high-dimensional representation. However, we do not rely on any initial estimate (although it might considerably speed up the convergence). Moreover, as we describe in the next sections, our convex formulation is entirely in the discrete domain and exploits the quantization of the disparity values.

5.3 A patch-based image formation model

Our first step is to rewrite the problem (5.2) as a patch matching problem. Let us define the *patch operator* $\mathcal{P}_{x,y}$ as the mapping that extracts the $W \times W$ patch whose top-left corner lies at (x, y) of an image I, i.e.,

$$\mathcal{P}_{x,y}(I) = \{I(x+x_0, y+y_0)\}_{x_0,y_0=0,\ldots,W-1}. \tag{5.3}$$

We define the output of the patch operator to be a patch rearranged as a column vector whose W^2 elements have been rearranged in lexicographical order. Consider extracting one patch from each view of a light field, except for the (i_0, j_0)-th one (for example, this could be the central view), given a disparity ρ and collecting all the patches in a matrix $\mathbf{Q}^{\rho}_{x,y} \in \mathbb{R}^{W^2 \times (nm-1)}$. This operation can be described via

$$\mathbf{Q}^{\rho}_{x,y} = \{\mathcal{P}_{x-\rho\Delta_x(p-i_0), y-\rho\Delta_y(q-j_0)}(L_{p,q}) : \forall (p,q) \neq (i_0, j_0)\}. \tag{5.4}$$

If ρ is the true disparity of a fronto-parallel object in space, then all the columns in $\mathbf{Q}^{\rho}_{x,y}$ will be identical to each other (in the ideal Lambertian case) and identical to the column vector $\mathcal{P}_{x,y}(L_{i_0,j_0})$. We also denote the latter vector with the symbol $\mathbf{y}_{x,y}$. More in general however, noise, non-Lambertianity, shadows, occlusions, inter reflections and so on need to be taken into account. Since we believe that most of the time the Lambertian approximation will hold, we consider all the other image distortions as infrequent and

use a sparse representation to model them, i.e.,

$$\mathbf{y}_{x,y} = \mathbf{Q}^{\rho}_{x,y}\mathbf{c}^{\rho}_{x,y} + \mathbf{e}_{x,y}, \tag{5.5}$$

where $\mathbf{c}^{\rho}_{x,y}$ is a $nm-1$ column vector and $\mathbf{e}_{x,y}$ is a W^2 column vector with few non-zero entries. The coefficients in $\mathbf{c}^{\rho}_{x,y}$ determine the linear combination of vectors in $\mathbf{Q}^{\rho}_{x,y}$ that generate $\mathbf{y}_{x,y}$. When the disparity ρ corresponds to the true solution, any $\mathbf{c}^{\rho}_{x,y}$ such that $\mathbf{1}^T\mathbf{c}^{\rho}_{x,y} = 1$ will satisfy the above equation. Vice versa, when the disparity is incorrect and the scene has sufficiently rich texture, there should not exist any vector $\mathbf{c}^{\rho}_{x,y}$ that satisfies (5.5). Thus, we propose to force the disparity ρ to take values only from the set $\{\rho_1, \rho_2, \ldots, \rho_D\}$ and extend (5.5) to

$$\mathbf{y}_{x,y} = [\mathbf{Q}^{\rho_1}_{x,y}\ \mathbf{Q}^{\rho_2}_{x,y}\ \ldots\ \mathbf{Q}^{\rho_D}_{x,y}][\mathbf{c}^{\rho_1}_{x,y}\mathbf{c}^{\rho_2}_{x,y}\cdots\mathbf{c}^{\rho_D}_{x,y}]^T + \mathbf{e}_{x,y} \doteq \mathbf{Q}_{x,y}\mathbf{c}_{x,y} + \mathbf{e}_{x,y}, \tag{5.6}$$

where the $W^2 \times (nm-1)D$ matrix $\mathbf{Q}_{x,y}$ and the $(nm-1)D$ vector $\mathbf{c}_{x,y}$ are implicitly defined by the equation to the right.

5.4 Depth estimation

Based on the model (5.5), a first formulation for estimating disparity through patch matching is

$$\min_{\mathbf{c},\mathbf{e}} \frac{1}{2}\sum_{x,y} \|\mathbf{y}_{x,y} - \mathbf{Q}_{x,y}\mathbf{c}_{x,y} - \mathbf{e}_{x,y}\|_2^2 + \mu\|\mathbf{e}_{x,y}\|_1, \tag{5.7}$$

where $\mu > 0$ is a constant determining the degree of sparsity of $\mathbf{e}_{x,y}$, $\|\mathbf{e}_{x,y}\|_1$ denotes the ℓ^1 norm of $\mathbf{e}_{x,y}$, and $\mathbf{c}$ and $\mathbf{e}$ are the column vectors obtained by stacking vertically all the vectors $\mathbf{c}_{x,y}$ and $\mathbf{e}_{x,y}$ respectively. Since the total number of patches within the image domain is $\tilde{M}\tilde{N}$, where $\tilde{M} = M - W + 1$ and $\tilde{N} = N - W + 1$, the $\mathbf{e}$ vector has $\tilde{M}\tilde{N}W^2$ elements and the $\mathbf{c}$ vector has $\tilde{M}\tilde{N}(nm-1)D$ elements.

As explained in the previous section, we aim at concentrating the coefficients of $\mathbf{c}_{x,y}$ on the patches belonging to just one disparity hypothesis. If this is the case, then, given $\mathbf{c}_{x,y}$, one can estimate the disparity at a pixel (x, y) by using

$$\hat{\rho} = \underset{\rho\in\{\rho_1,\ldots,\rho_D\}}{\operatorname{argmax}} \|\mathbf{c}^{\rho}_{x,y}\|_{2,1}. \tag{5.8}$$

The same problem can be written in the following compact form

$$\min_{\mathbf{c},\mathbf{e}} \frac{1}{2}\|\mathbf{y} - \mathbf{Q}\mathbf{c} - \mathbf{e}\|_2^2 + \mu\|\mathbf{e}\|_1, \tag{5.9}$$

where the column vector $\mathbf{y}$ has been obtained by stacking all the $\mathbf{y}_{x,y}$, and $\mathbf{Q}$ is a block diagonal matrix whose blocks are the matrices $\mathbf{Q}_{x,y}$. To encourage the concentration

of nonzero entries in a single disparity block of $\mathbf{c}_{x,y}$ we propose to minimize the mixed $\ell_{2,1}$ norm of $\mathbf{c}_{x,y}$. Finally, since the disparity is a smooth map, we add a vector-valued isotropic total variation (TV) regularization term

$$\|\nabla \mathbf{c}\|_{2,1} \doteq \sum_{x,y} \sqrt{\|\mathbf{c}_{x,y} - \mathbf{c}_{x+1,y}\|_2^2 + \|\mathbf{c}_{x,y} - \mathbf{c}_{x,y+1}\|_2^2}, \tag{5.10}$$

where ∇ denotes the finite gradient in the spatial domain (and can be written in matrix form). By minimizing this term we encourage $\mathbf{c}$ coefficients to be similar across the spatial domain. The complete minimization problem can be written as follows

$$\min_{\mathbf{c},\mathbf{e}} \frac{1}{2}\|\mathbf{y} - \mathbf{Q}\mathbf{c} - \mathbf{e}\|_2^2 + \mu\|\mathbf{e}\|_1 + \lambda\|\nabla \mathbf{c}\|_{2,1} + \gamma\|\mathbf{c}\|_{2,1}, \tag{5.11}$$

where $\lambda, \gamma > 0$ are two constants. This is a convex problem and therefore it has the desirable property of converging to the same global optimum given any initialization. The minimization of problem (5.11) presents several challenges due to its high dimensionality, which we address in the next section.

5.5 Primal-dual formulation

One immediate issue of a primal solver for problem (5.11) is that it requires inverting very large matrices that are not easily diagonalized. To avoid such computational difficulties, we consider the primal-dual method, which is a first order algorithm, it does not require matrix inversions and enjoys fast convergence rates [126].

Firstly, we rewrite problem (5.11) in a more compact way by combining all the unknowns $\mathbf{c}$ and $\mathbf{e}$ into a single variable $\mathbf{x}$, and by defining 3 new functions F_1, F_2, and F_3 as follows

$$F_1(\mathbf{A}\mathbf{x} - \mathbf{y}) \doteq \frac{1}{2}\|\mathbf{y} - \mathbf{Q}\mathbf{c} - \mathbf{e}\|_2^2 \tag{5.12}$$

$$F_2(\Pi_{\mathbf{e}}\mathbf{x}) \doteq \|\mathbf{e}\|_1 \tag{5.13}$$

$$F_3(\mathbf{B}\mathbf{x}) \doteq \|\nabla \mathbf{c}\|_{2,1} + \frac{\gamma}{\lambda}\|\mathbf{c}\|_{2,1}, \tag{5.14}$$

where $\mathbf{A} \doteq [\mathbf{Q}\ \mathbf{I}_d]$, with $\mathbf{I}_d$ the identity matrix, $\Pi_{\mathbf{e}}\mathbf{x} \doteq \mathbf{e}$ and $\mathbf{B} \doteq [\nabla^\top\ \frac{\gamma}{\lambda}\mathbf{I}_d]^\top \Pi_{\mathbf{c}}$, with $\Pi_{\mathbf{c}}\mathbf{x} \doteq \mathbf{c}$. Notice that all the above functions are convex in the variable $\mathbf{x}$. Then, our primal formulation becomes

$$\min_{\mathbf{x}} F_1(\mathbf{A}\mathbf{x} - \mathbf{y}) + \mu F_2(\Pi_{\mathbf{e}}\mathbf{x}) + \lambda F_3(\mathbf{B}\mathbf{x}). \tag{5.15}$$

To solve the primal problem we can compute the gradients of the cost function and set it to zero. An immediate observation is that the gradient will yield in the best case linear

systems with non-diagonal matrices. For example, the first term $F_1(\mathbf{Ax}-\mathbf{y})$ yields

$$\frac{\partial}{\partial \mathbf{x}} F_1(\mathbf{Ax}-\mathbf{y}) = \mathbf{A}^\top \mathbf{Ax} - \mathbf{A}^\top \mathbf{y}, \tag{5.16}$$

which requires dealing with the matrix $\mathbf{A}^\top \mathbf{A}$. To avoid that, we use the primal-dual method. This method is based on the Legendre-Fenchel (LF) transform. Given a function F, the LF transform yields a conjugate function F^* such that

$$F^*(\mathbf{z}) \doteq \sup_{\mathbf{x}} \langle \mathbf{x}, \mathbf{z} \rangle - F(\mathbf{x}). \tag{5.17}$$

The conjugate function F^* is by construction convex and when F is also convex, then the LF transform F^{**} of the conjugate F^* is again F. When the conjugate functions F_1^*, F_2^*, and F_3^* can be computed easily and possibly in closed-form, then it is convenient to consider the primal-dual problem

$$\begin{aligned} \min_{\mathbf{x}} \max_{\mathbf{z}_1,\mathbf{z}_2,\mathbf{z}_3} & < \mathbf{Ax}-\mathbf{y}, \mathbf{z}_1 > - F_1^*(\mathbf{z}_1) + \mu < \Pi_{\mathbf{e}}\mathbf{x}, \mathbf{z}_2 > - \mu F_2^*(\mathbf{z}_2) \\ & + \lambda < \mathbf{Bx}, \mathbf{z}_3 > - \lambda F_3^*(\mathbf{z}_3), \end{aligned} \tag{5.18}$$

which we write in more compact form as

$$\min_{\mathbf{x}} \max_{\mathbf{z}} < K\mathbf{x}, \mathbf{z} > - \hat{F}(\mathbf{z}) \tag{5.19}$$

where $\mathbf{K} \doteq [\mathbf{A}^\top \ \mu\Pi_{\mathbf{e}}^\top \ \lambda \mathbf{B}^\top]^\top$, $\mathbf{z} \doteq [\mathbf{z}_1^\top \ \mathbf{z}_2^\top \ \mathbf{z}_3^\top]^\top$, and $\hat{F}(\mathbf{z}) \doteq F_1^*(\mathbf{z}_1) + \mu F_2^*(\mathbf{z}_2) + \lambda F_3^*(\mathbf{z}_3)$. To solve the above saddle point problem, we need to define the proximity operator, which is our fundamental computational tool to deal with the conjugate functions.

5.5.1 Proximity operator

The main result that we will exploit here is Moreau's identity. Given the conjugate F^* of F we have that

$$\text{prox}_{\sigma F^*}(\mathbf{z}) = \mathbf{z} - \sigma \text{prox}_{F/\sigma}(\mathbf{z}/\sigma) \tag{5.20}$$

and hence we can compute the proximity operator of the conjugate function F^* directly by using the proximity operator of the function F.

5.5.2 Primal-dual algorithm

The primal-dual algorithm to solve problem (5.19) is defined in Algorithm 4.

In algorithm 4, while the bottom two iterations are straightforward, the first one on the dual variable $\mathbf{z}$ requires computing the proximity operator of the conjugate functions

Algorithm 4: The primal-dual algorithm to solve (5.19)

initialize $\theta \in (0,1]$, $\tau\sigma\|\mathbf{K}\|^2 < 1$
while *not converged* **do**
  $\mathbf{z}_1^{n+1} = \text{prox}_{\sigma F_1^*}\left(\mathbf{z}_1^n + \sigma(\mathbf{A}\bar{\mathbf{x}}^n - \mathbf{y})\right)$
  $\mathbf{z}_2^{n+1} = \text{prox}_{\sigma\mu F_2^*}\left(\mathbf{z}_2^n + \sigma\mu\Pi_{\mathbf{e}}\bar{\mathbf{x}}^n\right)$
  $\mathbf{z}_3^{n+1} = \text{prox}_{\sigma\lambda F_3^*}\left(\mathbf{z}^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n\right)$
  $\mathbf{x}^{n+1} = \mathbf{x}^n - \tau\mathbf{K}^\top\mathbf{z}^{n+1}$
  $\bar{\mathbf{x}}^{n+1} = \mathbf{x}^{n+1} + \theta(\mathbf{x}^{n+1} - \mathbf{x}^n)$

F_1^*, F_2^*, and F_3^*. The first two functions are relatively easy to obtain as the conjugate functions can be computed in closed-form

$$F_1^*(\mathbf{z}_1) = \frac{1}{2}\|\mathbf{z}_1\|_2^2, \qquad \{F_2^*(\mathbf{z}_2)\}_s = \begin{cases} 0 & \text{if } |\{\mathbf{z}_2\}_s)| \leq \mu \\ +\infty & \text{otherwise} \end{cases} \tag{5.21}$$

where $s = 1, \ldots, \tilde{M}\tilde{N}W^2$. Hence, we can readily obtain the first two steps of the primal-dual algorithm

$$\mathbf{z}_1^{n+1} = \frac{1}{\sigma+1}\left(\mathbf{z}_1^n + \sigma(\mathbf{A}\bar{\mathbf{x}}^n - \mathbf{y})\right), \qquad \{\mathbf{z}_2^{n+1}\}_s = \mathcal{H}_{\sigma\mu}\left(\left\{\frac{\mathbf{z}_2^n}{\sigma\mu} + \Pi_{\mathbf{e}}\bar{\mathbf{x}}^n\right\}_s\right), \tag{5.22}$$

where $s = 1, \ldots, \tilde{M}\tilde{N}W^2$ and $\mathcal{H}_{\sigma\mu}$ denotes the element-wise thresholding operator

$$\mathcal{H}_{\sigma\mu}(\mathbf{z}) \doteq \min\left\{\sigma\mu, |\mathbf{z}|\right\}\text{sign}(\mathbf{z}). \tag{5.23}$$

The last term F_3^* is more involved. We compute the update equation by exploiting Moreau's identity

$$\text{prox}_{\sigma\lambda F_3^*}(\mathbf{z}_3^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n) = \mathbf{z}_3^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n - \sigma\lambda\text{prox}_{F_3/(\sigma\lambda)}(\mathbf{z}_3^n/(\sigma\lambda) + \mathbf{B}\bar{\mathbf{x}}^n), \tag{5.24}$$

so that we only need to compute $\text{prox}_{F_3/(\sigma\lambda)}$. Notice that $F_3(\mathbf{z}_3)$ is the ℓ_1/ℓ_2 norm $\|\mathbf{z}_3\|_{2,1}$. Thus, we need to evaluate

$$\text{prox}_{F_3/(\sigma\lambda)}(\mathbf{z}_3^n/(\sigma\lambda) + \mathbf{B}\bar{\mathbf{x}}^n) = \underset{\mathbf{z}}{\text{argmin}} \;\; \frac{1}{2}\|\frac{1}{\sigma\lambda}\mathbf{z}_3^n + \mathbf{B}\bar{\mathbf{x}}^n - \mathbf{z}\|_2^2 + \frac{1}{\sigma\lambda}\|\mathbf{z}\|_{2,1}. \tag{5.25}$$

The solution is computed in closed-form and results in a block soft-thresholding

$$\text{prox}_{F_3/(\sigma\lambda)}(\mathbf{z}_3^n/(\sigma\lambda) + \mathbf{B}\bar{\mathbf{x}}^n) = \mathcal{S}_{1/(\sigma\lambda)}\left(\frac{1}{\sigma\lambda}\mathbf{z}_3^n + \mathbf{B}\bar{\mathbf{x}}^n\right) \tag{5.26}$$

with

$$\left\{\mathcal{S}_{1/(\sigma\lambda)}(\mathbf{z}_3)\right\}_b = \{\mathbf{z}_3\}_b \max\left\{0, 1 - \frac{1}{\sigma\lambda\|\{\mathbf{z}_3\}_b\|_2}\right\} \tag{5.27}$$

and where blocks are indexed by $b = 1, \ldots, (3\tilde{M}\tilde{N} - \tilde{M} - \tilde{N})D$, since $\mathbf{z}_3$ is a $(3\tilde{M}\tilde{N} - \tilde{M} - \tilde{N})D(nm-1)$ dimensional vector.[1] Finally, by plugging the last expression in the proximity operator of F_3^*, the last update equation becomes

$$\begin{aligned}\{\text{prox}_{\sigma\lambda F_3^*}(\mathbf{z}_3^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n)\}_b = &\{\mathbf{z}_3^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n\}_b \\ &\cdot\left(1 - \max\left\{0, 1 - \frac{1}{\|\{\mathbf{z}_3^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n\}_b\|_2}\right\}\right),\end{aligned} \tag{5.28}$$

where $b = 1, \ldots, (3\tilde{M}\tilde{N} - \tilde{M} - \tilde{N})D$.

In all update equations there are no matrix inversions and calculations are therefore highly parallelizable. The final algorithm is summarized in Table 5.

Algorithm 5: Primal-dual algorithm for disparity estimation from light field data.

1: **initialize** $\mathbf{z}_1 \in \mathbb{R}^{\tilde{M}\tilde{N}W^2 \times 1}$, $\mathbf{z}_2 \in \mathbb{R}^{\tilde{M}\tilde{N}W^2 \times 1}$, $\mathbf{z}_3 \in \mathbb{R}^{(3\tilde{M}\tilde{N} - \tilde{M} - \tilde{N})D(nm-1)\times 1}$
2: **while** *not converged* **do**
3: $\quad \mathbf{z}_1^{n+1} = (\mathbf{z}_1^n + \sigma(\mathbf{A}\bar{\mathbf{x}}^n - \mathbf{y}))/(\sigma + 1)$
4: $\quad \{\mathbf{z}_2^{n+1}\}_s = \mathcal{H}_{\sigma\mu}(\{\mathbf{z}_2^n/(\mu\sigma) + \Pi_{\mathbf{e}}\bar{\mathbf{x}}^n\}_s)$
5: $\quad \mathbf{z}_3^{n+1} = \text{prox}_{\sigma\lambda F_3^*}(\mathbf{z}^n + \sigma\lambda\mathbf{B}\bar{\mathbf{x}}^n)$
6: $\quad \mathbf{x}^{n+1} = \mathbf{x}^n - \tau\mathbf{K}^\top\mathbf{z}^{n+1}$
7: $\quad \bar{\mathbf{x}}^{n+1} = \mathbf{x}^{n+1} + \theta(\mathbf{x}^{n+1} - \mathbf{x}^n)$

5.5.3 Implementation details

Because of the discretization, the dimensionality of the problem is quite high. One approach to managing such dimensionality is to use block coordinate descent [154], where one works iteratively on different subsets of the variables. In this work, we consider a simple and efficient approximation: we consider restricting the possible disparities $\rho_1, \ldots, \rho_D$ to a small but carefully selected subset and always work with that subset. To gain additional freedom, at each pixel (x, y) we make a different choice of such subset. Our strategy is to evaluate the function

$$g_{x,y}(\rho) = \sum_{i,j} \sum_{p>i, q>j} \Phi(L_{i,j}(x,y) - L_{p,q}(x - \rho\Delta_x(p-i), y - \rho\Delta_y(q-j))) \tag{5.29}$$

for as many ρ values as possible. Then, we sort $g_{x,y}$ in ascending order and take the disparities corresponding to the first 5 values of $g_{x,y}$. We then also add 5 more disparity candidates by selecting the disparities of neighboring pixels (in a 4-neighborhood structure) corresponding to the smallest cost. The purpose of this second group of disparity

[1] The total variation term introduces 2 blocks for any pixel in Ω except for the left hand side column and the bottom row of pixels (total blocks is $2(\tilde{M}-1)(\tilde{N}-1)$). These two rows of pixels, except for the bottom right corner, introduce only one block (total blocks $(\tilde{M}-1)+(\tilde{N}-1)$). Finally, the block sparsity term introduces $\tilde{M}\tilde{N}D$ blocks.

candidates is to allow (spatially) smooth disparity estimates.

5.6 Experimental results

We study the performance and robustness of our light field disparity estimation framework on different datasets, Buddha [164], Watch [1], Amethyst and Truck from the Stanford light field archive.[2] We compare our results with two light field depth estimation schemes [86, 162], and convex formulations [125]. Our parameters are: $\mu = 0.6$ and $\gamma = 1$ for all datasets, and $\lambda = 0.1$ for Amethyst and Truck. We work with 5×5 pixels patches ($W = 5$). Our algorithm is also demonstrated in the limit case where there are only two views (stereo). The group sparsity constraint can still work quite successfully. Another important factor is the input image size. We find that the method works better with high resolution images. However, it can also perform reasonably well on low-resolution data. In contrast, [86, 162] are challenged with few views and/or low-resolution images. The runtime of our algorithm is higher than [86]. If parallelism is fully exploited the ideal running time is about 1-3 minutes depending on the resolution and number of views. In our experiments, we search through 200 disparity candidates to determine the 10 candidates.

Figure 5.1 compares our scheme with simple plane sweep disparity search (independently at each pixel). We observe that our scheme imposes the global smoothness on the estimated disparity while the plane sweep fails to provide a smooth disparity map. As expected, the number of views used in the disparity estimation problem improves the depth estimate considerably. In our approach, an increase in the number of views results in more samples per disparity candidate in the $\mathbf{Q}$ matrix, therefore a better chance of fitting data more reliably. This is clearly noticeable in Figure 5.1 and Figure 5.2. We compare qualitatively our disparity estimation algorithm with the techniques introduced in [66, 162] in Table 5.1. It is clear that our scheme provides a better reconstruction quality. In Figure 5.4, we illustrate how the patch size W has an immediate effect on the recovered depth map. As is well known, the larger the patch, the less noisy the depth estimate is. However, increases in patch size also affect the performance of the algorithm in the recovery of small details.

Table 5.1: Qualitative results for Buddha shown in Figure 5.1. The table shows the percentage of pixels with relative depth error of more than 0.2%, 0.5% and 1%.

4 views			3 views			[66]			[162]	
1%	0.5%	0.2%	1%	0.5%	0.2%	1%	0.5%	0.2%	1%	0.2%
0.13	0.33	1.9	0.139	0.33	1.99	1.15	2.44	15.05	2.9	60.4

[2] See http://lightfield.stanford.edu.

5.6.1 Convex labeling

Given $\mathbf{c}$, the disparity $\hat{\rho}$ is determined at each pixel (x, y) independently by solving

$$\hat{\rho}_{x,y} = \underset{\rho \in \{\rho_1, \ldots, \rho_D\}}{\operatorname{argmax}} \|\mathbf{c}^{\rho}_{x,y}\|_{2,1}. \tag{5.30}$$

However, one can introduce a convex labeling by fitting a convex function per patch to the estimated $\mathbf{c}_{x,y}$ and impose a smooth prior such as TV constraint on the recovered disparity to solve an inpainting problem.

$$\tilde{\rho} = \underset{\rho}{\operatorname{argmin}} \ \lambda \|\nabla \rho\|_2 - f_{\mathbf{c}}(\rho), \tag{5.31}$$

where $\lambda > 0$ is a constant and $f_{\mathbf{c}}$ is the convex function fitted to $\mathbf{c}_{x,y}$. This formulation can help to remove the disparity noise however it can smooth some details.

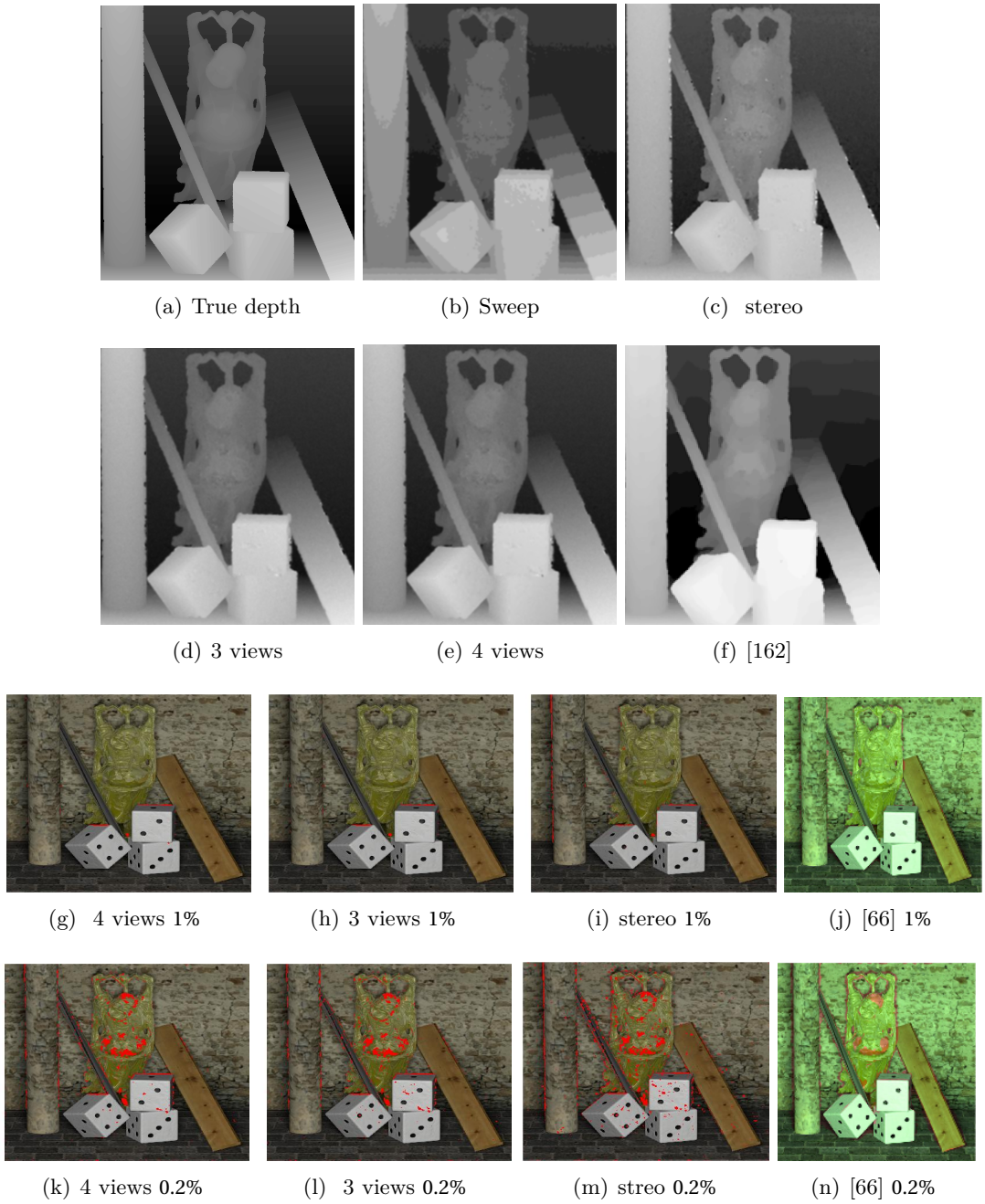
5.6.2 Multiview Depth Estimation

We test our light field depth estimation technique on the Middlebury dataset[3] with stereo methods [178], two light field depth estimation schemes [86, 162], and convex formulations [125, 126]. Our parameters are: $\mu = 0.6$ and $\gamma = 1$ for all datasets, then $\lambda = 0.5$ for Venus, $\lambda = 0.15$ for Cone.
One aspect we would like to point out is that the number of views used in the depth estimation problem improves the depth estimate considerably, this is clearly noticeable in Figure 5.6. Our algorithm is also demonstrated in the limit case where there are only two views (Figure 5.7). The group sparsity constraint can still work quite successfully.

5.7 Conclusions

We have presented a novel convex formulation to estimate depth from light field data. The method is based on a careful discretization of disparity values and exploits a linear patch-based formulation to represent patches in one view with patches in other views. The proposed model can easily be extended to handle simple departures from the ideal Lambertian model. For example, the current model can already handle contrast changes due to illumination (these changes would be reflected in the magnitude of the coefficients of $\mathbf{c}$). The problem of depth estimation is cast as a minimization problem subject to group sparsity constraints and spatial smoothing. To gain computational efficiency we use the primal-dual method. This results in an algorithm where each dual variable update can be computed easily, independently and efficiently. Our experiments show that this method competes well with the state of the art.

[3] http://vision.middlebury.edu/stereo/data/



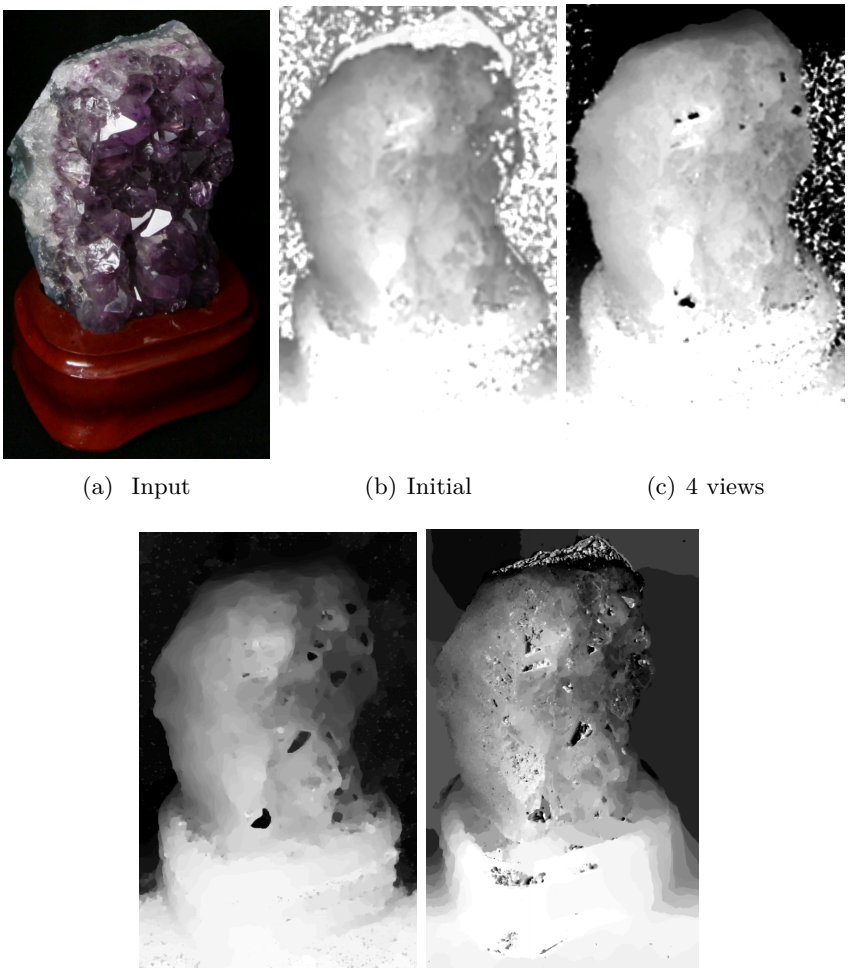
(a) True depth (b) Sweep (c) stereo

(d) 3 views (e) 4 views (f) [162]

(g) 4 views 1% (h) 3 views 1% (i) stereo 1% (j) [66] 1%

(k) 4 views 0.2% (l) 3 views 0.2% (m) streo 0.2% (n) [66] 0.2%

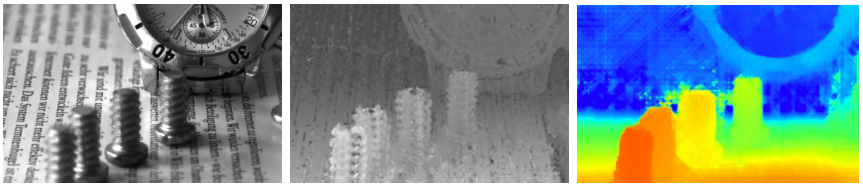
Figure 5.1: Disparity estimation of Buddha dataset. From left to right, top row shows: the center view, the ground truth, the depth map obtained by plane sweep depth search (independently at each pixel). Middle row: the estimated depth map using different number of views, and the depth map obtained from [162]. Bottom: the estimated disparity in areas with error more than 1% are highlighted in red. We observe that an increase in the number of views improves the reconstruction quality and our scheme provides sharpe edges while the depth map estimated using [162] blurs the edges and has staircasing artifacts.



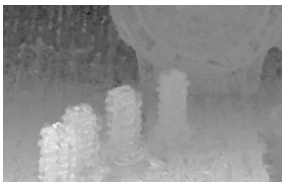
(a) Input (b) Initial (c) 4 views

(d) [162] (e) [126]

Figure 5.2: Disparity estimation of Amethyst dataset from a camera array. (a) One of the input images. (b) Initial depth estimate (plane sweep depth search) (c) Estimated disparitity using our scheme. (d-e) Estimated depth map using [162] and [126]. Notice how we obtain a reasonable estimate of the top part of the stone, while competing methods either fail or obtain a noisier estimate.

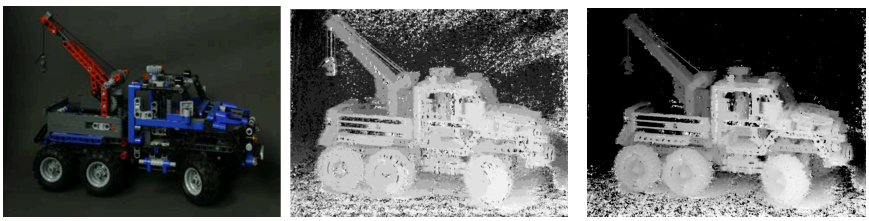


(a) Center view (b) 4 views (c) Manufacturer



(d) [162]

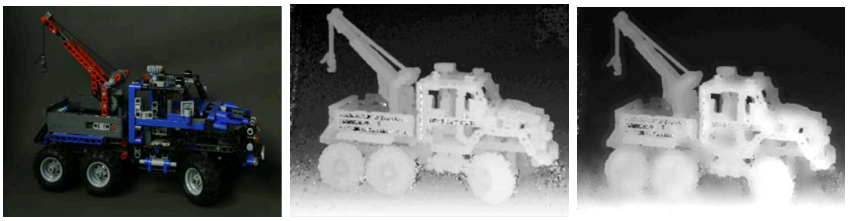
Figure 5.3: Depth estimation with the Raytrix plenoptic camera (handheld light field camera). We compare our algorithm with the reference depth provided by the manufacturer and [162]. Our scheme on a handheld light field camera yields a more detailed depth map.



(a) Input (b) Initial (c) $W = 5$

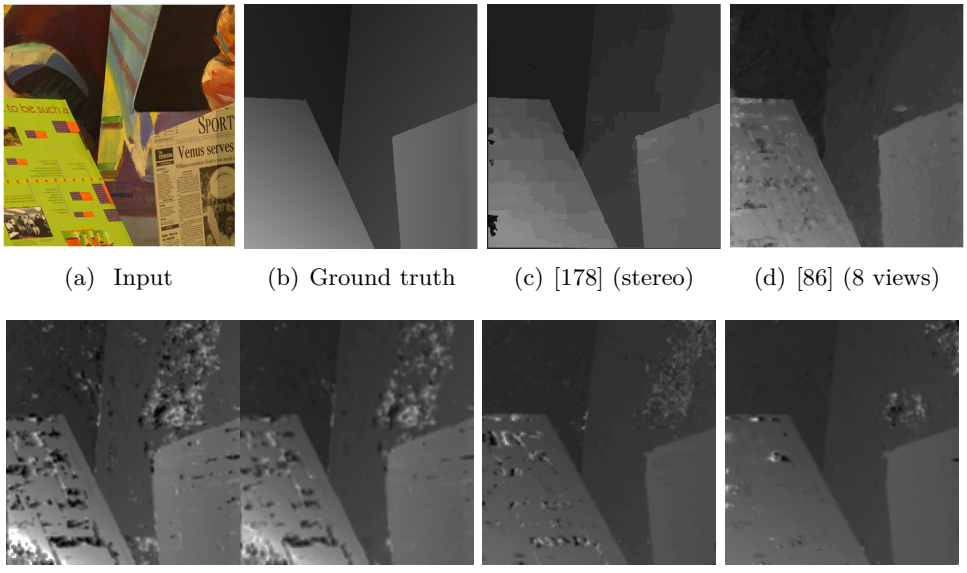
(d) $W = 8$ (e) [162] (f) [86]

Figure 5.4: Disparity estimation of Truck dataset acquired by a camera array. We assess the influence of patch size in our scheme. Increasing the patch size results in a less noisy, but also smoother, depth map. In comparison to [86, 162], our algorithm provides sharper edges with a noisier background. This is due to two main reasons: 1) The initial 10 disparity candidates selected among 200 candidates do not contain the true disparity value, which can be improved by working on 200 candidates using block coordinate descent [154]. 2) The selection of the highest coefficients in $\mathbf{c}$ may lead to noisy disparity which can be addressed by imposing smoothness in the final estimation of the disparity from the coefficients of $\mathbf{c}$.



(a) Input (b) Independent Labeling (c) Smooth Labeling

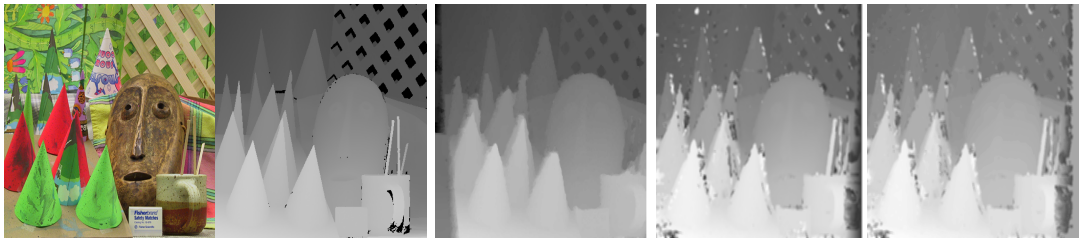
Figure 5.5: Comparison of smooth disparity estimation with independent disparity estimation. We observe that smooth labeling can remove some noise as the result of improper initial candidate selection (specially in smooth area), however it smooths out some details with respect to independent disparity labeling.



(a) Input (b) Ground truth (c) [178] (stereo) (d) [86] (8 views)

(e) Initial stereo (f) stereo (g) 4 views (h) 8 views

Figure 5.6: Disparity estimation from the multiview: Venus dataset. On the top row we show (left to right): one input image, the ground truth depth map, the estimate of [178] for the stereo case and that of [86] for 8 views. On the bottom row we show (left to right): our initial depth estimate (plane sweep depth search), our final result with 2, 4 and 8 views.



(a) Input (b) Ground truth (c) [125] (Stereo) (d) Sweep (e) stereo

Figure 5.7: Disparity estimation from stereo images: Cone dataset. Top image is one of the input images. Bottom row (left to right): the ground truth depth map, the estimate of [125] for the stereo case, our initial depth estimate (plane sweep depth search), and our final result with 2 views.

Chapter 6

Low-Power Compressive Multi-channel Cortical Recording

6.1 Introduction

Wireless monitoring of brain activity through implantable devices is a promising technology enabling advanced and cost-effective diagnosis and treatment of brain disorders such as stroke, Parkinson's disease, depression and epilepsy [110, 137]. Recording from multiple sites, however, introduces a major technological bottleneck as the large bandwidth requirement for data telemetry which is not easily achievable by state-of-the-art wireless technology. The increased power consumption of transmission for large recording arrays can cause major safety and biocompatibility concerns regarding the applicability of such devices. Thus, some type of data reduction prior to telemetry is needed to meet the requirements of an implantable device.

Compressive sensing has been recently studied in the context of biological signals (e.g. ECG [39, 105], EEG [30] and iEEG [70]) to tackle the data rate issue. When compared to thresholding and activity-dependent recording, CS has the advantage of preserving the temporal information and morphology of the signal for the entire recording period. It is also possible to apply CS along with other methods (such as interpacket redundancy removal, Huffman coding [105] or dynamic power management of the front-end LNA) in order to further relax the stringent energy and bandwidth requirements of implantable system.

While the majority of research presented in literature focus on power minimization of the implantable system, there is also a stringent need to minimize the circuit area in order to include the highest number of recording units into the available die area. Large-scale recording of cortical activity is particularly important in the case of diseases like epilepsy which spread over wide regions of cortical area. The state-of-the-art

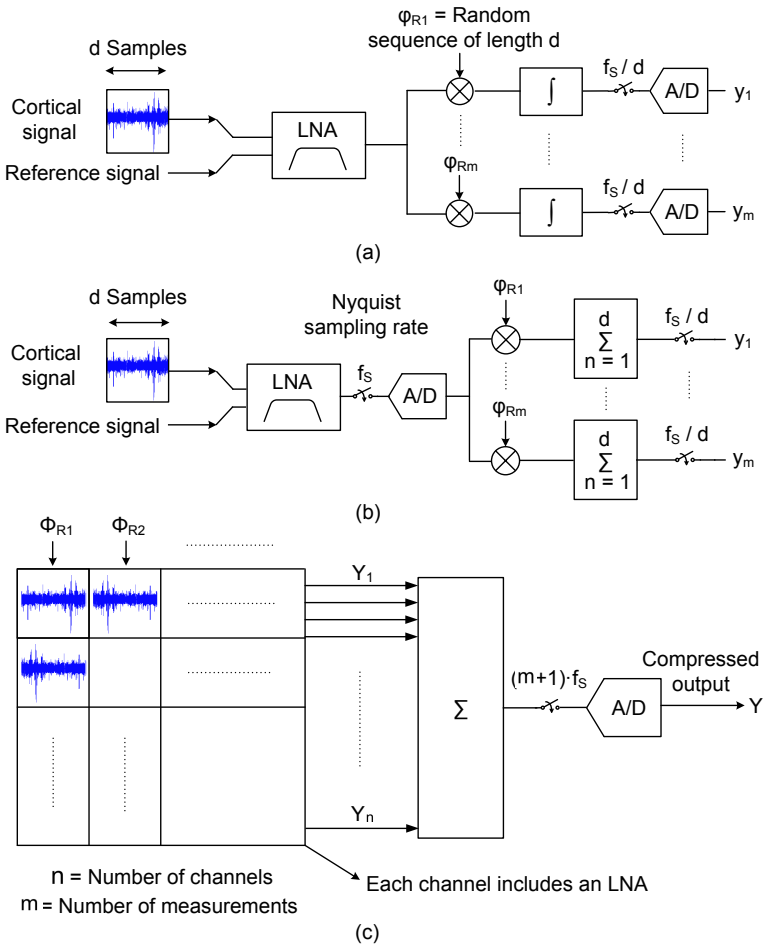


Figure 6.1: Block diagram of (a) the analog single-channel CS, (b) the digital single-channel CS and (c) the proposed multichannel CS architectures.

research targeting such applications progresses toward minimally invasive flexible and dense recording arrays with high-resolution recording capability of intracranial EEG (iEEG) signals [145, 168]. The high resolution (i.e., small spacing of recording sites) provides the capability of capturing higher frequency activity than traditionally recordable by large widely-spaced electrodes, giving a profound insight into the fundamental mechanisms underlying such abnormalities. Electrocorticographic signals recorded from human cortex can be used as an alternative to invasive spike recordings through penetrating electrodes, in order to control prosthetic limbs in BMIs as shown in [48].

The common microelectronic approach to CS ([30, 33, 91]) consists of on-the-fly compression of consecutive samples of each recording unit over time, either in analog (Figure 6.1(a)) or digital (Figure 6.1(b)) domain. Even though this approach results in a significant energy efficiency, its large area usage disqualifies the concept for a multichannel recording interface which should include the circuits supporting many channels in a lim-

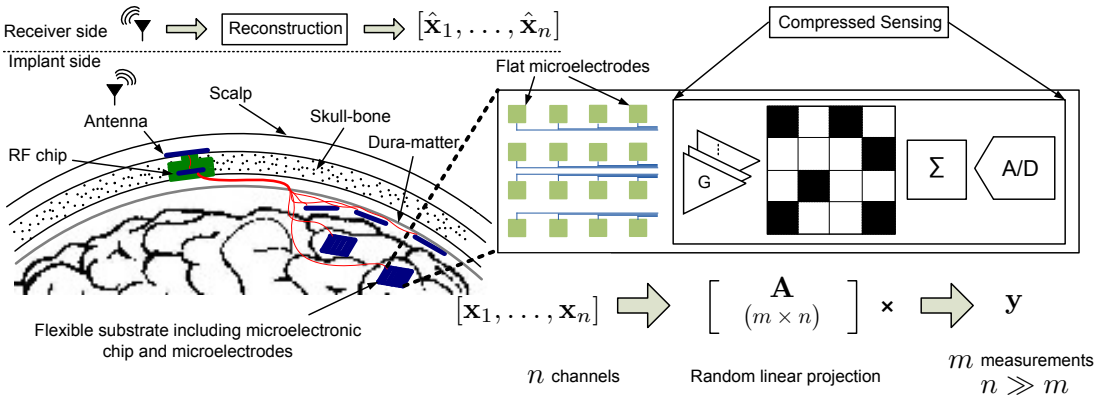


Figure 6.2: System-level view of the proposed multichannel compressive sensing method used in an implantable neural recording interface [141]. The neural signals sensed at the electrode sites are amplified, randomly projected, summed up and digitized through a single on-chip ADC. An RF unit placed within a burr hole in the skull transmits the compressed and digitized data originating from several recording chips to an external receiver and powers the implanted system.

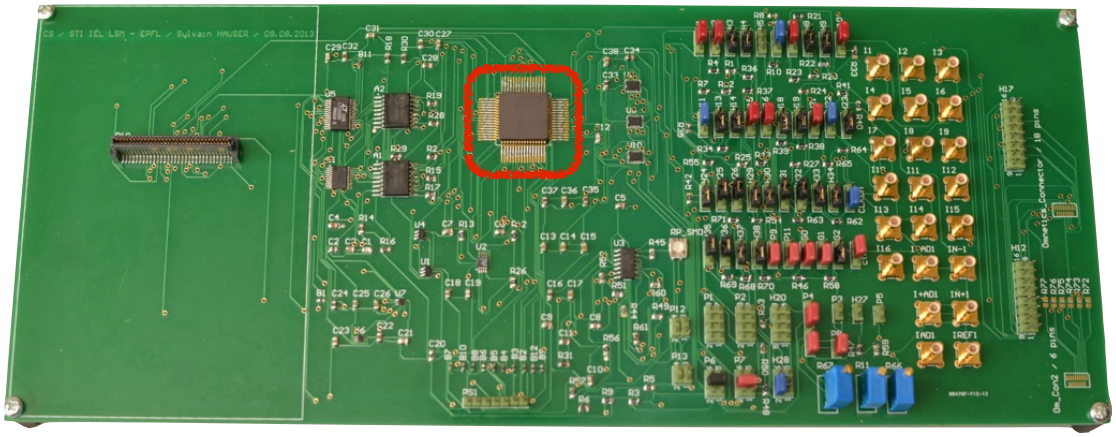


Figure 6.3: Measurement setup for clock and pre-recorded iEEG signal generation along with our proposed iEEG compressive acquisition chip

ited die area. To overcome this issue, a new multichannel measurement scheme (Figure 6.1(c)) along with an appropriate recovery scheme are proposed, which encode the whole array to a single compressed data stream. In the proposed approach, the compression is carried out in the analog domain, and in a multichannel fashion. This technique circumvents the need to place one ADC per channel and results in a significant area saving. Figure 6.3 shows our iEEG compressive acquisition chip on the measurement board. Based on our acquisition technique, a wireless monitoring system consisting of several recording/compressing units is proposed (Figure 6.2). Taking benefit of the area-

efficient implementation of CS, the number of recording units implantable on the cortex which satisfy the energy constraints of the system is scaled up by a factor equal to the compression ratio.

6.2 An Introduction to Compressive Sampling

Let $\mathbf{x} \in \mathbb{R}^n$ represents a vector with S non-zero coefficients, an $S-$sparse vector. According to compressive sensing, it is possible to recover $\mathbf{x}$ with $m \ll n$ non-adaptive linear measurements [9, 22, 23, 41]. The compressive acquisition is modeled similar to linear inverse problem where the general linear operation is replaced by a specific measurement\sensing matrix $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^m$. The compressive measurements are formed by

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{6.1}$$

The goal is to recover the vector $\mathbf{x}$ form the observation vector $\mathbf{y}$. Unfortunately, this problem is ill-posed and there are many solutions that fulfill the acquisition model (6.1). However, we know that the original signal is sparse, therefore we can exploit the prior knowledge on the signal and just look for the signals that are sparse and fulfill (6.1). This means that we do not look for any signal but the sparse constraint restricts the problem to a sparse solution therefore the signal recovery problem becomes well-posed given the measurement matrix is properly constructed. In general, the sparsity level of the signal is not known in advance. A natural way is to search for the sparset solution that satisfies the measurement constraint $\mathbf{y} = \mathbf{A}\mathbf{x}$. As explained in Section 2.2, the sparse recovery reads

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \ \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{6.2}$$

The solution to the problem (6.2) is equal to the original vector $\mathbf{x}$ if the measurement matrix is appropriately constructed.

In the CS literature, one of the largely used tools to ensure a successful recovery of a sparse or compressible vector $\mathbf{x}$ from the linear measurements is the Restricted Isometry Property (RIP) [26, 27].

Definition 6.2.1 (**Restricted Isometry Property**)**.** Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ satisfies the restricted isometry property for all S-sparse vector $\mathbf{x} \in \mathbb{R}^n$ and a small restricted isometry constant $\delta_S < 1$ if

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2. \tag{6.3}$$

For orthogonal matrices the isometry constant is equal to zero for any sparsity level. The RIP condition implies that the eigenvalues of the restricted matrix $\mathbf{A}_S^\top \mathbf{A}_S$ are in

$[1-\delta_S, 1+\delta_S]$ for any subset $\mathscr{D} \subseteq \{1,\dots,n\}$. RIP is a strong condition for practical designs, yet we can hope that if a measurement matrix satisfies RIP condition with a sufficiently small constant δ_S, then a perfect reconstruction of the $S-$sparse vector $\mathbf{x}$ can be acquired from the observations $\mathbf{y}$ even in the presence of small perturbation.

For ℓ_0 minimization if one was able to find the $S-$sparse vector through ℓ_0 minimization and if $\mathbf{A}$ satisfies the RIP of order $2S$ then there is a unique solution to $\mathbf{y} = \mathbf{Ax}$. However, the relaxed ℓ_1 minimization requires a stronger sufficient condition to guarantee the exact recovery of the original vector $\mathbf{x}$.

Theorem 6.2.1 ([128])**.** *Let matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ satisfies RIP of order $2S$ with constant $\delta_S \leq \frac{1}{3}$. Then every $S-$sparse vector $\mathbf{x} \in \mathbb{R}^n$ is the unique solution to the problem* (6.2)*.*

The above theorem guarantees that if the measurement matrix satisfies the RIP of order $2S$ with δ_S sufficiently sparse, then the exact recovery of the original signal is possible. However, this theorem is for a perfect scenario where the signal is exactly $S-$sparse and the acquisition is noise-free, which is not realistic in real problems. The following theorem shows that if we consider a stronger condition on the measurement matrix, we can recover a meaningful solution from the compressive acquisition in the presence of measurement noise and if the signal is not exact-sparse.

Theorem 6.2.2 ([24])**.** *Let $\mathbf{A}$ be an $m-$by$-n$ matrix with the restricted isometry constant $\delta_{2s} < \sqrt{2}-1$. Suppose that the noisy measurement $\mathbf{y} = \mathbf{Ax} + \mathbf{n}$ where $\|\mathbf{n}\|_2 \leq \epsilon$ represents the additive measurement noise, then the recovered vector from the compressive measurements can be recovered by*

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}}{\operatorname{argmin}} \quad \|\mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{y}-\mathbf{Ax}\|_2 \leq \epsilon. \tag{6.4}$$

Then the following relation

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq C_0\epsilon + C_1 \frac{\|\mathbf{x}-\mathbf{x}_s\|_1}{\sqrt{s}}, \tag{6.5}$$

for some constants C_0, C_1 depends only on δ_{2S}. The vector $\mathbf{x}_S$ is the best $S-$term approximation[1]

In (6.1), the first term is originated from the presence of measurement noise and the second term measures the deviation of the vector $\mathbf{x}$ from the exact sparse representation $\mathbf{x}_S$.

[1]The best $S-$term approximation of a vector $\mathbf{x} \in \mathbb{R}n$ reads

$$\underset{\mathbf{x_S}\in\mathbb{R}^{\mathbf{n}}}{\operatorname{argmin}} \quad \|\mathbf{x}-\mathbf{x}_2\|_2 \quad \text{s.t.} \quad \|\mathbf{x}_s\|_0 \leq S.$$

This approximation keeps the S highest entries of $\mathbf{x}$ and thresholds the remaining coefficients to zero.

6.2.1 Random Sampling

We discussed the sufficient condition based on RIP to recover an exact sparse or compressible vector form compressive measurements. In the previous section, we have presented sufficient conditions based on the RIP that guarantee stable and accurate reconstructions of sparse or compressible signals. However, we have not mentioned so far how to construct such good sensing matrices. In this section, we present several matrices constructed using a random process and satisfying the RIP with high probability.

Sub-Gaussian Matrices

Gaussian matrices and Bernoulli matrices are example of sub-Gaussian matrices. A Gaussian matrix is formed by identically and independently sampling a Gaussian distribution $\mathcal{N}(0, 1/m)$, where m is the number of measurements. Likewise, a Bernoulli matrix takes i.i.d. samples in $\{-1/\sqrt{m}, 1/\sqrt{m}\}$ with equal probability.

Theorem 6.2.3 ([8])**.** *Let matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ be a sub-Gaussian matrix. The matrix $\mathbf{A}$ will satisfy the RIP of order S and constant $\delta_S \leq \delta$ with overwhelming probability if*

$$m \geq CS\log(n/S), \tag{6.6}$$

for some constant C.

This result indicates that the number of required measurements to recover a signal sampled by sub-Gaussian matrices is proportional to the sparsity of the signal. Therefore, the sub-Gaussian matrices are optimal for compressive acquisition.

We should mention that the sub-Gaussian measurement matrices are universal with respect to the sparsity domain, i.e. the number of measurements require to recover the signal is independent of the sparsity domain. This property has advantage in real applications since we do not need to adapt the sensing strategy of the signal to the sparsity basis.

6.2.2 Sub-Sampled Orthonormal Matrices

The sub-Gaussian matrices though have nice features such as universality and optimal number of measurements, they are not limited for practical applications. Since the sub-Gaussian matrices are hard to implement in hardware and the recovery algorithms using these matrices is slow. One way to achieve better sensing matrices is to randomly select m column vector of an orthonormal $\Omega \in \mathbb{C}^{n\times n}$ as the discrete Fourier and Hadamard matrices. Practically, these sensing matrices are important because they can be stored efficiently and computation with these matrices is much faster that sub-Gaussian ma-

trices. However, the number of measurements required to recovery a signal sensed by these matrices is higher than sub-Gaussian matrices and unlike sub-Gaussian matrices they are not universal.

Let consider the vector $\mathbf{x}$ is sparse in some orthonormal basis $\Phi \in \mathbb{R}^{n\times n}$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a sub-sampled measurement matrix, then

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\Phi\alpha, \tag{6.7}$$

where α is an $S-$sparse vector. In order to check whether the sensing matrix satisfies the RIP, we introduce mutual coherence.

Definition 6.2.2 (**Mutual Coherence**)**.** Let Φ, Ψ be two orthonormal matrices. The mutual coherence is defined as

$$\mu(\Phi, \Psi) \triangleq \max_{1\leq i,j\leq n} |<\Phi_i, \Psi_j>| \tag{6.8}$$

The mutual coherence satisfies $n^{-\frac{1}{2}} \leq \mu(\Phi, \Psi) \leq 1$ [128]. The minimum coherence is obtained when both matrices are incoherent, e.g. Fourier transform and the identity matrix (canonical basis) are maximally incoherent. The minimum coherence is when at least one column vector in both matrices is the same.

The relation between mutual coherence and RIP of a sub-sampled matrix follows.

Theorem 6.2.4 ([128])**.** *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be an uniformly sub-sampled matrix form the orthonormal matrix $\Omega \in \mathbb{C}^{n\times n}$, and $\Phi \in \mathbb{R}^{n\times n}$ an orthonormal basis. With probability at least $1 - n^{-\gamma \log^3(n)}$, if*

$$m \geq C\delta^{-2} n\mu^2(\mathbf{A}, \Phi) S \log^4(n), \tag{6.9}$$

for some constant $C, \gamma > 0$, then the matrix $\sqrt{\frac{n}{m}}\mathbf{A}^\top\Phi$, satisfies the RIP of order S with constant $\delta_s \leq \delta$.

The above theorem indicates that the number measurements for exact recovery of sparse signal is proportional to $\mu^2(\mathbf{A}, \Phi)$. This means when the sensing matrix is not universal the number of measurements is quadratic with respect to the mutual coherence. The ideal situation is when the mutual coherence is minimum, in this case the number of measurements is optimal as with sub-Gaussian matrices. However, when the mutual coherence is maximum $\mu = 1$, the number of required measurements is equal to the dimension of the vector.

6.3 Multichannel Neural Compressive Acquisition

As discussed, compressive sensing exploits the known structures in the signals to lower the required sampling ratio below the Nyquist rate, while providing a signal recovery of acceptable quality.

In the case of multichannel neural recording, measurement limitations such as die area and power consumption suggest the use of a multichannel compression technique rather than acquiring each channel separately. Therefore, it is important to consider a measurement scheme which fulfills the physical constraints of the system. Let $\mathbf{X} \in \mathbb{R}^{d \times n}$ represent the multichannel iEEG signal where d is the dimension of the signal in each channel and in a defined time-window called *compression block* and n is the number of channels. We define a reshaping operator $\mathcal{P} : \mathbb{R}^{d \times n} \rightarrow \mathbb{R}^{n \cdot d}$ which transposes the input matrix and vectorizes the resulting matrix by concatenating its columns after each other. The linear compressive measurements are obtained by acquiring $m = p/d$ measurements from columns of $\mathbf{X}$, i.e. m measurement at each time-sample from all channels, where $p \ll d \times n$ is the total number of measurements. Hence, the multichannel sensing matrix $\mathbf{A}_{\mathrm{MC}} \in \mathbb{R}^{(md) \times (dn)}$ is represented as follows

$$\mathbf{A}_{\mathrm{MC}} = \begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_d \end{bmatrix} \qquad \mathbf{A}_i = \begin{bmatrix} a_i^{1,1} & \cdots & a_i^{1,n} \\ a_i^{2,1} & \cdots & a_i^{2,n} \\ \vdots & \ddots & \vdots \\ a_i^{m,1} & \cdots & a_i^{m,n} \end{bmatrix} \tag{6.10}$$

where $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ and a_i is uniformly selected from $\{0, 1\}$ to approximate a measurement matrix similar to the Bernoulli matrix. The schematic of neural compressive acquisition is demonstrated in Figure 6.4. The multichannel measurement vector $\mathbf{y} \in \mathbb{R}^p$ is defined as

$$\mathbf{y} = \mathbf{A}_{\mathrm{MC}}\ \mathcal{P}(\mathbf{X}) \tag{6.11}$$

6.4 Multichannel Neural Recovery from Compressive Measurements

Wideband neural signals consist of high amplitude spikes followed by a long period of low activity. As a consequence, the neural signals have a sparse structure in time domain. The lower frequency EEG signals have a sparse representation in Gabor or wavelet domains [70]. The multichannel neural signals have high inter-channel dependencies, as the signals recorded by the adjacent channels, depending on the spatial resolution and pitch of the electrodes, are delayed or scaled version of each other. Therefore, it is important to consider a model in order to employ the cross-correlations in the

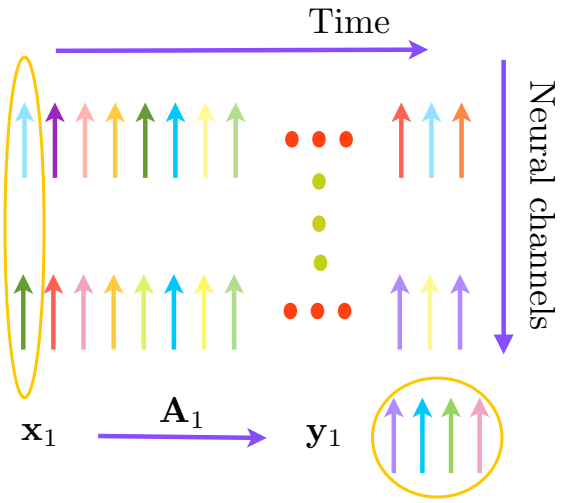


Figure 6.4: Neural signal acquisition model. A 16-channel neural signal is compressed to a 4-compressive measurements.

multichannel iEEG signals to efficiently reconstruct the underlying neural data.

6.4.1 Sparse Recovery

The simplest neural recovery from the compressive acquisition is to exploit the sparsity of the neural signals in the Gabor transform. The recovery of the multichannel iEEG signal from the compressive measurements explicitly employs a multichannel Gabor transform Φ_{MC}. The underlying neural signals in channels share similar structures. Hence, the transform domain of the multichannel signal is a block-diagonal matrix which presents the Gabor transform along the diagonal and is defined as

$$\Phi_{\text{MC}} = \mathbf{I}_n \otimes \Phi = \begin{bmatrix} \Phi & 0 & \cdots & 0 \\ 0 & \Phi & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi \end{bmatrix}, \tag{6.12}$$

$\Phi \in \mathbb{R}^{k \times d}$ is the Gabor transform, $\mathbf{I}_n$ is the identity matrix of size $n \times n$, and $\otimes$ represents the Kronecker product. Then we use ℓ_1 minimization to recover the neural signals $\mathbf{X}$ from the compressive measurements $\mathbf{y}$ using the following convex minimization

$$\underset{\mathbf{X} \in \mathbb{R}^{\mathbf{d} \times \mathbf{n}}}{\text{argmin}} \ \|\Phi_{\text{MC}}^{\top} \mathbf{X}_{\text{vec}}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}_{\text{MC}} \mathcal{P}(\mathbf{X}) \tag{6.13}$$

where $\mathbf{X}_{\text{vec}}$ is the vector form of $\mathbf{X}$ which includes the concatenation of its columns. The measurement consistency in the recovery algorithm ($\mathbf{y} = \mathbf{A}_{\text{MC}}\ \mathcal{P}(\mathbf{X})$) pertains to obtaining a signal $\mathbf{X}$ which is in accordance with the measurements $\mathbf{y}$ through the measurement

matrix $\mathbf{A}_{\mathrm{MC}}$, i.e. (6.13) recovers a signal ($\mathbf{X}$) which is sparse in the Gabor transform and satisfies the measurement consistency constraint.

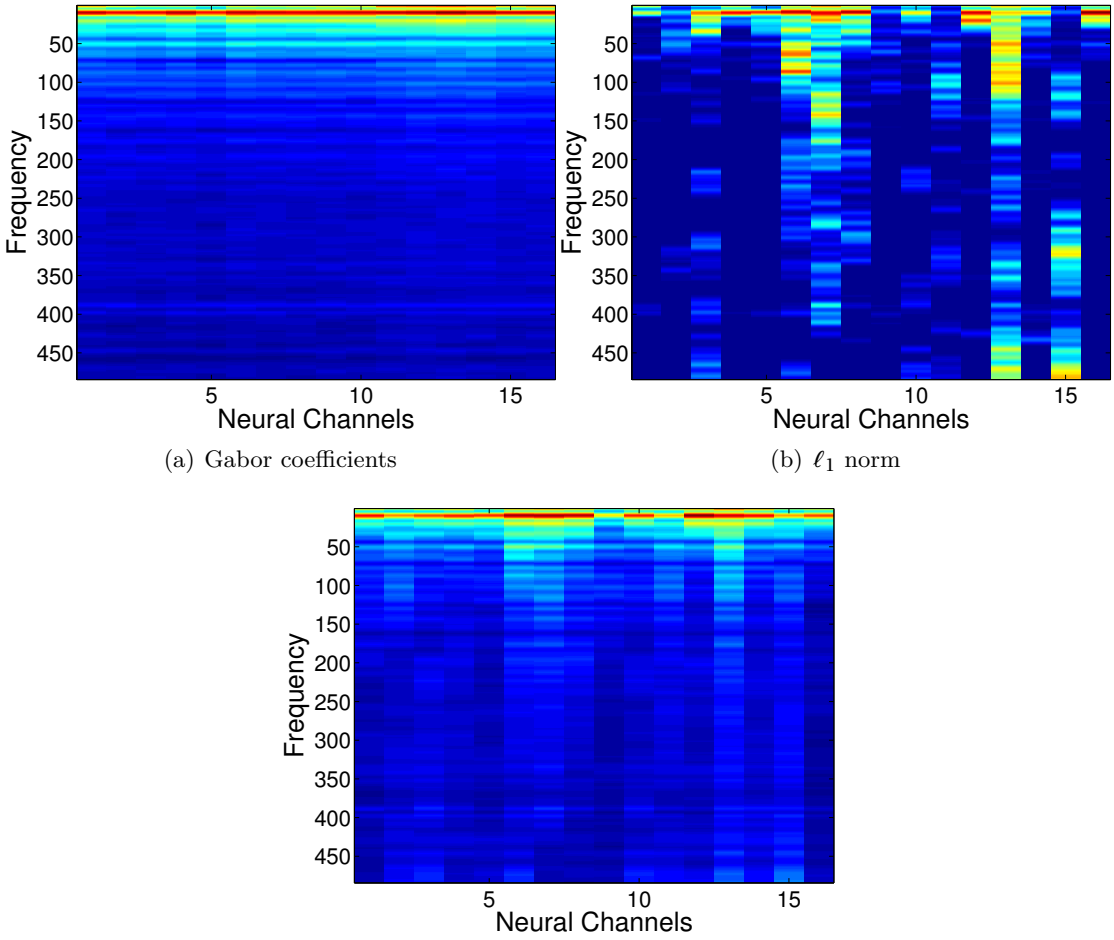
6.4.2 Mixed Norm Recovery

The multichannel neural signals have high inter-channel dependency, as the signals recorded by the adjacent channels are delayed or scaled version of each other, depending on the spatial resolution and pitch of the electrodes which indicates the propagation of neural activity within the brain. The dependent structure of multichannel neural signals suggests the design of a recovery model which exploits the similarity of neural signals. The Gabor coefficients of iEEG signals recorded by adjacent channels in a Gaussian window are shown in Figure 6.5(a). The Gabor coefficients are observed to follow similar activity among neural channels for each frequency.

As discussed in Section 2.3, we explained that the sparse prior induces coefficient-wise sparsity without considering the inter-channel correlation between coefficients. However, the neural recovery should also model the cross-correlations of iEEG signals to improve the reconstruction quality. An appropriate model for multichannel neural signals should highlight the group structure of Gabor coefficients, i.e. the model should lead to a sparsity on the number of active frequencies and promote similar activity on the neural channels for the selected frequency.

We model the dependency of neural signals using the group Lasso (the mixed $\ell_{2,1}$ norm). As explained in Section 2.3.1, the group Lasso discards or retains a group of coefficients together, since the same threshold is applied to the ℓ_2 norm of each group. However, ℓ_1 norm shrinks each coefficient independently. Consequently, the ℓ_1 recovery does not model the block structure of the neural signals.

However, the group Lasso respects the group structure of neural signals and imposes sparsity on the group of coefficients rather than each coefficient independently. Thus, the mixed $\ell_{2,1}$ norm promotes dense blocks for a sparse number of frequencies. Figure 6.5 compares the recovered Gabor coefficients of multichannel neural signal employing sparse and group Lasso recovery in a Gabor window with the Gabor coefficients of the original neural signal. In Figure 6.5, we observe that the recovered Gabor coefficients using the group Lasso yield the same structure as of the original multichannel neural signals. Furthermore, the sparse recovery (Figure does not respect the group structure of neural signals and results in recovery of Gabor coefficients which are sparse and independently spread along different frequencies for each neural channel. This behaviour is explained by the fact that the sparse recovery does not consider the group structure of neural signals. The solution to the multichannel neural recovery using the mixed $\ell_{2,1}$ norm is



(a) Gabor coefficients

(b) ℓ_1 norm

(c) $\ell_{2,1}$ mixed norm

Figure 6.5: The structure of Gabor coefficients of multichannel neural signals in a Gabor window. (a) Original neural singals: the Gabor coefficients represent group structure along frequencies. (b) ℓ_1 norm: the Gabor coefficients are scattered and recovery does not respect the group structure. (c) Group Lasso: the joint recovery preserves the group structure of Gabor coefficients.

obtained by replacing the ℓ_1 norm by the mixed norm as

$$\Psi_{\mathrm{MC}} \underset{\alpha_{\mathrm{MC}} \in \mathbb{R}^{\mathbf{d}\cdot\mathbf{n}}}{\operatorname{argmin}} \|\alpha_{\mathrm{MC}}\|_{2,1} \quad \text{s.t.} \quad \mathbf{y} = \Phi_{\mathrm{MC}}\, \mathcal{P}(\Phi_{\mathrm{MC}}\, \alpha_{\mathrm{MC}}). \tag{6.14}$$

6.5 Microelectronic Architecture

In this section we briefly describe our sensing architecture more detail is included in [70, 142]. The system architecture of the proposed spatial compression scheme to realize the multichannel sampling strategy presented in the previous Section is depicted in Figure 6.6.

Each channel of the recording scheme contains a low-noise amplifier (LNA) for boosting the low-amplitude recorded signals in the front-end of the system. The amplified signals of the individual channels are sampled on C_S and kept constant during m measurements. The linearity of the track-and-hold circuit is guaranteed by using PMOS source-to-bulk connected source-followers. The sampled signal charges the holding capacitor in the first half cycle of the clock. In the second half, the holding capacitors of all channels are connected to the integrating capacitor, based on the random value controlling the in-pixel switch. Thus, the signals of all channels in the array are multiplied by the instantaneous random value and summed together on C_{INT} ($C_{INT} \gg C_H$). The compressed voltage $V_{out}(n)$ can be written as:

$$\frac{C_H \phi_{R1}(n) V_1(n-1/2) + \ldots + C_H \phi_{Rn2}(n) V_{n2}(n-1/2)}{C_H \phi_{R1}(n) + \ldots + C_H \phi_{Rn2}(n) + C_{INT}}, \tag{6.15}$$

where $V_i(n)$ is the tracked level of the signal originating from channel number i at time nT, with T being the period of the clock signal. $\phi_{Ri}(n)$ is the level (1 or 0) of the random sequence applied to i^{th} channel at time nT and n is the number of channels.

In the proposed method, compressive samples are acquired from different locations and electrodes in the spatial domain, rather than over time. As a significant advantage, this design encodes the full array to one single data which is digitized using a single ADC. As a benefit of compressive sensing, the sampling rate of the latter ADC is n/m times smaller than the sampling rate of the unique ADC which is required in a non-compressed but time-multiplexed topology. Thus, the cost of implementation in terms of in-pixel area and power is much less than previous topologies, including non-compressed and single channel compressed schemes. Using a differential topology (Figure 6.6), the non-linearity and DC components caused by the source follower buffer circuit are partially removed. As an alternative, an active integrator can be used to perform the full array randomized integration and boost the signal level at the cost of an additional operational transconductance amplifier. The required speed of this amplifier is proportional to the size of the array which dictates the measurement number m. The compressed signal (Vout in Figure 6.6) passes through a variable-gain amplifier (VGA) to further boost the level of the signal and drive the ADC.

6.5.1 Pseudo Random Matrix Generation

The actual implementation of compressive sensing requires an efficient generation of the measurement matrix in terms of power consumption and area overhead. In a single channel approach, each channel needs to be loaded with q sequences, building the rows of the measurement matrix. Multiplication and integration in the analog domain (or summation in digital domain) is performed in q paths. In the proposed spatial compression scheme on the other hand, each channel is loaded with only one sequence. The measurement matrix supporting the first m measurements required for recovering the

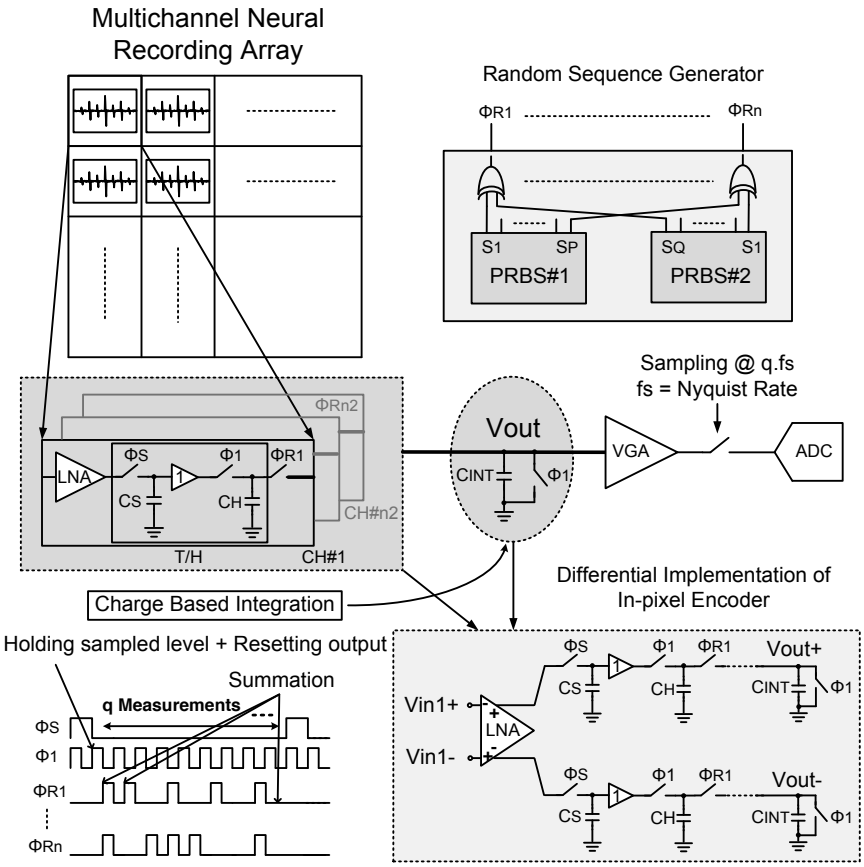


Figure 6.6: Proposed multichannel compressive acquisition scheme for iEEG recording.

first sample of each channel is created by taking the first m values of the in-channel sequences. Pseudorandom sequences which exhibit low coherence with any fixed sparsity basis [33] are a proper choice for the implementation of the measurement matrix.

In this design, the sequence generation is achieved by XORing the multiple outputs of maximum different length Pseudo Random Bit Sequence (PRBS) generators (Figure 6.6). Considering a test recording array of 4×4 and a value of m equal to 4 (Ratio Compression = 4), the 16 sequences driving the individual channels are generated by XORing the states of a 4-bit PRBS generator with another 5-bit PRBS generator. True Single-Phase Clocked (TSPC) flip-flops are used resulting in very low power consumption and a compact implementation. A small number of 9 flip-flops and 16 XOR gates is sufficient to generate the required sequences for 16 channels. While a single channel compression block has to be physically designed for a specific predefined q and redesigned for different compression ratios, the proposed scheme is easily adaptable for different values of m by simply adjusting the clock frequency. The same circuit can be used for different compression ratios and the only change is in the reconstruction code which receives m as a parameter. Therefore, the proposed scheme can be perfectly tuned

to find the appropriate compression ratio based on the diagnostic and medical considerations which impose the acceptable level of loss in the recovered data with respect to the original neural signal.

6.6 Experimental Results

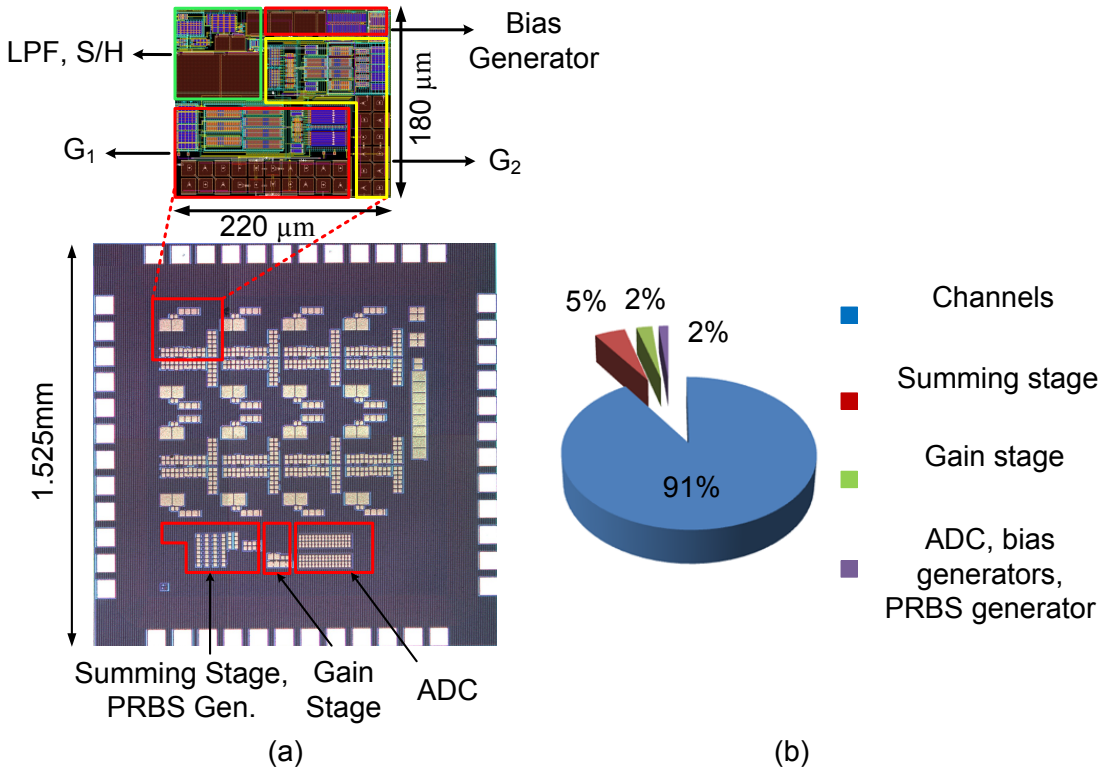


Figure 6.7: (a) Microphotograph of the chip and an individual channel's layout, (b) power breakdown of a 16-channel compressive sensing array.

The microphotograph of the fabricated chip is shown in Figure 6.7(a). The total current consumption of the chip including the buffers and bias generators is 140 μA drawn from a 1.2 V power supply, corresponding to effective current of 8.75 μA per channel. The achieved power density of the system is 7.2 mW/cm^2, significantly below the safety limit of 80 mW/cm^2 [156] for an implantable system. The contribution of different blocks of the system to the total power consumption is shown in Figure 6.7(b).

In order to demonstrate the effectiveness of the proposed acquisition model, a long segment of multichannel iEEG signal recorded from subdural strip and greed electrodes implanted on the left temporal lobe of a patient with medically refractory epilepsy have been used as the input. The signals are recorded during an invasive pre-surgical evaluation phase to pinpoint the areas of the brain involved in seizure generation and

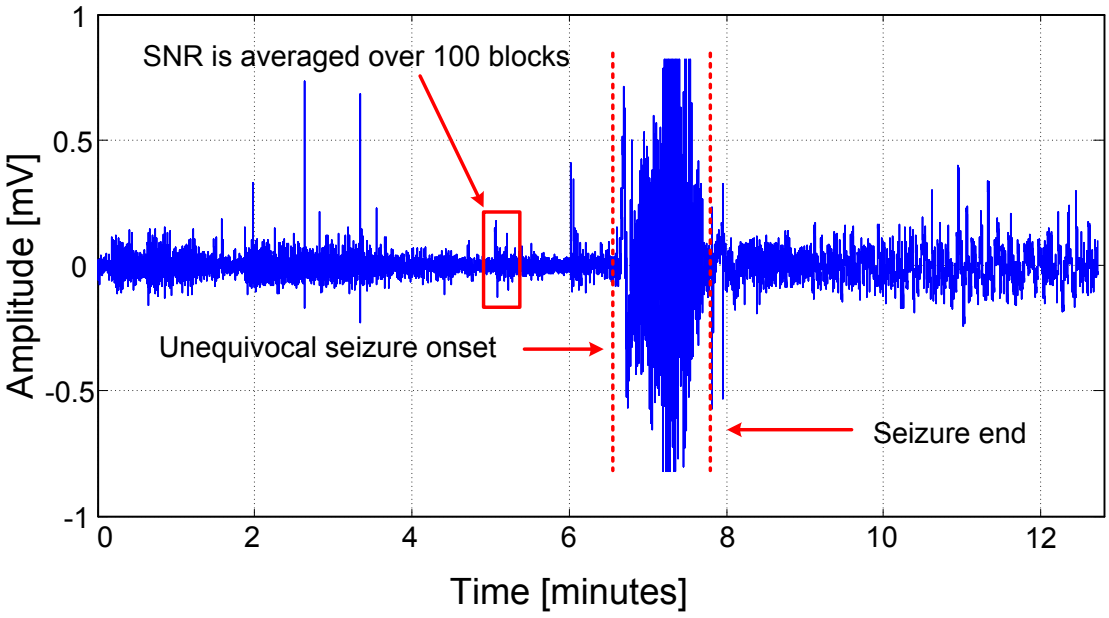
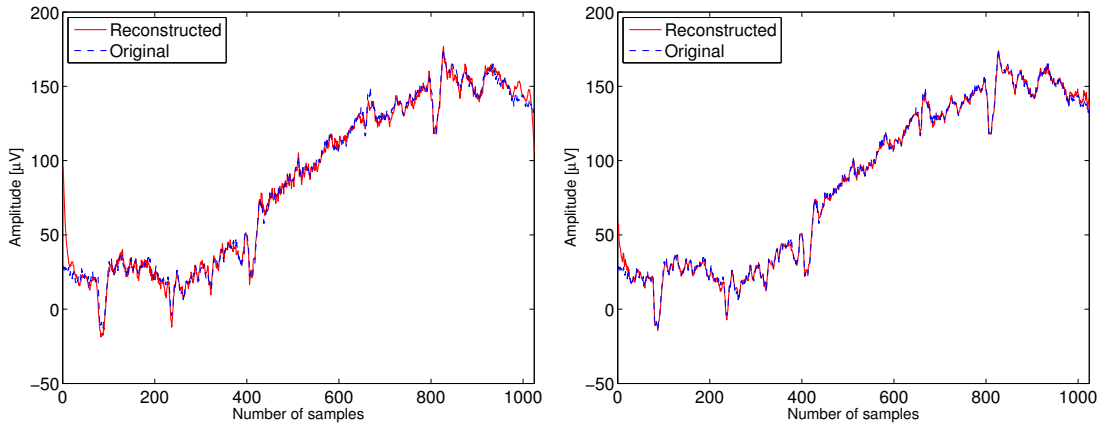


Figure 6.8: One channel of human intracranial EEG recording using strip and greed electrodes implanted on the left temporal lobe. The recovery SNR is calculated by averaging over 100 blocks of signal in the low-voltage fast activity region.

to study the feasibility of a resection surgery. This data includes minutes of pre-ictal, ictal and post-ictal activities sampled at 32 kS/s, using Neuralynx. The signals recorded by 16 adjacent channels of a greed of the electrodes are applied into the proposed CS system.



(a) Sparse recovery

(b) Group Lasso recovery

Figure 6.9: Comparison of recovery performance using different reconstruction methods for a block of length 1024 and compression ratio of 4; (a) sparse recovery $\text{SNR}_{\text{CH1}} = 21.3$ dB; (b) group Lasso recovery $\text{SNR}_{\text{CH1}} = 28.04$ dB.

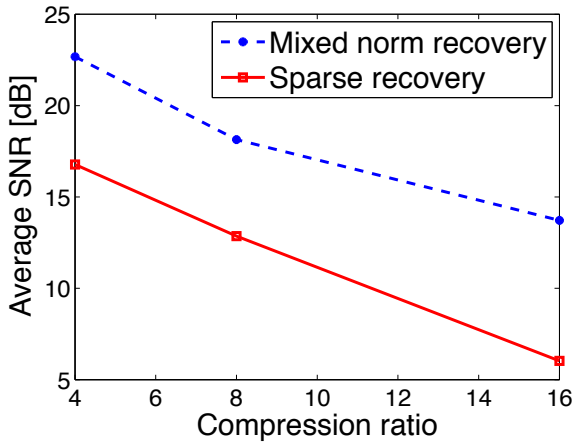


Figure 6.10: Comparison of mixed norm and sparse recovery performance for different compression ratios. SNRs are averaged over 20 compression blocks.

6.6.1 Recovery Performance

As previously explained, we employ Gabor transform as the sparsity domain of neural signals for multichannel neural recovery based on sparse and mixed norm methods. The recovery SNR of the reconstructed signal ($\hat{x}$) with respect to the original signal (x) is calculated from

$$\text{SNR} = -20\log_{10}\|x-\hat{x}\|_2/\|x\|_2 \tag{6.16}$$

for each recording channel (e.g. SNR_{CH1} represents the recovery SNR of first channel). The mean SNR of 16 channels are averaged over 100 blocks of the signal, as shown in Figure 6.8. The performance of the circuit is validated for low-voltage fast activities which are shown to be associated with seizure onset. The reconstructed signals versus the original signals corresponding to one block of a single channel data using ℓ_1 and $\ell_{2,1}$ norm minimization are shown in Figure 6.9(a) and Figure 6.9(b). The length of each compression block (d) is equal to 1024 samples and is equivalent to 256 msec at a 4 kHz sampling frequency. The digitized data after ADC is used for recovery. As shown in these figures, applying the $\ell_{2,1}$ recovery on the compressed data produced by the adjacent channels results in an improved recovery performance, compared to the sparse recovery. The averaged SNRs using the ℓ_1 and $\ell_{2,1}$ recovery are 16.64 and 21.80 dB, respectively. Based on the statistical analysis reported in [67], a minimum SNR of 10.45 dB (corresponding to a PRD of 30 %) is acceptable to maintain the diagnostically important data in the recovered signal, e.g. for successful seizure detection. Reducing the number of measurements to $m = 1$, i.e. CR = 16 results in average SNR = 13.72 dB, using $\ell_{2,1}$ recovery. Thus, the system is able to successfully recover low-voltage iEEG signals compressed by a ratio as high as 16. Figure 6.10 presents the average reconstruction SNR for sparse and joint recovery for different compression ratios. The achieved compression

ratios vary within the range of $16/15$ up to 16, corresponding to $\mathrm{CR} = 16/m$, where m is an integer number between 1 and 15.

As confirmed by the average SNR, the system is potentially able to recover the signal over the entire recording period, with some tolerable loss (i.e. $\mathrm{SNR} > 10.45$ dB). However, the amplitude and frequency of the recorded signals can significantly vary, depending on the distance between the microelectrodes and their surface area [20]. The amplitude of the very high frequency oscillations ($250-500$) recorded using macroelectrodes [79] is much smaller ($5-30$ versus $100-1250$ μV) than human fast ripples recorded from smaller microelectrodes [20]. The amplitude of the fast ripples recorded by the high-density arrays will expectedly fall within the range of tens-hundreds of microvolts, which is efficiently recovered at the output of the system, as shown in Figure 6.9.

The required time for the reconstruction of the $1024-$length epochs of the multichannel signal is 1.12 second per channel, using a 2.66 GHz processor with 4 GB of RAM. To achieve a real-time performance, the speed of the reconstruction could be improved by hardware implementation of the algorithm on FPGA and using custom acceleration techniques.

6.6.2 Effect of Circuit Non-idealities and Non-adjacent Channels

A second database consisting of multichannel intracranially recorded signals of the slices of a rat somatosensory cortex under bicuculline, which blocks the synaptic inhibition and consequently mimics the epilepsy, is applied to the CS recording system. This signal includes epileptiform burst activity and extracellularly detected spikes. However, the verifiable signals of this database are associated with electrodes randomly located on the array. Consequently, these channels exhibit limited synchronous activity during the seizure, compared to the previous database. This effect is reflected in relatively lower recovery SNR of neural signals presented in Table 6.1. SNR is evaluated for each channel and for the multichannel signal, by comparing $\mathbf{X}_{\mathrm{vec}}$ with the reconstructed multichannel data stream (defined as $\mathrm{SNR}_{\mathrm{T}}$).

Table 6.1: Recovery Quality In the Presence of Noise.

Performance Metric	Gabor Transform, ℓ_1 recovery, CR = 4				
	×QN[a] ×cktN[b]	✓QN ×cktN	×QN ✓cktN	✓QN ✓cktN	*MAT LAB*
$\mathrm{SNR}_{\mathrm{CH1}}$	14.83 [dB]	14.64 [dB]	14.75 [dB]	14.32 [dB]	15.12 [dB]
$\mathrm{SNR}_{\mathrm{T}}$	10.82 [dB]	10.52 [dB]	10.76 [dB]	10.23 [dB]	10.97 [dB]

[a] Excluding the quantization noise.
[b] Excluding the noise of the circuit.

In order to study the effect of circuit non-idealities (such as quantization and thermal noise) on the recovery performance, a comparison of SNR is also presented in Table 6.1. The reconstruction results are compared by including and excluding different noise sources in simulations. The results are compared to the recovery performance of the compressed signal generated by matrix multiplication in MATLAB, using the same matrix as the output of the on-chip PRBS generator. CS can improve the attainable signal fidelity in the presence of sensor noise as shown in [31]. Although the reconstruction performance of a CS system is not as good as a simple quantizer for noiseless inputs [61], for more practical noisy signals recorded by the sensors, CS achieves a better performance, i.e. lower energy and improved PRD (Percentage Root-Mean Squared Difference) [31]. The CS system filters some of the input noise during reconstruction [31], and consequently is a correct choice for noisy environments such as a neural interface. Otherwise expressed, the recovery algorithm discards the coefficients below a certain threshold in the sparse representation of the signal. The discarded coefficients can be interpreted as filtered noise.

The recovery performance of the proposed system is marginally affected by the quantization noise of the ADC, as confirmed by the results of simulations presented in Table 6.1. Excluding the circuit noise in simulations (thermal and flicker) results in a negligible improvement of the recovery performance which confirms the robustness of the CS system against non-idealities induced by the circuit. Consequently, the specifications related to the resolution of the ADC, the required noise performance of the analog front-end and the summing stage preceding the ADC and therefore the total power consumption and area of the chip can be further relaxed without jeopardizing the recovery performance.

6.6.3 Architecture Performance Comparison

Table 6.2 summarizes the performance of the system and presents a comparison with published works. In this table, compression power and area refer to the extra power consumption and area usage of the signal digitization, compression and thresholding blocks which are commonly added to the total power consumption and area of the analog front-end. The authors in [33] apply compressive sensing on a single-channel pre-recorded EEG data by acquiring measurements in the digital domain. The power saving is significant while the area overhead is not addressed. Due to the youthfulness of the field and the lack of similar electronic architectures that use CS in brain implants, we have compared our results with a Discrete Wavelet Transform (DWT)-based design [81] for intra-cortical implants and several additional systems based on spike/AP detection [60, 64, 118, 132] for implantable neural recording applications. While the design in [81] mainly addresses the area-efficiency of the implantable system and proposes an architecture that sequentially evaluates the DWT of the multichannel data in the digital domain, our results outperform this approach in terms of area and power efficiency. In addition, high compression ratios are achieved by means of the following

thresholding and redundancy removal stages, while the DWT by itself does not result in any data compression. Thresholding, however, results in a significant loss of the signal in non-spiking regions while a more precise recovery is achieved at much lower compression ratios (e.g. at $CR = 2$ in [81]). The chip includes several memory registers containing threshold values of different channels and additional blocks such as controllers, address generator and buffer units which degrade the power and area efficiency of the system. Some of the reported spike detector systems achieve significant data reduction [64, 132] with negligible overhead in terms of compression power and area [132]. However, the patient-specific threshold setting in such systems can result in design complexity in a real neural interface in addition to the loss of signal in non-spiking regions. Furthermore, the transmitted signal may not be acceptable to the clinicians who usually prefer to have access to the entire iEEG data, even though somewhat lossy, for a thorough neurological examination.

Table 6.2: Comparison of system performance with published literature.

Parameter	[33]	[132]	[64]	[60]	[81]	Ours
Tech. [μm CMOS]	0.09	0.13	0.5	0.18	0.5	0.18
Power supply [V]	0.6	1.2	3.3	1.8	-	1.2
Comp.[a] method	DCS	PWL SD[b]	SD	AP det.	DWT SD	MCS
Number of channels	1	1	100	16	32	16
Comp. area [mm^2]	0.103	0.080	< 0.160	> 0.0475	0.18	0.008
Comp. power [μW]	1.9	1.18	27	> 96	95	0.95
Sampling rate [kS/s]	≤20	90	15	30	25	4
CR	≤10	125	150	48	≤20	≤16

[a] Compression
[b] Spike Detection

6.7 Conclusion

A new multichannel architecture for compressive recording of cortical signals at the surface of the cortex is proposed. In addition to the area efficiency, the proposed method is easily adaptable to different compression ratios, depending on the sparsity of the input signals. The power efficiency resulting from the compressive sensing methodology in addition to the minimal area cost, make this approach highly relevant for power- and area-constrained multichannel sparse signal acquisition. This approach can be investigated in other applications than neural recording, which require data recording from multiple nodes. Extensive system-level analysis and simulations confirm the relevance and efficiency of the system for high-density recording applications, compared to alternative compression methods.

Chapter 7

Conclusion

This thesis argues that many problems in machine learning and signal processing produce massive amount of data but with inherent structures. We believe the data deluge requires a set of new strategies to establish connection between data acquisition and processing. As a first step toward this goal, this thesis presents solutions that prove fruitful focus on this aspect. Our approach leverages data structures to propose accurate algorithmic schemes based on convex optimization tools to gain insights from high-dimensional data.

In this thesis, we show that considering data structures to design feasible algorithms is a key that leads to rigorously reduce the acquisition time, sampling rate and transmission power and also provides promising methods to extract information and recover the missing data points in the data flood. We believe the structure modeling of high-dimensional data stretches far beyond the applications addressed in this thesis. The structures do not restrict to the ones identified in this thesis and in fact we are at the early stage of modeling the structures and yet there are many complicated structures to be investigated such as underlying structures in social networks and connection between brain regions.

In the following, we explain the possible future direction of the applications discussed in this thesis. We hope that this thesis inspires follow-on work to leverage data structures to acquire information from high-dimensional data.

7.1 Future Work

7.1.1 Toward Ultimate Plenoptic Camera

The light field cameras explained in this thesis assume that some light field dimensions are constant. However, as discussed in this thesis, there are strong correlations in all

light field dimensions which are not yet explored. For instance, multi-spectral light fields shown in Figure 7.1 demonstrates correlations between spectral, spatial and angular dimensions. Therefore, leveraging the correlations in all light field dimensions should be the purpose of future light field photography. However, prior to addressing the correlations in all light field dimensions, we need to develop new acquisition techniques to sample the plenoptic function. Similar to the coded acquisition, the computational photography techniques based on compressive sensing framework can be used to capture the full light field function.

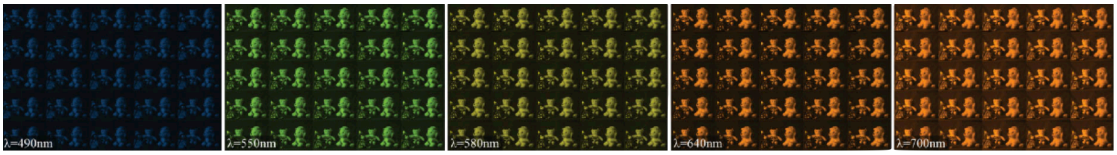


Figure 7.1: Multi-spectral light fields with 10 spectral channels and 5×5 views. The light field is captured using a X-Y translation table and a multi-spectral camera [165].

As discussed in this thesis, camera arrays or lenslet arrays provide accurate reconstruction of light fields. Light field views acquired with these methods are arranged on a regular lattice and the displacement between them is linear. Thus modeling the correlations between views can avoid the scene geometry. Another promising avenue of research is to acquire light fields using unstructured grids [38]. This acquisition model requires an interactive system to guide the user to cover the scene from different viewpoints. However, to derive the correlations model between light field views, we need to consider more sophisticated structures and the modeling cannot discard the scene geometry.

7.1.2 Cortical Recording

There is a wide possibility of hardware improvements for the explained cortical signals acquisition scheme such as investigation for more advance power efficient acquisition implementation. Another promising line of research is an adaptive measurement scheme to incorporate the interaction between the encoder and decoder in the acquisition process to design an ultra-low power acquisition process. A part from the encoder investigation, on the decoder side deploying models to better leverage the correlations between neural channels can be further explored. For instance, the neural channels can be modeled as several independent active source where the measurements are acquired from the linear combination of these sources. In addition, for epilepsy seizure detection one can re-design the neural recording process such that directly acquire the seizure at the encoder and transfer the seizure to the decoder or the seizure detection can be performed directly on measurements or a coarse reconstruction of neural channels instead of their full reconstruction.

Appendix A

Tensor Algebra

This section briefly reviews some concepts and notations for tensor algebra required for light field photography. We refer the interested readers to relevant literature [34, 87].

A *tensor* is a multi-dimensional array of data which is the generalization of matrices to higher dimensions. The number of dimensions of tensors is called mode or order. A mode N tensor is denoted as $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$. A vector is a mode-1 tensor and a matrix is a mode 2 tensor.

Fibers are associated to vectors in a tensor. A fiber is defined by fixing all modes except one in a tensor. For example, mode-k fibers are all vectors derived by fixing $\{n_1, \ldots, n_{k-1}, n_{k+1}, \ldots, n_N\}$.

Slices are obtained by fixing all but two tensor indices. Slices are two-dimension section of tensors, we adopt the convention that the first unfixed index is the row index and the second is the column index of the slice.

A *rank one* mode-N tensor is constructed from the outer product of N vectors

$$\mathcal{X} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}, \tag{A.1}$$

$\circ$ represents the vector outer product. Each element of the tensor is the product of the corresponding vector elements, i.e.

$$x_{i_1, i_2, \ldots, i_N} = u^{(1)}_{i_1} \circ u^{(2)}_{i_1} \circ \cdots \circ u^{(N)}_{i_N}, \tag{A.2}$$

A.0.3 Matricization and Tensor-Matrix Product

There are many ways to assemble a tensor to a matrix, a convenient way to *unfold* a tensor to a matrix along mode-k is a matrix of dimension $n_k \times (n_1 \cdots n_{k-1} n_{k+1} \cdots n_N)$. The

mode-k unfolding of tensor $\mathscr{X}$ is denoted as $\mathbf{X}_{(k)}$. Mode-k tensor unfolding maps tensor element $(i_1, i_2, \ldots, i_N)$ to matrix element (i_k, j) such that

$$j = 1 + \sum_{m=1, m\neq k}^{N} (i_m - 1) J_m, \quad J_m = \prod_{p=1, p\neq k}^{m-1} n_p. \tag{A.3}$$

The *tensor-matrix product*, also known as *mode-k product*, of a tensor $\mathscr{X} \in \mathbb{R}^{n_1 \times \cdots \times n_N}$ and a matrix $\mathbf{A} \in \mathbb{R}^{m \times n_k}$ is denoted as

$$\mathscr{Y} = \mathscr{X} \times_k \mathbf{A} \in \mathbb{R}^{n_1 \times n_2 \cdots \times n_{k-1} \times m \times n_{k+1} \times \cdots \times n_N}. \tag{A.4}$$

Each element of the product computes

$$y_{i_1,\ldots,i_{k-1},j_k,i_{k+1},\ldots,i_N} = \sum_{i_k=1}^{n_k} x_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_N} a_{j_k,i_k} \tag{A.5}$$

for $j_k = 1, 2, \ldots, m$. The tenor-matrix product can be expressed in tensor unfolding format

$$\mathbf{Y}_{(k)} = \mathbf{A}\mathbf{X}_k. \tag{A.6}$$

For distinct modes in a series the order of multiplication does not matter, i.e.

$$\mathscr{X} \times_k \mathbf{A} \times_j \mathbf{B} = \mathscr{X} \times_j \mathbf{A} \times_k \mathbf{B}. \tag{A.7}$$

A.0.4 Kronecker and Khatri-Rao Product

The *Kronecker product* of matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{j \times k}$, denoted as $\mathbf{A} \otimes \mathbf{B}$, results in a matrix of dimension $mj \times nk$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B} & \mathbf{A}_{12}\mathbf{B} & \ldots & \mathbf{A}_{1n}\mathbf{B} \\ \mathbf{A}_{21}\mathbf{B} & \mathbf{A}_{22}\mathbf{B} & \ldots & \mathbf{A}_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1}\mathbf{B} & \mathbf{A}_{m2}\mathbf{B} & \ldots & \mathbf{A}_{mn}\mathbf{B} \end{bmatrix}. \tag{A.8}$$

The *Khatri-Rao product* between $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{j \times n}$ is defined by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \mathbf{a}_2 \otimes \mathbf{b}_2 \cdots \mathbf{a}_n \otimes \mathbf{b}_n], \tag{A.9}$$

$\mathbf{a}_\cdot$ and $\mathbf{b}_\cdot$ are column vectors of $\mathbf{A}$ and $\mathbf{B}$. The Khatri-Rao and Kronecker products are identical for vectors.

A.0.5 Inner Products and Tensor Norms

Inner product of two tensors $\mathscr{X}$ and $\mathscr{Y}$ of the same size is defined as

$$\langle \mathscr{X}, \mathscr{Y} \rangle = \sum_{i_1=1}^{n_1} \cdots \sum_{i_n=1}^{N} x_{i_1,\ldots,i_n} y_{i_1,\ldots,i_n}. \tag{A.10}$$

The induced tensor norm from inner product is defined as

$$\|\mathscr{X}\|_{\mathrm{F}} = \sqrt{\langle \mathscr{X}, \mathscr{X} \rangle}. \tag{A.11}$$

A.0.6 Tensor Decomposition

Unlike matrices, tensor rank is not uniquely defined. In this section, we discuss different tensor decomposition techniques to approximate a tensor with a few components.

CANDECOMP/PARAFAC decomposition

The CP decomposition factorizes a tensor into a sum of rank-one tensors

$$\mathscr{X} = \sum_{i=1}^{k} \mathbf{u}_i^{(1)} \circ \mathbf{u}_i^{(2)} \circ \cdots \circ \mathbf{u}_i^{(N)}. \tag{A.12}$$

The rank of tensor $\mathscr{X}$ is equal to the minimum number of rank one tensors that sum to $\mathscr{X}$.

For matrices the best k-rank approximation is given by the first k components of the matrix SVD. That is for the matrix $\mathbf{A}$ the best k-rank approximation is

$$\hat{\mathbf{A}} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \circ \mathbf{v}_i \tag{A.13}$$

For tensor approximation the best (k-1)-rank approximation is not a factor in the best k-rank approximation [87]. This yields that the tensor decomposition is not sequential.

To determine the best rank approximation of a tensor most algorithms try for different ranks until the desired approximation. Having fixed the best rank approximation, we can factorize a tensor into CP components by *Alternating Least Square* (ALS) [87]. To express the decomposition, we choose a mode-3 tensor and the best k-rank approximation boils down to

$$\min_{\mathscr{X}} \|\mathscr{X} - [[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}]]\|_{\mathrm{F}}, \quad [[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}]] = \sum_{i=1}^{k} \mathbf{u}_i^{(1)} \circ \mathbf{u}_i^{(2)} \circ \mathbf{u}_i^{(3)}. \tag{A.14}$$

The ALS algorithm fixes $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ and solves the minimization for $\mathbf{U}^{(1)}$, and then

fixes $\mathbf{U}^{(3)}$ and $\mathbf{U}^{(1)}$ computes the minimization for $\mathbf{U}^{(2)}$ and follows the same procedure for $\mathbf{U}^{(3)}$. The algorithm repeats these steps till a convergence is achieved. Having fixed all but $\mathbf{U}^{(1)}$, we can employ tensor unfoldings to have the following minimization

$$\min_{\mathbf{F}} \|\mathscr{X} - \mathbf{U}^{(1)}(\mathbf{U}^{(3)} \odot \mathbf{U}^{(2)})^{\top}\|_{\mathrm{F}}, \tag{A.15}$$

Khatri-Rao product simplifies the minimization to

$$\mathbf{U}^{(1)} = \mathbf{X}_{(1)}(\mathbf{U}^{(3)} \odot \mathbf{U}^{(2)})(\mathbf{U}^{(3)^{\mathrm{T}}}\mathbf{U}^{(3)} * \mathbf{U}^{(2)^{\mathrm{T}}}\mathbf{U}^{(2)})^{\dagger}, \tag{A.16}$$

$*$ represents the matrix elementwise product. The CP decomposition for higher-order tensor is explained in Algorithm 6.

Algorithm 6: ALS algorithm to compute CP decomposition of mode-N tensor with k components [87].

Initialization: $\mathbf{X}_{(i)} \in \mathbb{R}^{n_i \times k}$ for $n = 1, \ldots, N$
while *not converged* **do**
  for $n = 1, \ldots, N$ **do**
    $\mathbf{P} = \mathbf{U}^{(1)^{\mathrm{T}}}\mathbf{U}^{(1)} * \cdots * \mathbf{U}^{(n-1)^{\mathrm{T}}}\mathbf{U}^{(n-1)} * \mathbf{U}^{(n+1)^{\mathrm{T}}}\mathbf{U}^{(n+1)} * \mathbf{U}^{(N)^{\mathrm{T}}}\mathbf{U}^{(N)}$;
    $\mathbf{U}^{(n+1)} = \mathbf{X}_{(n)}(\mathbf{U}^{(N)} \odot \mathbf{U}^{(N)} \odot \cdots \odot \mathbf{U}^{(n+1)} \odot \mathbf{U}^{(n-1)} \odot \cdots \odot \mathbf{U}^{(1)})\mathbf{P}^{\dagger}$

Tucker Decomposition

The Tucker decomposition is a form of tensor decomposition that factorizes the tensor into a core tensor multiplied by a matrix along each mode. The Tucker decomposition of mode-N tensor [87] is :

$$\mathscr{X} = \mathscr{C} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)}. \tag{A.17}$$

The elementwise Tucker decomposition is defined as:

$$x_{i_1 i_2 \ldots i_N} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_N=1}^{R_N} c_{r_1 r_2 \ldots r_N} u_{r_1}^{(1)} u_{r_2}^{(2)} \cdots u_{r_N}^{(N)}.$$

Tucker decomposition in matrix format is calculated by

$$\mathbf{X}_{(n)} = \mathbf{U}^{(n)}\mathbf{C}_{(n)}(\mathbf{U}^{(1)} \otimes \cdots \mathbf{U}^{(n-1)} \otimes \mathbf{U}^{(n+1)} \otimes \cdots \otimes \mathbf{U}^{(N)})^{\mathrm{T}}. \tag{A.18}$$

The n-rank of a tensor $\mathscr{X}$, denoted as $\mathrm{rank}_n(\mathscr{X})$, is the column rank of $\mathbf{X}_{(n)}$, i.e. $\mathscr{X}$ has $\mathrm{rank}_{(n)} = (R_1, R_2, \ldots, R_N)$. The n-rank defines unfolding rank of a tensor and it should not be confused by the tensor rank which is the minimum number of rank one tensors.

Tucker decomposition is known as *higher-order SVD* (HOSVD) [92] . HOSVD is a generalization of matrix SVD to higher order tensor decomposition. The HOSVD decomposition applies matrix SVD to tensor unfolding along each mode. The HOSVD decomposition of a mode-N tensor is defined in Algorithm 7.

Algorithm 7: HOSVD decomposition of a mode-N tensor with $\text{rank}_n(r_1, r_2, \ldots, r_N)$ [87].

for $n = 1, \ldots, N$ **do**
  $\mathbf{U}^{(n)}$ = first r_n singular values of $\mathbf{X}_{(n)}$
$\mathscr{C} = \mathscr{X} \times_1 \mathbf{U}^{(1)^{\mathrm{T}}} \times_2 \mathbf{U}^{(2)^{\mathrm{T}}} \times_3 \cdots \times_N \mathbf{U}^{(N)^{\mathrm{T}}}$

A.0.7 Square Norm

The definition of $\text{rank}_{(N)}$ motivates to recover a low-rank matrix by the convex minimization of $\sum_i \lambda_i ||\mathbf{X}_{(i)}||_*$ from linear measurements. This approach for tensor recovery from sum of nuclear norms obtained from tensor unfoldings has widely used in [3, 52, 92, 102]. Oymak et al. [119] has shown thought it seems trivial to recover the simultaneous structures of an object by combining the convex relaxations of each structure, the recovery is not more successful than the best single regularizer. Mu et al. [112] used this proof to show that the sum of nuclear norms for tensor unfolding does not represent the tensor structure and the number of required measurements to recover the low-rank tensor is the same as number of measurements required to the tensor unfolded along just one mode. The notion of *Square norm*, which is matrix reshaping of a tensor unfolding, for convex low-rank tensors recovery is introduced in [112]. However, the number of required measurements for convex surrogate of tensor rank recovery is higher than non-convex model [112, 129]. The reason can be explained by the fact that unlike nuclear norm for matrix recovery which is the tightest convex envelop to matrix rank [131], the square norm for tensor rank is not tight.

Bibliography

[1] Raytrix. http://http://www.raytrix.de/.

[2] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, 1991.

[3] B. Ajdin, M. Finckh, C. Fuchs, J. Hanika, and H. Lensch. Compressive higher-order sparse and low-rank acquisition with a hyperspectral light stage. Technical Report MSU-CSE-00-2, Universitätsbibliothek Tübingen, February 2012.

[4] A. Ashok and M. A. Neifeld. Compressive light field imaging. In *SPIE DSS*, pages 76900Q–76900Q, 2010.

[5] A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *IJCV*, 2012.

[6] D. Babacan, R. Ansorge, M. Luessi, P. Ruiz, R. Molina, and A. Katsaggelos. Compressive light field sensing. 2012.

[7] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, et al. Structured sparsity through convex optimization. *Statistical Science*, 27(4):450–468, 2012.

[8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[9] R. G. Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[10] R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 2008.

[11] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *Information Theory, IEEE Transactions on*, 56(4):1982–2001, 2010.

[12] T. Basha, S. Avidan, A. Hornung, and W. Matusik. Structure and motion from scene registration. In *CVPR*. IEEE, 2012.

[13] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer Science & Business Media, 2011.

[14] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[15] T. Bishop and P. Favaro. The light field camera: extended depth of field, aliasing and superresolution. *PAMI*, 2012.

[16] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *Proc. ICCP*, pages 1–9. IEEE, 2009.

[17] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

[18] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *IJCV*, 1987.

[19] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 2001.

[20] A. Bragin, I. Mody, C. L. Wilson, and J. Engel. Local generation of fast ripples in epileptic brain. *The Journal of neuroscience*, 22(5):2012–2021, 2002.

[21] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy. Wave optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.

[22] E. Candès, J. Romberg, and T. Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[23] E. J. Candès. Compressive sampling. In *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, pages 1433–1452, 2006.

[24] E. J. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[25] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. of Comput. math*, 9(6):717–772, 2009.

[26] E. J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[27] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52 (12):5406–5425, 2006.

[28] E. J. Candès and T. Tao. The Power of Convex Relaxation: Near-optimal Matrix Completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[29] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, 2011. ISSN 0004-5411.

[30] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic. Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors. *Solid-State Circuits, IEEE Journal of*, 47(3):744–756, 2012.

[31] F. Chen, F. Lim, O. Abari, A. Chandrakasan, and V. Stojanovic. Energy-aware design of compressed sensing systems for wireless sensors under performance and reliability constraints. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 60(3):650–661, 2013.

[32] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[33] X. Chen, Z. Yu, S. Hoyos, B. M. Sadler, and J. Silva-Martinez. A sub-nyquist rate sampling receiver exploiting compressive sensing. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 58(3):507–520, 2011.

[34] A. Cichocki, R. Zdunek, A. H. Phan, and S. ichi Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, 2009.

[35] P. L. Combettes and J.-C. Pesquet. Proximal Splitting Methods in Signal Processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. 2011.

[36] E. D'Angelo. Patch-based methods for variational image processing problems. *PhD Thesis*, 2013.

[37] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[38] A. Davis, M. Levoy, and F. Durand. Unstructured light fields. *Computer Graphics Forum*, 31(2pt1), 2012.

[39] A. M. Dixon, E. G. Allstot, D. Gangopadhyay, and D. J. Allstot. Compressed sensing system considerations for ecg and emg wireless biosensors. *Biomedical Circuits and Systems, IEEE Transactions on*, 6(2):156–166, 2012.

[40] D. Donatsch, S. A. Bigdeli, P. Robert, and M. Zwicker. Hand-held 3d light field photography and applications. *The Visual Computer*, 2014.

[41] D. Donoho. Compressed Sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[42] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):83–91, 2008.

[43] D. Dudley, W. M. Duncan, and J. Slaughter. Emerging digital micromirror device (dmd) applications. In *Micromachining and Microfabrication*, pages 14–25. International Society for Optics and Photonics, 2003.

[44] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[45] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *CVPR*. IEEE, 2009.

[46] O. Faugeras and R. Keriven. *Variational principles, surface evolution, PDE's, level set methods and the stereo problem*. IEEE, 2002.

[47] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[48] M. S. Fifer, S. Acharya, H. L. Benz, M. Mollazadeh, N. E. Crone, and N. V. Thakor. Towards electrocorticographic control of a dexterous upper limb prosthesis. *IEEE pulse*, 3(1):38, 2012.

[49] M. A. Figueiredo and R. D. Nowak. An em algorithm for wavelet-based image restoration. *Image Processing, IEEE Transactions on*, 12(8):906–916, 2003.

[50] A. W. Fitzgibbon, Y. Wexler, A. Zisserman, et al. Image-based rendering using image-based priors. In *ICCV*, volume 3, pages 1176–1183, 2003.

[51] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM Journal on Numerical Analysis*, 46(2):577–613, 2008.

[52] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

[53] T. Georgiev and C. Intwala. Light field camera design for integral view photography. *Adobe System, Inc*, 2006.

[54] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala. Spatio-angular resolution tradeoffs in integral photography. *Rendering Techniques*, 2006: 263–272, 2006.

[55] T. Georgiev, C. Intwala, and D. Babacan. Light-field capture by multiplexing in the frequency domain. Technical report, Citeseer, 2007.

[56] B. Goldluecke and D. Cremers. An approach to vectorial total variation based on geometric measure theory. In *CVPR*, 2010.

[57] B. Goldluecke and M. A. Magnor. Joint 3d-reconstruction and background separation in multiple views using graph cuts. In *CVPR*. IEEE, 2003.

[58] J. Goodman. Introduction to fourier optics. 2008.

[59] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. SIGGRAPH*, pages 43–54, 1996.

[60] B. Gosselin, A. E. Ayoub, J.-F. Roy, M. Sawan, F. Lepore, A. Chaudhuri, and D. Guitton. A mixed-signal multichip neural recording interface with bandwidth reduction. *Biomedical Circuits and Systems, IEEE Transactions on*, 3(3):129–141, 2009.

[61] V. K. Goyal, A. K. Fletcher, and S. Rangan. Compressive sampling and lossy compression. *Signal Processing Magazine, IEEE*, 25(2):48–56, 2008.

[62] A. Gramfort and M. Kowalski. Improving m/eeg source localizationwith an inter-condition sparse prior. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 141–144. IEEE, 2009.

[63] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, and M. Kowalski. Time-frequency mixed-norm estimates: sparse m/eeg imaging with non-stationary source activations. *NeuroImage*, 70:410–422, 2013.

[64] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher. A low-power integrated circuit for a wireless 100-electrode neural recording system. *Solid-State Circuits, IEEE Journal of*, 42(1):123–133, 2007.

[65] S. Heber and T. Pock. Shape from light field meets robust pca. In *Computer Vision–ECCV 2014*, pages 751–767. Springer, 2014.

[66] S. Heber, R. Ranftl, and T. Pock. Variational shape from light field. In *EMMCVPR*. Springer, 2013.

[67] G. Higgins, S. Faul, R. P. McEvoy, B. McGinley, M. Glavin, W. P. Marnane, and E. Jones. Eeg compression using jpeg2000: How much loss is too much? In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 614–617. IEEE, 2010.

[68] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar. Video from a Single Coded Exposure Photograph using a Learned Over-Complete Dictionary. In *Proc. IEEE ICCV*, 2011.

[69] M. Hosseini Kamal, M. Golbabaee, and P. Vandergheynst. Light Field Compressive Sensing in Camera Arrays. In *Proc. ICASSP*, pages 5413 –5416, 2012.

[70] M. Hosseini Kamal, M. Shoaran, Y. Leblebici, A. Schmid, and P. Vandergheynst. Compressive multichannel cortical signal recording. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 4305–4309. Ieee, 2013.

[71] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. *The Journal of Machine Learning Research*, 12:3371–3412, 2011.

[72] A. Humayun, O. Mac Aodha, and G. J. Brostow. Learning to find occlusion regions. In *CVPR*. IEEE, 2011.

[73] F. Ives. Parallax stereogram and process of making same., 1903. US Patent 725,567.

[74] R. Jenatton, J.-Y. Audibert, and F. Bach. Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research*, 12:2777–2824, 2011.

[75] R. Jenatton, A. Gramfort, V. Michel, G. Obozinski, F. Bach, and B. Thirion. Multi-scale mining of fmri data with hierarchical structured sparsity. In *Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on*, pages 69–72. IEEE, 2011.

[76] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for hierarchical sparse coding. *The Journal of Machine Learning Research*, 12:2297–2334, 2011.

[77] H. Ji, C. Liu, Z. Shen, and Y. Xu. Robust video denoising using low rank matrix completion. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1791–1798. IEEE, 2010.

[78] H. Ji, S. Huang, Z. Shen, and Y. Xu. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM J Im Sciences*, 4(4):1122–1142, 2011.

[79] J. Jirsch, E. Urrestarazu, P. LeVan, A. Olivier, F. Dubeau, and J. Gotman. High-frequency oscillations during human focal seizures. *Brain*, 129(6):1593–1608, 2006.

[80] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[81] A. M. Kamboh, K. G. Oweiss, and A. J. Mason. Resource constrained vlsi architecture for implantable neural data compression systems. In *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, pages 1481–1484. IEEE, 2009.

[82] S. B. Kang and R. Szeliski. Extracting view-dependent depth maps from a collection of images. *IJCV*, 58(2):139–163, 2004.

[83] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *CVPR*. IEEE, 2001.

[84] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. LeCun. Learning invariant features through topographic filter maps. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1605–1612. IEEE, 2009.

[85] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *Information Theory, IEEE Transactions on*, 56(6):2980–2998, 2010.

[86] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 32(4):73:1–73:12, 2013.

[87] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[88] M. Kowalski and B. Torrésani. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, image and video processing*, 3(3):251–264, 2009.

[89] M. Kowalski, K. Siedenburg, and M. Dorfler. Social sparsity! neighborhood systems enrich structured shrinkage operators. *Signal Processing, IEEE Transactions on*, 61(10):2498–2511, 2013.

[90] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *IJCV*, 2000.

[91] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud. Theory and implementation of an analog-to-information converter using random demodulation. In *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pages 1959–1962. IEEE, 2007.

[92] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl*, 21:1253–1278, 2000.

[93] J. Lawrence, A. Ben-Artzi, C. DeCoro, W. Matusik, H. Pfister, R. Ramamoorthi, and S. Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. *ACM Trans. Graph. (SIGGRAPH)*, 25(3), 2006.

[94] K. Lee and Y. Bresler. ADMiRA: Atomic Decomposition for Minimum Rank Approximation. *IEEE Transactions on Information Theory*, 56(9):4402–4416, 2010.

[95] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM Transactions on Graphics (TOG)*, volume 26, page 70. ACM, 2007.

[96] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1964–1971. IEEE, 2009.

[97] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. SIGGRAPH*, pages 31–42, 1996.

[98] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM Transactions on Graphics (TOG)*, volume 27, page 55. ACM, 2008.

[99] G. Lippmann. La Photographie Intégrale. *Academie des Sciences*, 146:446–451, 1908.

[100] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman. Sift flow: Dense correspondence across different scenes. In *ECCV*. Springer, 2008.

[101] G. Liu, Z. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *ICML*, 2010.

[102] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *Pattern Analy. and Machine Intell., IEEE Tran. on*, 35(1):208–220, 2013.

[103] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *Proc. ICCP*, pages 1–8, 2009.

[104] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

[105] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst. Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes. *Biomedical Engineering, IEEE Transactions on*, 58(9):2456–2466, 2011.

[106] R. F. Marcia and R. M. Willett. Compressive coded aperture video reconstruction. In *Proc. European Signal Processing Conf.(EUSIPCO)*, 2008.

[107] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph. (SIGGRAPH)*, 32(4):46, 2013.

[108] K. Mitra and A. Veeraraghavan. Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior. In *Proc. IEEE CCD*, pages 22–28, 2012.

[109] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *American Control Conference (ACC), 2010*, pages 2953–2959. IEEE, 2010.

[110] M. Mollazadeh, K. Murari, G. Cauwenberghs, and N. V. Thakor. Wireless micropower instrumentation for multimodal acquisition of electrical and chemical neural activity. *Biomedical Circuits and Systems, IEEE Transactions on*, 3(6):388–397, 2009.

[111] J. J. Moreau. Fonctions Convexes Duales et Points Proximaux dans un Espace Hilbertien. *Reports of the Paris. Academy of Sciences*, 255:2897–2899, 1962.

[112] C. Mu, B. Huang, J. Wright, and D. Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. *arXiv preprint*, 2013.

[113] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S. K. Nayar. Programmable aperture camera using lcos. In *Computer Vision–ECCV 2010*, pages 337–350. Springer, 2010.

[114] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3): 301–321, 2009.

[115] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.

[116] G. Obozinski, L. Jacob, and J.-P. Vert. Group lasso with overlaps: the latent group lasso approach. *arXiv preprint arXiv:1110.0413*, 2011.

[117] G. Obozinski, M. J. Wainwright, M. I. Jordan, et al. Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*, 39(1):1–47, 2011.

[118] R. H. Olsson and K. D. Wise. A three-dimensional neural recording microsystem with implantable data compression circuitry. *Solid-State Circuits, IEEE Journal of*, 40(12):2796–2804, 2005.

[119] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *arXiv preprint*, 2012.

[120] J. Y. Park and M. B. Wakin. A multiscale framework for compressive sensing of video. In *Proc. PCS*, pages 1–4, 2009.

[121] P. Peers, K. vom Berge, W. Matusik, R. Ramamoorthi, J. Lawrence, S. Rusinkiewicz, and P. Dutré. A compact factored representation of heterogeneous subsurface scattering. *ACM Trans. Graph. (SIGGRAPH)*, 25(3):746–753, 2006.

[122] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2233–2246, 2012.

[123] C. Perwass and L. Wietzke. Single Lens 3D-Camera with Extended Depth-of-Field. In *Proc. SPIE 8291*, pages 29–36, 2012.

[124] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *ICCV*, pages 1762–1769. IEEE, 2011.

[125] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *ECCV*. Springer, 2008.

[126] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM J. on Imag. Sciences*, 2010.

[127] N. S. Rao, R. D. Nowak, S. J. Wright, and N. G. Kingsbury. Convex approaches to model wavelet sparsity patterns. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1917–1920. IEEE, 2011.

[128] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[129] H. Rauhut, R. Schneider, and Z. Stojanac. Low rank tensor recovery via iterative hard thresholding. In *Proc. 10th Inter. Conf. on Sampling Theory and App.*, 2013.

[130] B. Recht. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 12:3413–3430, 2011.

[131] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[132] A. Rodriguez-Perez, J. Ruiz-Amaya, M. Delgado-Restituto, and A. Rodriguez-Vazquez. A low-power programmable neural spike detection channel with embedded calibration and data compression. *Biomedical Circuits and Systems, IEEE Transactions on*, 6(2):87–100, 2012.

[133] V. Roth and B. Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the 25th international conference on Machine learning*, pages 848–855. ACM, 2008.

[134] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[135] A. C. Sankaranarayanan, P. K. Turaga, R. G. Baraniuk, and R. Chellappa. Compressive acquisition of dynamic scenes. In *Proc. ECCV*, pages 129–142. 2010.

[136] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk. Cs-muvi: Video compressive sensing for spatial-multiplexing cameras. In *Proc. ICCP*, pages 1–10, 2012.

[137] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron*, 52(1):205–220, 2006.

[138] P. Sen and S. Darabi. Compressive rendering: A rendering application of compressed sensing. *IEEE TVCG*, 17(4):487–499, 2011.

[139] P. Sen, S. Darabi, and L. Xiao. Compressive rendering of multidimensional scenes. In *Video Processing and Computational Video*, pages 152–183. Springer, 2011.

[140] P. M. Shankar, W. C. Hasenplaugh, R. L. Morrison, R. A. Stack, and M. A. Neifeld. Multiaperture imaging. *Appl. Opt.*, 45(13):2871–2883, 2006.

[141] M. Shoaran, C. Pollo, Y. Leblebici, and A. Schmid. Design techniques and analysis of high-resolution neural recording systems targeting epilepsy focus localization. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5150–5153. Ieee, 2012.

[142] M. Shoaran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid. Compact low-power cortical recording architecture for compressive multichannel data acquisition. 2014.

[143] B. M. Smith, L. Zhang, H. Jin, and A. Agarwala. Light field video stabilization. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 341–348. IEEE, 2009.

[144] N. Srebro and R. Salakhutdinov. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In *Advances in Neural Information Processing Systems*, pages 2056–2064, 2010.

[145] M. Stead, M. Bower, B. H. Brinkmann, K. Lee, W. R. Marsh, F. B. Meyer, B. Litt, J. Van Gompel, and G. A. Worrell. Microseizures and the spatiotemporal scales of human partial epilepsy. *Brain*, page awq190, 2010.

[146] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *Signal Processing, IEEE Transactions on*, 57(8):3075–3085, 2009.

[147] X. Sun, X. Mei, M. Zhou, H. Wang, et al. Stereo matching with reliable disparity propagation. In *3DIMPVT*. IEEE, 2011.

[148] R. Szeliski and D. Scharstein. Symmetric sub-pixel stereo matching. In *ECCV*. Springer, 2002.

[149] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (tombo): Concept and experimental verification. *Appl. Opt.*, 40(11):1806–1813, 2001.

[150] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[151] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[152] A. N. Tikhonov and V. Y. Arsenin. Solutions of ill-posed problems. 1977.

[153] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50(10):2231–2242, 2004.

[154] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.*, 2001.

[155] B. A. Turlach, W. N. Venables, and S. J. Wright. Simultaneous variable selection. *Technometrics*, 47(3):349–363, 2005.

[156] J. J. Van Gompel, S. M. Stead, C. Giannini, F. B. Meyer, W. R. Marsh, T. Fountain, E. So, A. Cohen-Gadol, K. H. Lee, and G. A. Worrell. Phase i trial: safety and feasibility of intracranial electroencephalography using hybrid subdural electrodes containing macro-and microelectrode arrays. *Neurosurgical focus*, 25(3): E23, 2008.

[157] M. A. O. Vasilescu and D. Terzopoulos. TensorTextures: Multilinear image-based rendering. *ACM Trans. Graph. (SIGGRAPH)*, 23:336–342, 2004.

[158] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):69, 2007.

[159] A. Veeraraghavan, D. Reddy, and R. Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE TPAMI*, 33(4):671 –686, 2011.

[160] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (SIGGRAPH Asia)*, 32(6):166, 2013.

[161] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. G. Baraniuk. Compressive imaging for video representation and coding. In *Picture Coding Symposium*, volume 1, 2006.

[162] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4d light fields. In *CVPR*. IEEE, 2012.

[163] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. PAMI*, 2013.

[164] S. Wanner, S. Meister, and B. Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *Vision, Modeling & Visualization*, pages 225–226. The Eurographics Association, 2013.

[165] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich. Computational plenoptic imaging. In *Computer Graphics Forum*, volume 30, pages 2397–2426. Wiley Online Library, 2011.

[166] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH)*, 31:1–11, 2012.

[167] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):765–776, 2005.

[168] G. A. Worrell, A. B. Gardner, S. M. Stead, S. Hu, S. Goerss, G. J. Cascino, F. B. Meyer, R. Marsh, and B. Litt. High-frequency oscillations in human temporal lobe: simultaneous microwire and clinical macroelectrode recordings. *Brain*, 131 (4):928–937, 2008.

[169] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2):210–227, 2009.

[170] J. Wright, A. Ganesh, K. Min, and Y. Ma. Compressive principal component pursuit. *Information and Inference*, 2(1):32–68, 2013.

[171] Z. Xu and E. Y. Lam. A high-resolution lightfield camera with dual-mask design. In *SPIE Optical Engineering+Applications*, pages 85000U–85000U, 2012.

[172] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

[173] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV$-\ell_1$ range image integration. In *ICCV*, pages 1–8. IEEE, 2007.

[174] C. Zhou and S. K. Nayar. Computational cameras: convergence of optics and processing. *Image Processing, IEEE Transactions on*, 20(12):3322–3340, 2011.

[175] C. Zhou, O. Cossairt, and S. Nayar. Depth from diffusion. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1110–1117. IEEE, 2010.

[176] Y. Zhou, R. Jin, and S. Hoi. Exclusive lasso for multi-task feature selection. In *International Conference on Artificial Intelligence and Statistics*, pages 988–995, 2010.

[177] R. Ziegler, S. Bucheli, L. Ahrenberg, M. Magnor, and M. Gross. A bidirectional light field-hologram transform. In *Computer Graphics Forum*, volume 26, pages 435–446. Wiley Online Library, 2007.

[178] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In *ACM Transactions on Graphics (TOG)*, volume 23, pages 600–608. ACM, 2004.

Mahdad Hosseini Kamal

Website: http://lts2www.epfl.ch/people/mahdad
E-mail: hosseinikamal.m@gmail.com
Phone: (+41)78-888-7553
Nationality: Iran

Education

Oct. 2010-Mar. 2015 — **PhD in Electrical Engineering**, Signal Processing Laboratory (LTS2), EPFL, Switzerland
Dissertation: Structure Modeling of High Dimensional Data: New Algorithms and Applications

Sept. 2007-Apr. 2010 — **M.Sc. in Communication Systems**, "Specialization in Wireless Communications", EPFL, Switzerland
Thesis: Estimating and Learning the Trajectory of Mobile Phones

Sept. 2003-jun. 2007 — **B.Sc. in Electrical Engineering**, "Specialization in Telecommunications", K. N. Toosi University of Technology, Iran
Thesis: A Power Efficient Routing Protocol for Wireless Sensor Networks

Professional Experience

Jun. 2010-Mar. 2015 — **Research Assistant**, Signal Processing Lab (LTS2), EPFL, Switzerland

Oct. 2013-Jan. 2014 — **Visiting Scholar**, Camera Culture Group, Media Lab, MIT, USA

Jun. 2013-present — **Gamaya**, VP of Algorithms & Camera Design, Switzerland

Sept. 2010-Dec. 2014 — **Teaching Assistant**, various bachelor and master level courses, EPFL, Switzerland

Feb. 2009-Apr. 2010 — **Research Intern**, Nokia Research Center, Switzerland

Jul.-Sept. 2006 — **R&D Intern**, Micromodje Industries, Iran

Patents

Mar. 2015

- Compact Low-power Recording Architecture for Multichannel Acquisition of Biological Signals and Methods for Compressing Said Biological Signal data, US Patent Application n° 14/668,313

Competences

Career Competences

- Big data analysis and algorithmic skills for high dimensional data
- Distributed optimization, compressive sensing, image deconvolution
- Signal and image analysis, optimization, dimensionality reduction
- Light field imaging, camera design, disparity estimation
- Hyper-spectral image acquisition
- Clustering and classification of large data sets, feature extraction, pattern discovery and object recognition

Soft Skills/Leadership

- Good communication, organizational, language skills
- Capable of working independently and in a team
- Teamwork, project management, leadership
  - Teaching assistant for 8 undergraduate/graduate courses
  - Supervising Semester and internship projects
  - Member of board of directors of "Iranian Students Association at EPFL (IRSA)" in 2013-2014

Computer Skills

- C/C++, java, Python, R, MATLAB, SQL, MS Office, LaTeX

Language Skills

- **English**: fluent
- **French:** good, level B2 certificate
- **Persian:** excellent, native speaker

Publications

Journals

- A Convex Solution to Disparity Estimation from Light Fields via the Preconditioned Primal-Dual Method, **M. Hosseini Kamal**, P. Favaro, P. Vandergheynst, in Preparation, SIAM J. of Imaging Science
- Tensor Low-rank and Sparse Light Field Photography, **M. Hosseini Kamal**, B. Heshmat , R. Raskar, P. Vandergheynst, G. Wetzstein, submitted to Journal of Computer Vision and Image Understanding, 2015
- Compact Low-Power Cortical Recording Architecture for Compressive Multichannel Data Acquisition, M. Shoaran, **M. Hosseini Kamal**, C. Pollo, P. Vandergheynst, A. Schmid, IEEE TBioCas, 2014
- Compressive Image Acquisition in Modern CMOS IC Design, N. Katic, **M. Hosseini Kamal**, A. Schmid and P. Vandergheynst, Y. Leblebici, Intern. J. of Circuit Theory and App., 2013

Conferences

- A Convex Solution to Disparity Estimation from Light Fields via the Primal-Dual Method, **M. Hosseini Kamal**, P. Favaro, P. Vandergheynst, EMMCVPR, 2014
- Computationally Efficient Background Subtraction in the Light Field Domain, A. Ghasemi, **M. Hosseini Kamal**, M. Vetterli, IS&T/SPIE Electronic Imaging, 2014
- Multichannel Blind Deconvolution using Low-rank and Sparse Decomposition, **M. Hosseini Kamal**, P. Vandergheynst, SPARS, 2013
- Joint Low-rank and Sparse Light Field Modeling for Dense Multiview Data Compression, **M. Hosseini Kamal**, P. Vandergheynst, ICASSP, 2013
- Compressive Multichannel Cortical Signal Recording, **M. Hosseini Kamal**, M. Shoaran, Y. Leblebici, A. Schmid, P. Vandergheynst, ICASSP, 2013
- Column-Separated Compressive Sampling Scheme for Low Power CMOS Image Sensors, N. Katic, **M. Hosseini Kamal**, M. Kilic, A. Schmid, P. Vandergheynst, Y. Leblebici, NEWCAS, 2013
- Power-Efficient CMOS Image Acquisition System based on Compressive Sampling, N. Katic, **M. Hosseini Kamal**, M. Kilic, A. Schmid, P. Vandergheynst, Y. Leblebici, MWSCAS, 2013
- High Frame-Rate Low-Power Compressive Sampling CMOS Image Sensor Architecture, N. Katic, **M. Hosseini Kamal**, M. Kilic, A. Schmid, P. Vandergheynst, Y. Leblebici, GLSVLSI, 2013
- Interconnected Network of Cameras, **M. Hosseini Kamal**, H. Afshari, Y. Leblebici, A. Schmid, P. Vandergheynst, IS&T/SPIE Electronic Imaging, 2013
- Light Field Compressive Sensing in Camera Arrays, **M. Hosseini Kamal**, M. Golbabaee, P. Vandergheynst, ICASSP 2012