Nr. 43 31. October 2017



Preprint-Series: Department of Mathematics - Applied Mathematics

A Framework for Compresssive Time-of-Flight S. Antholzer, C. Wolf, M. Sandbichler, M. Dielacher and M. Haltmeier



Technikerstraße 13 - 6020 Innsbruck - Austria Tel.: +43 512 507 53803 Fax: +43 512 507 53898 https://applied-math.uibk.ac.at

A Framework for Compressive Time-of-Flight 3D Sensing

Stephan Antholzer, Christoph Wolf, Michael Sandbichler, Markus Dielacher and Markus Haltmeier

Abstract—Spatially and temporally highly resolved depth information enables numerous applications including human-machine interaction in gaming industry or safety functions in the automotive industry. In this paper we address this issue using Time-of-flight (ToF) 3D cameras which are compact devices providing highly resolved depth information. Practical restrictions often require to reduce the amount of data to be read out and transmitted. Using standard ToF cameras, this can only be achieved by lowering the spatial or temporal resolution. To overcome such a limitation, we propose a compressive ToF camera design that allows to reduce the amount of data while keeping high spatial and temporal resolution. This uses the theory of compressive sensing and sparse recovery. We propose efficient block-wise reconstruction algorithms based on ℓ_1 -minimization. We apply the developed reconstruction methods to data captured by a real ToF camera system and evaluate them in terms of reconstruction quality and computational effort.

Index Terms—Compressed sensing, Time-of-Flight Imaging, 3D Imaging, sparse recovery, image reconstruction.

S. Antholzer, C. Wolf, M. Sandbichler, and M. Haltmeier are with the Department of University of Innsbruck, Technikerstraße 13, 6020 Innsbruck, Austria. E-mail: {stephan.antholzer, michael.sandbichler, markus.haltmeier}@uibk.ac.at

M. Dielacher Infineon Technologies Austria, Babenbergerstraße 10, 8020 Graz, Austria. E-mail: markus.dielacher@infineon.com

I. INTRODUCTION

Time-of-Flight (ToF) camera systems rely on the time of flight (or travel time) of an emitted and reflected light beam to create a depth image of a scenery. They offer many advantages over traditional systems (e.g. lidar) such as compact design, registering depth and intensity images at a high frame rate, and low power consumption [12]. This makes them ideal for mobile usage for example using a ToF camera on a mobile phone. On such devices, the computational resources we can use for the required image reconstruction algorithms are limited. While there are several technologies allowing 3D imaging, in this paper we will focus on cameras that use a modulated light source to calculate the phase shift (encoding the depth image) between the emitted and received signal [16].

High spatial and temporal resolution requires a large amount of data to be read out and transferred from ToF cameras. In order to determine a depth image, at typically four different phase images per frame have to be collected in the ToF camera. However, even from four phase images the depth image is unique only up to a certain maximal distance from the camera. To measure larger distances one needs additional phase images that have to be read out and transferred. Also in multi-camera systems, where the depth image is calculated outside the camera, the

amount of data can be very high. If the data rate is a limiting factor, either the spatial or the temporal resolution has to be reduced in a conventional ToF camera.

To address this issue, in this article we propose a compressive ToF camera that allows a reduced amount of data to be transferred while preserving high spatial and temporal resolution. Instead of individual pixels of the phase images, the compressive ToF camera reads out combinations of pixel values that are transferred to an external processor. We thereby only use combinations of elements in the same row, which is compatible with existing ToF camera designs. Note that a completely different compressive ToF camera design has been proposed in [20], [19], [17]. In order to reconstruct the original phase and depth images we use techniques from sparse recovery and ℓ_1 -minimization.

Outline

In Section II we give a short overview on ToF imaging. In Section III we present the type of measurements that we propose for compressive ToF imaging. We thereby start with details on the classical (noncompressive) and the new (compressive) designs. Additionally, we prove that the used matrices fulfill the RIP-property under suitable conditions. In Section IV we give details on the numerical algorithm and present extensive studies of our two-step reconstruction approach of recovering the depth image from the compressed measurements.

A preliminary version of this paper has been presented at the International Conference Sampling Theory and Applications (SampTA) 2017 in Tallinn [3].



Figure I.1: BASIC PRINCIPLES OF A TOF CAMERA. Phase images are collected by sampling the crosscorrelation of the emitted with the reflected light pulse. From these phase images one computes a depth image mapping the distance from the scene to the camera.

II. BASICS OF 3D IMAGING USING TOF CAMERAS

A ToF camera measures the distance of a scenery to the camera. By sending out a diffuse light pulse and measuring the reflected signal the camera is able to record the depth informations of the entire scenery at once. To acquire depth information the sent out light is modulated and can be generated by an LED. The scenery reflects the light which is recorded by the camera as depicted in Figure I.1. The emitted pulse can be modeled as a time-dependent function $g(t) = C \cos(\omega t)$, where C is the amplitude, ω the modulation frequency (or carrier frequency), and t the time variable. The signal is reflected, and the camera receives, for any individual pixel $i \in \{1, ..., n\}$, a phase and amplitude shifted signal

$$f_i(t) = B_i + A_i \cos(\omega t - \varphi_i).$$

Here φ_i is the phase shift depending on the distance d_i between the camera and the scene mapped at pixel i, A_i the amplitude depending on the reflectivity, and B_i an offset. The phase shift is related to the distance d_i via the relation $d_i = \varphi_i c/(2\omega)$.

At each pixel of the ToF camera, the crosscorrelation between the reference and the reflected signal is measured, where the cross-correlation between two signals $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}$ is given by

$$c_{f,g}(s) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)g(t+s) \,\mathrm{d}t \,.$$
(II.1)

In our case $c_{f,g}(\cdot)$ can be calculated analytically [1], [24], [16] which yields

$$c_{f_{i},g}(s) = \frac{A_i C}{2} \cos(\omega s + \varphi_i) + K_i.$$
(II.2)

Here K_i are some constants accounting for noise and the background image generated by ambient light. By sampling the cross-correlation function at the sampling points $s = 0, \pi/(2\omega), \pi/\omega, 3\pi/(2\omega)$ we get four so called phase images

$$\mathbf{p}^{(1)} = -\frac{\mathbf{A}C}{2}\cos(\boldsymbol{\varphi}) + \mathbf{K}$$
$$\mathbf{p}^{(2)} = -\frac{\mathbf{A}C}{2}\sin(\boldsymbol{\varphi}) + \mathbf{K}$$
$$\mathbf{p}^{(3)} = -\frac{\mathbf{A}C}{2}\cos(\boldsymbol{\varphi}) + \mathbf{K}$$
$$\mathbf{p}^{(4)} = -\frac{\mathbf{A}C}{2}\sin(\boldsymbol{\varphi}) + \mathbf{K}$$

Here we have set $\varphi = (\varphi_j)_{j=1}^n \in \mathbb{R}^n$ and $\mathbf{K} = (K_j)_{j=1}^n$, $\mathbf{A} = (A_j)_{j=1}^n \in \mathbb{R}^n$, and all operations are taken point-wise. Under the common assumption that K_i is independent of the pixel location we can estimate the phase shifts φ by

$$\hat{\boldsymbol{\varphi}} = \arg\left(\mathbf{p}^{(1)} - \mathbf{p}^{(3)} + i\left(\mathbf{p}^{(4)} - \mathbf{p}^{(2)}\right)\right). \quad (\text{II.3})$$

Here $\alpha = \arg(z) \in [0, 2\pi)$ denotes the argument of the complex number z defined by $z \triangleq re^{i\alpha}$. In particular, the depth image is given by $\hat{\mathbf{d}} = \hat{\boldsymbol{\varphi}} c/(2\omega)$.

Since the phase shifts are contained in $[0, 2\pi)$, the maximal distance that can be found by (II.3) is $d_{\text{max}} = (c/\omega)\pi$. Larger distances are falsely identified, taking values in the interval $[0, d_{\text{max}})$. To overcome this ambiguity, several methods have been proposed in the literature (see, for example, [16], [11]). One such approach consists in capturing two sets of phase images with different modulation frequencies $\omega_1 \neq \omega_2$, and then comparing the two depth images. In this paper we will not address the ambiguity problem further since the compressive ToF camera that we propose below can be extended to multiple modulation frequencies in a straightforward manner. Further research, however, is needed to thoroughly investigate the possibility of using machine learning to solve the ambiguity issue.

III. Compressive ToF sensing and image reconstruction

In this section we present the proposed compressive ToF 3D sensing design compatible with existing ToF cameras. Additionally, we describe an efficient block-wise reconstruction procedure based on sparse recovery.

A. Compressive ToF sensing

As mentioned in the introduction, in a conventional ToF camera, all pixel values of all phase images have to be read out and large amounts of data have to be transferred. To reduce the amount of data, in this paper we propose a compressive ToF camera, which reads out and transmits linear combinations instead of individual pixel values of the phase image. Our proposed compressive ToF camera design is based on the existing non compressive ToF camera design, which should allow to engineer and build the new camera with low effort. The only difference between the two designs is in the way the pixels of the sensors are read out. For the compressive ToF camera, we propose to read out linear combinations of neighbouring pixels.

The data collected by the compressive ToF camera can be written in the form

$$\forall i \in \{1, 2, 3, 4\}: \quad \mathbf{y}^{(i)} = \mathbf{M}\mathbf{p}^{(i)}$$

3

Here $\mathbf{M} \in \mathbb{R}^{m \times n}$ is the measurement matrix, $\mathbf{p}^{(i)} \in \mathbb{R}^n$ are the phase images and $\mathbf{y}^{(i)} \in \mathbb{R}^m$ the read out data with $m \ll n$. To reconstruct the depth image from the compressed readouts we propose the following two-step procedure: First, we estimate the differences $\mathbf{p}^{(1)} - \mathbf{p}^{(3)}$ and $\mathbf{p}^{(4)} - \mathbf{p}^{(2)}$ from

$$\mathbf{y}^{(1)} - \mathbf{y}^{(3)} = \mathbf{M} \left(\mathbf{p}^{(1)} - \mathbf{p}^{(3)} \right),$$
 (III.1)

$$\mathbf{y}^{(4)} - \mathbf{y}^{(2)} = \mathbf{M} \left(\mathbf{p}^{(4)} - \mathbf{p}^{(2)} \right),$$
 (III.2)

using sparse recovery. In a second step we recover the depth image by applying (II.3) to the estimated differences.

Any of the equations (III.1), (III.2) is an underdetermined system of the form $\mathbf{y} = \mathbf{M}\mathbf{x}$, for which in general no unique solution exists. To obtain solution uniqueness, the vector $\mathbf{x} \in \mathbb{R}^n$ needs to satisfy certain additional requirements. In the recent years sparsity turned out to be a powerful property for this purpose. Recall that \mathbf{x} is called *s*-sparse, if it has at most *s* nonzero entries. Assuming sparsity, the vector \mathbf{x} is recovered by solving the ℓ_1 -minimization problem

$$\underset{\mathbf{z} \in \mathbb{R}^n}{\text{minimize}} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{M}\mathbf{z} = \mathbf{y} \,. \qquad \text{(III.3)}$$

In order for (III.3) to uniquely recover \mathbf{x} , the matrix \mathbf{M} needs to fulfill certain properties. One sufficient condition is the restricted isometry property (RIP). The matrix \mathbf{M} is said to satisfy the *s*-RIP with constant $\delta > 0$, if

$$(1-\delta) \|\mathbf{z}\|_{2}^{2} \leq \|\mathbf{M}\mathbf{z}\|_{2}^{2} \leq (1+\delta) \|\mathbf{z}\|_{2}^{2}$$

holds for all s-sparse $\mathbf{z} \in \mathbb{R}^n$. If the s-RIP constant is sufficiently small, then (III.3) uniquely recovers any sufficiently sparse vector (see, for example, [10], [13], [7]).

Although some results for deterministic RIP matrices exist [18], [6], matrices satisfying the RIP



Figure III.1: STANDARD READOUT VERSUS COM-PRESSIVE READOUT. Left: Standard sequential readout using 16 ADCs. Right: In the compressive readout, in each row, instead of 16 sequential readouts per ADC, m/K combinations of pixel values are read out.

are commonly constructed in some random manner. Realizations of Gaussian or Bernoulli random matrices are known to satisfy the RIP with high probability [13]. Rademacher random variables take the values -1 and 1 with equal probability. It has been shown [8] that if $m \ge C\delta^{-2}s\log(n/s)$, then $\delta_s \le \delta$ with high probability for both types of matrices. Remarkably, this bound on m is optimal [13]. This means we can not expect to have universal recovery guarantees for a smaller number of measurements. Thus, if the number of measurements is lower than this bound, we can find a *s*-sparse vector that can not be recovered.

More generally, all matrices with independent subgaussian entries satisfy the RIP with high probability. A subgaussian random variable X is defined by the following property

$$\mathbb{P}(|X| \ge t) \le \beta e^{-\kappa t^2} \quad \text{for all } t > 0 \qquad (\text{III.4})$$

with constants $\beta, \kappa > 0$. It is easy to show that Rademacher and Gaussian random variables are subgaussian. Subgaussian matrices **M** are also universal [13] which means that for any unitary matrix

 $\Psi \in \mathbb{C}^{n \times n}$ the matrix $\mathbf{M}\Psi$ also satisfies the RIP. Thus one can also recover signals that are not sparse in the standard basis but for which $\Psi^* \mathbf{x}$ is sparse, where Ψ^* is the conjugate transpose. This property is very useful in applications since many natural signals have sparse representations in certain bases different from the standard basis. In general, if the restricted isometry constant of $\mathbf{M}\Psi$ is small then we can recover the signal by solving (III.3) with $\mathbf{M}\Psi$ instead of \mathbf{M} , in the noisy case,

$$\underset{\mathbf{z}\in\mathbb{C}^{n}}{\text{minimize}} \|\boldsymbol{\Psi}^{*}\mathbf{z}\|_{1} \quad \text{such that } \|\mathbf{M}\mathbf{z}-\mathbf{y}\|_{2} \leq \eta \,.$$
(III.5)

Here η equals to the noise level of the measurements. Similar results hold if Ψ is a frame (see [9], [15], [23], [28]).

In practical applications such unstructured matrices can not always be used. Either there are restrictions on the matrix preventing us from using a random matrix with i.i.d. entries or the storage space is limited, such that storing a full matrix would be too expensive. There are different methods for constructing structured compressed sensing matrices that satisfy the RIP. For example, such matrices can be constructed by random subsampling of an orthonormal matrix [13, Chapter 12] or deterministic convolution followed by random subsampling [25] or using a random convolution followed by deterministic subsampling [27]. In the next section we will examine the latter type since its application to ToF imaging and the existing camera designs is more accessible.

B. Compressive 3D Sensing Using Block Partial Circulant Matrices

The hardware requirements in our case prevent us from using arbitrary matrices since for the analogto-digital converters (ADC) the weights 0 and $\pm c$ for some fixed constant $c \in \mathbb{R}$ should be used. Further, any individual ADC can only be wired with a limited number of pixels (compare Fig. III.1) which imposes a particular block-structure of the measurement matrix. Thus the measurement matrices that we use in our approach take the block-diagonal form

$$\mathbf{M} = \mathbf{M}_1 \otimes \cdots \otimes \mathbf{M}_K := \begin{pmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & & \mathbf{M}_K \end{pmatrix}.$$
(III.6)

Here, each sub-matrix $\mathbf{M}_k \in \mathbb{R}^{m_k \times n_k}$ operates on a certain subset $\Omega_k \subsetneq \{1, \ldots, n\}$ with $n_k = |\Omega_k|$ elements coming from a single row in the image. For simplicity we consider the case that $n_k = n/K$ and $m_k = m/K$ for each k. The particular measurements in each row block are constructed in a certain random manner satisfying the requirements above. In the following subsection we consider partial circulant matrices for possible block entries.

A particularly useful class of row-wise measurements in the compressive ToF camera can be modeled by partial circulant matrices. A circulant matrix $\mathbf{C}_{\mathbf{v}} \in \mathbb{R}^{n \times n}$ associated with $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ is defined by

$$\forall i, j \in \{1, \dots, n\}: \quad (\mathbf{C}_{\mathbf{v}})_{i,j} = v_{i \ominus j}$$

where $j \ominus i := (j - i) \mod n$ is the cyclic subtraction. In particular, for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $\mathbf{C}_{\mathbf{v}}\mathbf{w} = \mathbf{v} * \mathbf{w}$, where $(\mathbf{v} * \mathbf{w})_j := \sum_{i=1}^n v_{j \ominus i} w_i$ is the circular convolution. For any subset $\Omega \subseteq \{1, \ldots, n\}$, the projection matrix $\mathbf{R}_{\Omega} \in \mathbb{R}^{|\Omega| \times n}$ is defined by $\mathbf{R}_{\Omega}\mathbf{v} := (\mathbf{v}_i)_{i \in \Omega}$.

Definiton III.1. The partial circulant matrix associated to $\mathbf{v} \in \mathbb{R}^n$ and $\Omega \subseteq \{1, \ldots, n\}$ is defined by $\frac{1}{\sqrt{|\Omega|}} \mathbf{R}_{\Omega} \mathbf{C}_{\mathbf{v}}.$

Further, recall that a random vector \mathbf{v} with values in $\{\pm 1\}^n$ is called a Rademacher vector if it has independent entries taking the values ± 1 with equal probability. Partial circulant matrices satisfy the RIP. Such results have been obtained first in [27] and have later been refined in [22] using the theory of suprema of chaos processes. These results have formulated for sparsity in the standard basis. For our purpose we formulate such a result for the general orthonormal bases.

Theorem III.2 ([22, Theorem 1.1]). Consider the partial circulant matrix $\mathbf{M} = \frac{1}{\sqrt{m}} \mathbf{R}_{\Omega} \mathbf{C}_{\mathbf{v}}$ associated to a vector \mathbf{v} with iid subgaussian entries and a subset $\Omega \subseteq \{1, \ldots, n\}$ containing m elements. If, for some $s \leq n$ and $\delta \in (0, 1)$, we have

$$m \ge C\delta^{-2}\mu^2 s(\log s)^2 (\log n)^2,$$

then, with probability at least $1 - n^{-\log n \log^2 s}$, the s-RIP constant of $\mathbf{M}\Psi$ is at most δ , where the constant μ is given by $\mu = \max_{i,j} |\langle \mathbf{F}_{j-}, \Psi_{i-} \rangle|$. Here \mathbf{F} is the discrete Fourier matrix and $\Psi \in \mathbb{C}^{n \times n}$ is any unitary matrix.

Proof. For the case that Ψ is the identity matrix the result is derived in the original paper [22]. The generalization to arbitrary Ψ can be shown analogously to the original results. Such a proof is worked out in [2], [4].

Theorem III.2 shows that random partial circulant matrices yield stable recovery of sparse vectors using (III.3). Recall that the proposed compressive ToF camera readout uses block diagonal measurement matrices of the form (III.6). Taking each block as a random partial circulant matrix and applying Theorem III.2 yields the following result. **Theorem III.3.** Let $\mathbf{M} \in \mathbb{R}^{m \times n}$ be of the form (III.6), where each block on the diagonal is a partial circulant matrix $\mathbf{M}_k = \frac{1}{\sqrt{m_k}} \mathbf{R}_{\Omega_k} \mathbf{C}_{\mathbf{v}_k}$ associated with independent Rademacher vectors \mathbf{v}_k and subsets $\Omega_k \subseteq \{1, \ldots, n/K\}$ having $m_k = m/K$ elements that are selected independently and uniformly at random. If, for some $s \in \mathbb{N}$ and $\delta \in (0, 1)$, we have

$$m \ge KC\delta^{-2}\mu^2 s(\log s)^2 (\log(n/K))^2,$$

then, with probability at least $(1 - (n/K)^{-\log(n/K)\log^2 s})^K$, $\mathbf{M}_k \Psi$ has the s-RIP constant of at most δ for all $k = 1, \ldots, K$. Here the constant μ is given by $\mu = \max_{i,j} |\langle \mathbf{F}_{j-}, \Psi_{i-} \rangle|$ and $\Psi \in \mathbb{C}^{(n/K) \times (n/K)}$ is any unitary matrix.

Proof. As $m/K \ge C\delta^{-2}s(\log s)^2(\log(n/K))^2$, we can apply Theorem III.2 with n and m replaced by n/K and m/K to each block. Thus the restricted isometry constant of each block is at most δ with probability at least $1 - (n/K)^{-\log(n/K)\log^2 s}$. As the generating Rademacher vectors for each block are independent, the *s*-RIP constants of all blocks are uniformly bounded by δ with probability at least $(1 - (n/K)^{-\log(n/K)\log^2 s})^K$.

Theorem (III.3) yields stable recovery via (III.5) if the vector \mathbf{x} is $\boldsymbol{\Psi}$ -block sparse, meaning that $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$ with $\boldsymbol{\Psi}^* \mathbf{x}_k \in \mathbb{R}^{n/K}$ being *s* sparse for all $k = 1, \ldots, K$.

C. Image reconstruction by block- ℓ_1 minimization

As presented in Section III-A, the depth image is recovered from compressed readouts by first estimating the differences $\mathbf{p}^{(1)} - \mathbf{p}^{(3)}$ and $\mathbf{p}^{(4)} - \mathbf{p}^{(2)}$ from (III.1) and (III.2), which are underdetermined systems of equations of the form $\mathbf{y} = \mathbf{M}\mathbf{x}$, and then applying (II.3) to the estimated differences. In this subsection we present how to efficiently solve these underdetermined systems using block-wise ℓ_1 - \mathbf{B}_{n/b^2} with diagonal blocks minimization.

Suppose that the measurement matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ has a block diagonal form (III.6), with diagonal blocks $\mathbf{M}_k \in \mathbb{R}^{(m/K) \times (n/K)}$ operating on a subgroup of pixels from individual lines. This type of measurement matrices reflects the current ToF camera architecture illustrated in Fig. III.1. Assuming the sparsifying basis Ψ to be block diagonal with diagonal blocks $\Psi_k \in \mathbb{R}^{(n/K) \times (n/K)}$, the full ℓ_1 minimization problem (III.5) can be decomposed into K smaller ℓ_1 -minimization problems of the form

$$\underset{\mathbf{z}\in\mathbb{R}^{n/K}}{\text{minimize}} \|\boldsymbol{\Psi}_{k}^{*}\mathbf{z}\|_{1} \quad \text{subject to } \|\mathbf{M}_{k}\mathbf{z}_{k}-\mathbf{y}_{k}\|_{2} \leq \eta/K.$$
(III.7)

Here $\mathbf{y}_k \in \mathbb{R}^{m/K}$ are the data from a single block. If all \mathbf{M}_k satisfy the Ψ -RIP (i.e. $\mathbf{M}_k \Psi_k$ satisfies the RIP), then (III.7) stably and robustly recovers any Ψ -block-sparse vector $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_K]$ with $\mathbf{y} = \mathbf{M}\mathbf{x}$. Theorem III.3 shows that this, for example, is the case if \mathbf{M}_k are realized as random partial circulant matrices.

By solving (III.7) we exploit sparsity within a single row-block. While one can expect some rowsparsity, (III.7) does not fully exploit the level of sparsity present in two-dimensional images. As shown in [30] using row-sparsity yields artifacts in the reconstructed image. In this work we therefore follow a different approach that is described next. For that purpose we consider an additional partition of all pixels

$$\{1,\ldots,n\} = \bigcup_{\ell=1,\ldots,n/b^2} B_\ell, \qquad \text{(III.8)}$$

where B_{ℓ} corresponds to all indices in squared blocks of size $b \times b$ with b := n/K. Then the measurement matrix can be written in the form $\mathbf{M} = \mathbf{B}_1 \otimes \cdots \otimes$

$$\mathbf{B}_{\ell} := \begin{pmatrix} \mathbf{M}_{\ell} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{\ell+b} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{M}_{\ell+(b-1)b} \end{pmatrix} \in \mathbb{R}^{b^2 \times b^2}$$
(III.9)

for $\ell = 1, \ldots, n/b^2$. We further assume that the sparsifying basis $\Psi = \Psi_1 \otimes \cdots \otimes \Psi_{n/b^2}$ is block diagonal with $\Psi_k \in \mathbb{R}^{b^2 \times b^2}$. In such a situation, (III.3) can be decomposed into n/b^2 smaller ℓ_1 minimization problems,

$$\underset{\mathbf{z}_{\ell} \in \mathbb{R}^{b^2}}{\text{minimize}} \| \boldsymbol{\Psi}_{\ell}^* \mathbf{z}_{\ell} \|_{1} \quad \text{subject to } \| \mathbf{B}_{\ell} \mathbf{z}_{\ell} - \mathbf{y}_{\ell} \|_{2} \leq \varepsilon .$$
(III.10)

The advantage of (III.10) over (III.7) is that Ψ_{ℓ} can now be chosen as a two dimensional wavelet or cosine transform, which are well known to provide sparse representation of images. On the other hand, (III.5) is still decomposed into smaller subproblems which enables efficient numerical implementations. The optimization problems (III.10) can be solved in parallel which further decreases computation times. Using a global sparsifying transformation might be better in terms of sparsity, but the resulting problem is less efficient to solve. In the future we will investigate optimal compromises between sparsity and computational efficiency.

For the actual numerical implementation we use ℓ_1 -Tikhonov regularization

$$\underset{\mathbf{z}_{\ell} \in \mathbb{R}^{b^2}}{\text{minimize } \lambda \|\mathbf{z}_{\ell}\|_1 + \|\tilde{\mathbf{B}}_{\ell} \mathbf{z}_{\ell} - \mathbf{y}_{\ell}\|_2^2 }$$
(III.11)

where $\tilde{\mathbf{B}}_{\ell} = \mathbf{B}_{\ell} \Psi_l$ can be calculated by $(\Psi_{\ell}^{\mathsf{T}} \mathbf{B}_{\ell}^{\mathsf{T}})^{\mathsf{T}}$. The two problems (III.10), (III.11) are equivalent [14] in the sense that every solution of (III.10) is also a solution of (III.11) for λ depending on ε and vice versa. For minimizing the unconstrained ℓ_1 -problem (III.11) we use the fast iterative soft thresholding

7

algorithm (FISTA) introduced in [5] which is an efficient algorithm for ℓ_1 -minimization.

IV. EXPERIMENTAL RESULTS

In this section we present some experimental results using raw data captured by an existing standard ToF camera. An example of such raw data (four phase images) is shown in Figure IV.1. From the raw data of the standard ToF camera, we generated the compressive sensing measurements synthetically.

A. Compressed ToF sensing

For compressive ToF sensing, we initialized the measurement matrices M (the block circulant matrices; see Section III-B) randomly with the entries of a random vector generating the partial circulant blocks taking values in $\{-1, 1, 0\}$ with equal probability. The blocks have size $m \times 14$ which implies that the compression ratio is 14/m. In the experiments we observed that usually not all blocks of our measurement matrix yield adequate reconstruction properties. This indicates that the size of the single blocks is not large enough to guarantee recovery in each block with high probability. Using bigger blocks would overcome this issue (according to Theorem III.3), but this is not possible with our camera design. We therefore propose the following alternative strategy. We start with a set of several candidates for the blocks of the measurement matrix from which we choose the ones with the lowest reconstruction error on a set of test images.

For the following results we have chosen the parameters in the FISTA algorithm by hand and did not perform extensive parameter optimization. On most images the parameter choice had a moderate influence on the reconstruction error. Thus we used $\lambda = 0.05$ for all presented results with ℓ_1 minimization. For the basis Ψ we use the 2D-Haar wavelet transform and, as described in Section III-C, we executed the reconstruction block wise with a block size of 28×28 .

To measure the error between the uncompressed depth image $\mathbf{d} \in \mathbb{R}^{l \times n} = \mathbb{R}^{168 \times 224}$ and the reconstructed depth image $\tilde{\mathbf{d}} \in \mathbb{R}^{168 \times 224}$ we use the relative mean absolute error (RMAE) and the peak signal to noise ratio (PSNR) defined by

MAE(
$$\mathbf{d}, \tilde{\mathbf{d}}$$
) := $\frac{1}{ln} \sum_{i=1}^{l} \sum_{j=1}^{n} |d_{i,j} - \tilde{d}_{i,j}|$

$$\text{RMAE}(\mathbf{d}, \tilde{\mathbf{d}}) \coloneqq \frac{\text{MAE}(\mathbf{d}, \tilde{\mathbf{d}})}{\max_{i,j}(d_{i,j})} 100\%$$

$$\text{PSNR}(\mathbf{d}, \tilde{\mathbf{d}}) \coloneqq 10 \log \left(ln \frac{\max_{i,j}(d_{i,j})}{\sum_{i,j} (d_{i,j} - \tilde{d}_{i,j})^2} \right) \text{dB} \,.$$



Figure IV.1: PHASE IMAGES as captured by the existing ToF camera (chair-image). The vertical line artefacts, as well as the noise are inherent to the measurement procedure.

B. Example 1: Chair image

For our first set of examples we consider phase images of a chair shown in Figure IV.1, which is less than 1 meter away from the camera (see Figure IV.2, left). In the FISTA reconstruction of the chair-image

we see some artefacts around the main objects. They are a result from the small blocks in the measurement matrix and varying sparsity level of the images. The wavelet coefficients are not exactly zero but only rapidly decreasing and on blocks around edges the decay is slower than elsewhere. These artefacts at image contours could be reduced by adding a noise reduction term to (III.11), for example a total variation term [29], [31], [21], which results in an improved image quality. However, the reconstruction speed would be slower.

In Figure IV.3 we can see the reconstructed differences of the phase images. In the uncompressed differences of the phase images shown in Figure IV.3 (bottom), one clearly sees how the noise \mathbf{K} (see (II.2)) contained in the original phase images (shown in Figure IV.1) is cancelled out. This demonstrates that it is indeed beneficial to capture four phase images instead of two. The phase images shown in Figure IV.1 contain significantly more noise and artefact than the differences shown in Figure IV.3.

C. Example 2: Books image

In the second example set we consider an image from a couple of books and folders (see Figure IV.4, left). The objects are around 0.5 to 1.2 meters away from the camera and the background consists of a wall, which is around 1.5 meters away. In the booksimage, the FISTA yields less artefact compared to the chair image since the image consists of larger piecewise constant regions. As can be seen in Figure IV.5, this results in faster decaying wavelet coefficients.

D. Very limited data

To investigate the reconstruction quality when only using a very small amount of data, we generated a measurement matrix with m = 3. This results in a



Figure IV.2: TOP. Depth image calculated from the phase images of Figure IV.1. On the top left we see wrapping artefact, because the objects exceed the maximal measurable distance. BOTTOM. Reconstructed depth image after 10 iterations of FISTA. The colors give the distance from the camera in meters. The compression ratio is 50% and the relative mean absolute error 3.53%.

compression ratio of around 21%. In this example we also increased the probability for zeros to 2/3 and the resulting matrix had around 57% zeros. This means that the images can be captured very quickly since zero entries in the measurement matrix imply that the camera can skip the corresponding pixel. However this results in some missing data which can never be recovered. The reconstructed depth image is shown in Figure IV.6, where one notices clear vertical artefacts. The vertical structure of the artifacts is



Figure IV.3: TOP. Reconstructed differences of phase images $\mathbf{p}^{(1)} - \mathbf{p}^{(3)}$ and $\mathbf{p}^{(4)} - \mathbf{p}^{(2)}$ with a compression ratio of 50%. BOTTOM. Uncompressed differences. One clearly sees that the uncompressed phase differences contain less artefacts than the original phase images $\mathbf{p}^{(i)}$ itself (shown in Figure IV.1).

because we us the same block in the measurement matrix for measurements in each column. As a consequence, the resulting artefacts are more regular than for the case where we use different blocks for each row.

E. Error analysis

In Table I we show the average mean absolute error evaluated on a set of 26 test images for various compression ratios. The images consist of the chair-image and books-image and other similar images captured with the ToF camera in an office and an apartment. We used the same FISTA parameter $\lambda = 0.05$ and ran 10 iterations for all the samples.

V. CONCLUSION

In this paper, we proposed a compressive ToF camera design that reduces the required amount of data to be read out and transferred. The proposed compressed ToF camera uses measurements within



Figure IV.4: TOP. Depth image of our second example (the books-image) calculated from the uncompressed phase images. BOTTOM. Reconstructed depth image after 10 iterations of FISTA with the same settings as in Figure IV.2 (the compression ratio is again 50%). The relative mean absolute error is 0.62%.

Measurements m	Average RMAE	Average PSNR
12	0.95	68.17
7	1.38	65.84
5	1.95	58.19
3	6.19	36.57

Table I: Average errors and signal to noise ratio on a set of test images of the depth images.

10



Figure IV.5: ORDERED WAVELET COEFFICIENTS of the blocks from the differences of the phase images (left: $\mathbf{p}^{(1)} - \mathbf{p}^{(3)}$ and right: $\mathbf{p}^{(4)} - \mathbf{p}^{(2)}$) from our two example images. We only plotted the largest 10% of the coefficients. As expected, the coefficients for the chair-image decrease slower.



Figure IV.6: DEPTH IMAGE recovered from very limited data. The compression ratio is around 21%. Additionally, more than half of the entries of the measurement matrix equals zero. The relative mean absolute error is 8.98%.



Figure IV.7: TOP. Average relative MAE error of the reconstructed depth images depending on the compression ratio m/n. We used the same procedure and data as in Table I. BOTTOM. Same as on top, but now using the PSNR for evaluating the reconstruction quality.

rows of the image which yields a block-diagonal measurement matrix. Random partial circulant matrices as diagonal blocks have been shown to be compatible with current camera architecture. However, their asymptotic recovery guarantees do not directly apply in our case. To fix this, one can increase the block size, which is not really practical for the ToF camera. On the other hand our experimental results still clearly demonstrate that it is possible to recover the original images for small block sizes. For that purpose, we proposed and implemented a strategy to increase the compressed sensing ability of the random partial circulant matrices. For image reconstruction we use different squared blocks that allow to exploit sparsity of the phase images in the two dimensional wavelet basis. Future work has to be done to improve the image quality and make it more consistent across different images and to increase the reconstruction speed. For that purpose we will investigate the use of machine learning based algorithms for compressed sensing [26].

Acknowledgments

Stephan Antholzer and Markus Haltmeier acknowledge support of the Austrian Science Fund (FWF), project P30747. The work of Michael Sandbichler is supported by the FWF, Grant no. Y760.

References

- M. Albrecht. Untersuchung von Photogate-PMD-Sensoren hinsichtlich qualifizierender Charakterisierungsparameter und -methoden. PhD thesis, University of Siegen, 2007.
- [2] S. Antholzer. Nonlinear compressive time-of-flight 3D imaging. Master's thesis, University of Innsbruck, 2017.
- S. Antholzer, C. Wolf, M. Sandbichler, M. Dielacher, and M. Haltmeier. Compressive time-of-flight imaging. In Sampling Theory and Applications (SampTA), 2017 International Conference on, pages 556–560. IEEE, 2017.
- [4] D. Aschenbrücker. Sparse recovery with random convolutions. Master's thesis, University of Bonn, 2015.
- [5] A. Beck and M. Teboulle. A fast iterative shrinkagethresholding algorithm for linear inverse problems. SIAM J. Imaging Sci., 2(1):183–202, 2009.
- [6] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit constructions of RIP matrices and related problems. *Duke Math. J.*, 159(1):145–185, 2011.
- [7] T. Cai and A. Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory*, 60:122 – 132, 2014.
- [8] E. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406– 5425, 2006.

- [9] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. Appl. Comput. Harmon. Anal., 31(1):59–73, 2011.
- [10] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
- [11] D. Droeschel, D. Holz, and S. Behnke. Multifrequency phase unwrapping for time-of-flight cameras. In 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1463–1469, 2010.
- [12] S. Foix, G. Alenya, and C. Torras. Lock-in time-offlight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.
- [13] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.
- [14] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of *ℓ*¹regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.
- [15] M. Haltmeier. Stable signal reconstruction via l¹minimization in redundant, non-tight frames. *IEEE Trans. Signal Process.*, 61(2):420–426, 2013.
- [16] M. E. Hansard, S. Lee, O. Choi, and R. Horaud. Time-of-Flight Cameras – Principles, Methods and Applications. Springer Briefs in Computer Science. Springer, 2013.
- G. A. Howland, P. Zerom, R. W. Boyd, and J. C. Howell. Compressive sensing lidar for 3D imaging. In *CLEO: 2011 Laser Science to Photonic Applications*, pages 1–2, May 2011.
- [18] M. A. Iwen. Simple deterministically constructible rip matrices with sublinear Fourier sampling requirements. In 2009 43rd Annual Conference on Information Sciences and Systems, pages 870–875, 2009.
- [19] A. Kadambi and P. T. Boufounos. Coded aperture compressive 3-d lidar. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1166–1170, April 2015.
- [20] A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal. Codac: A compressive depth acquisition camera framework. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5425–5428, 2012.
- [21] F. Krahmer, C. Kruschel, and M. Sandbichler. Total variation minimization in compressed sensing. arXiv preprint arXiv:1704.92195, 2017.
- [22] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67(11):1877–1904, 2014.

- [23] F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. SIAM J. Math. Anal., 47(6):4606–4629, 2015.
- [24] R. Lange. 3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology. PhD thesis, University of Siegen, 2000.
- [25] K. Li, L. Gan, and C. Ling. Convolutional compressed sensing using deterministic sequences. *IEEE Transactions* on Signal Processing, 61(3):740–752, 2013.
- [26] A. Mousavi and R. G. Baraniuk. Learning to invert: Signal recovery via deep convolutional networks. arXiv preprint arXiv:1701.03891, 2017.
- [27] H. Rauhut, J. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Applied* and Computational Harmonic Analysis, 32(2):242–254, 2012.
- [28] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inf. Theory*, 54(5):2210 –2219, 2008.
- [29] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. Variational methods in imaging, volume 167 of Applied Mathematical Sciences. Springer, New York, 2009.
- [30] C. Wolf. Framework for compressed sensing in time-offlight based 3D imaging. Master's thesis, University of Innsbruck, 2016.
- [31] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.