

A Novel Approach for Solving MPI for Multi-Target ToF Imaging using Subdivision-based Nested Compressed Sensing

Sanhita Guha[†], Faisal Ahmed[‡], and Miguel Heredia Conde[‡]

[†]Fraunhofer Institute for High Frequency Physics and Radar Techniques (FHR), Wachtberg, Germany

[‡]Center for Sensorsystems (ZESS), University of Siegen, Paul-Bonatz-Straße 9-11, 57076 Siegen, Germany

Emails: sanhita.guha@fhr.fraunhofer.de • faisal.ahmed@uni-siegen.de • heredia@zess.uni-siegen.de

Abstract—Time-of-Flight (ToF) cameras have become a popular imaging modality for macroscopic scene detection. In particular, amplitude-modulated continuous wave (AMCW) ToF cameras use the phase difference between sent and received signals for object depth reconstruction. However, various sources of multipath reflections exist in practice, causing each ToF pixel to erroneously receive a superposition of multiple reflections instead of a single bounce. This leads to distortions in the phase difference and consequently, errors in the depth maps. Compressed Sensing methods have emerged as an effective approach to solve this multipath interference (MPI) problem. However, it has two major disadvantages—large sensing matrix size leading to high computational load, and a high mutual coherence causing reconstruction failure. This paper introduces a subdivision-based nested compressed sensing algorithm that aims to alleviate these known disadvantages. Measurements at multiple modulation frequencies are used to isolate the k interfering signals in the time domain. Simulation results are presented with a noise performance analysis and results based on real multi-target measurement data are also discussed.

Index Terms—time-of-flight, multi-path interference, depth resolution improvement, compressed sensing

I. INTRODUCTION

There is an increasing demand for high-resolution remote sensing systems across different parts of the wavelength spectrum. The Time-of-Flight (ToF) camera is one such remote sensing system for macroscopic scene detection that has gained popularity in the recent years and is being used in a wide range of applications such as vehicle monitoring and obstacle detection in the automotive industry, safe patient monitoring for healthcare, 3D scanning, and human-machine interactions [1], [2], [3]. Specifically, the amplitude-modulated continuous wave (AMCW) ToF cameras have had a wide commercial impact due to their small size and high signal-to-noise (SNR) ratio [4].

In AMCW ToF cameras, the intensity of the emitted sinusoidal signal is modulated at one or several frequencies of the order of 10-100 MHz. In an ideal case, this signal is reflected by a sparse number of objects in the scene, causing specific time-shifts or phase-shifts in the signal, corresponding to the

object depths. The pixels of the ToF camera capture these shifts in phase, which can then be scaled by the modulation frequency to obtain the object depths.

ToF systems work on the assumption that each pixel receives a signal reflected by a single object. However, in practice, the signal received by each pixel corresponds to the superposition of signals from multiple scatterers, leading to multipath interference (MPI). As a result, the detected phase shift is distorted, which ultimately leads to an inaccurate depth estimation. Therefore, in ToF systems, depth resolution improvement equates to better multipath interference mitigation.

A flurry of research work has appeared on different ways of tackling this MPI problem. Some of these research directions include sensor modification [5], [6], spectral estimation [7], [8] and deep-learning based approaches [9]. Under the umbrella of spectral estimation, research on compressed sensing (CS) based optimization has shown promising results [10], [11], [12]. However, the main challenges have been the computational load due to the size of the sensing matrix, and its high coherence, leading to a failure in the l_1 norm minimization.

The aim of this paper is to tackle the MPI problem using a new subdivision-based nested CS approach. It aims to use multiple submatrices constructed from a main structured sensing matrix. This serves to reduce the coherence as well as the computational load of the optimizations. The depth-regions of interest can then be identified based on the variance of the sub-estimates. Re-applying CS on the combination of these interest areas can provide better estimates with a much lower computational cost. The performance of the algorithm is tested on a simulated multi-target scene, as well as on real ToF data.

The paper is structured as follows. Section II discusses the ToF operation and generation of the measurements at different modulation frequencies. Section III describes the CS formulation in detail. Section IV discusses the CS specific challenges encountered in ToF MPI mitigation and introduces the subdivision-based CS algorithm to overcome these problems. Section V provides results based on simulations as well as real data, demonstrating the effectiveness of the proposed method. Section VI discusses the conclusions drawn and the scope for future work.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860370.

II. PROBLEM STATEMENT

In this work, the focus is placed on a multi-frequency AMCW ToF system. In this section, the mathematical formulation of the frequency-domain sensing model is described.

The scene response function in the phase domain is given by

$$e(\phi) = \sum_{k=1}^P a_k \delta(\phi - \phi_k), \quad (1)$$

where $\delta(\phi - \phi_k)$ is a Dirac delta function centered at ϕ_k , which is the phase shift that the transmitted signal undergoes when following the path $k \in [1, P]$. P represents the number of paths which is assumed to be low or 'sparse'. a_k is the attenuation factor due to non-unit reflectance of target. The received signal at each pixel can then be expressed as a convolution of the transmitted periodic illumination signal $s(\phi)$, and the scene response $e(\phi)$, i.e.,

$$r(\phi) = s(\phi) \star e(\phi). \quad (2)$$

As described in [11], the final measurement at each pixel is obtained by a cross-correlation of $r(\phi)$ with the PMD control signal p_{A-B} , yielding

$$m(\phi) = p_{A-B} \otimes r(\phi) \quad (3)$$

Substituting (2) in (3),

$$\begin{aligned} m(\phi) &= p_{A-B} \otimes (s(\phi) \star e(\phi)) \\ &= \sum_{k=1}^P a_k (p_{A-B} \otimes s(\phi)) (\phi - \phi_k). \end{aligned} \quad (4)$$

The per-path amplitude and the phase terms in (4) are the only two parameters of interest and can be determined by $2P + 1$ measurements [11]. If these two measurements have a phase shift of $\pi/2$, they can be used to construct the Fourier coefficients of the scene response function as follows:

$$\begin{aligned} y(j) &= m_j(0) + i m_j(\pi/2) \\ &= \sum_{k=1}^P a_k (\cos(-j\phi_k) - i \sin(-j\phi_k)) \\ &= \sum_{k=1}^P a_k e^{ij\phi_k}, \end{aligned} \quad (5)$$

where j represents the index specifying the frequency in the multi-frequency acquisition mode. Considering f_0 as the base operating frequency, the data acquisition is carried out for different frequencies given by $f_j = j f_0, j = 1, \dots, M$. Therefore (5) can be simplified to give

$$\begin{aligned} y(j) &= \sum_{k=1}^P a_k e^{ij(2\pi f_0 t_k)} \\ &= \sum_{k=1}^P a_k e^{i2\pi f_j t_k}. \end{aligned} \quad (6)$$

Thus, (6) provides a simplified expression of the measurements at every pixel in terms of discrete time and frequency

samples. The aim is to retrieve the terms a_k and $t_k, \forall k \leq P$, that can correctly explain y . The matrix pencil method, a robust variant of Prony's method, is a standard way to achieve this goal [16]. Compressed Sensing (CS) based approaches may improve the depth resolution by promoting sparse solutions in a finer grid. However, existing CS algorithms pose certain challenges that are discussed later.

III. COMPRESSED SENSING APPROACH

In practice, the number of objects of interest that reflect the illumination signal are sparse in comparison to the ambient dimension in time/depth domain. Therefore, it is logical to follow a CS-based approach for ToF depth resolution improvement.

A general CS problem is expressed as

$$\vec{y} = \mathbf{A}\vec{x} + \vec{e},$$

where $\vec{y} \in \mathbb{C}^J$ containing the measurements in frequency domain, $\mathbf{A} \in \mathbb{C}^{J \times K}, J \ll K$, is the sensing matrix, and $\vec{x} \in \mathbb{C}^K$ is an s -sparse reflectivity vector constructed from the sparse reflectivities of the objects in the scene. \vec{e} is white Gaussian noise in the measurement \vec{y} . (6) can be expanded into this general CS formulation, giving

$$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(J) \end{bmatrix}}_{\vec{y}} = \underbrace{\begin{bmatrix} e^{j2\pi(f_1, t_1)} & \dots & e^{j2\pi(f_1, t_K)} \\ e^{j2\pi(f_2, t_1)} & \dots & e^{j2\pi(f_2, t_K)} \\ \vdots & \ddots & \vdots \\ e^{j2\pi(f_M, t_1)} & \dots & e^{j2\pi(f_M, t_K)} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(K) \end{bmatrix}}_{\vec{x}} + \vec{e}. \quad (7)$$

The sensing matrix \mathbf{A} resembles a horizontal discrete Fourier transform (DFT) matrix, oversampled in time. Such a sensing matrix construction aligns with the structured CS approach [13], and is different from the traditionally used random sensing matrices. The sparse reflectivity vector \vec{x} consists of the real-valued reflectivities $a(k)$ from each object in the scene. Based on this framework, the solution for \vec{x} is given by

$$\min \|\vec{x}\|_1 \quad s.t. \quad \|\mathbf{A}\vec{x} - \vec{y}\|_2 \leq \eta, \quad (8)$$

where η is the upper bound on the norm of the measurement error, i.e., $\|\vec{e}\|_2 \leq \eta$. For the ToF application, if the location of the non-zero reflectivities can be determined accurately, the corresponding columns of \mathbf{A} can be identified as the active columns, thereby identifying the correct 'times of flight', t_k . These time instances then directly provide the depth information for each object, since $d_k = \frac{ct_k}{2}$. In order to get an improved depth resolution using this CS formulation, the ToF camera must acquire a sufficient number of measurements at different modulation frequencies. Theoretically, given a sparsity s and a finely-resolved depth grid of length N , the relation [14]

$$m \geq 2s \ln(N)$$

gives the number of multi-frequency measurements required.

IV. SUBDIVISION BASED ITERATIVE SOFT THRESHOLDING ALGORITHM

The use of popular CS algorithms for the depth resolution improvement in ToF systems poses certain challenges. In the CS formulation described in (9), a depth resolution improvement is possible only when CS algorithms support a finely-spaced time (or depth) grid. However, this is challenging due to the following reasons:

- 1) Given a constant depth to be covered, a finely-spaced time grid increases the size of the sensing matrix, leading to a large computational load.
- 2) With a finely-spaced time grid, the degree of similarity between adjacent columns, which is measured by their normalized inner product, increases, and, as a result, the coherence of the sensing matrix increases, leading to an ill-posed CS problem. In such a scenario, conventional CS methods fail to provide a good estimate.

In order to solve these issues, a sub-division-based nested CS algorithm is introduced. A different algorithm based on this idea was applied for radar range resolution improvement in [15].

A. Subdivision-based Nested CS

It must be noted that in order to get (real-valued) reflectivities using complex measurement vector \vec{y} , a symmetric version of the DFT matrix is used as the sensing matrix \mathbf{A} , i.e., the time grid is always centered around 0. Based on this sensing matrix structure, the subdivision-based nested CS algorithm is outlined in Algorithm 1 and is described as follows:

- 1) Determine the factor $k_{\text{sub}} = \frac{K}{J}$ that dictates the difference between the number of rows and columns of \mathbf{A} .
- 2) Divide \mathbf{A} into k sensing matrices \mathbf{A}_{sub} that cover different parts of the time or depth grid, while maintaining a symmetric structure. Thus,

$$\mathbf{A}_{\text{sub}}(i) = \begin{bmatrix} \mathbf{A}(:, 1:k_{\text{sub}}:\text{mid}-i) \\ \mathbf{A}(:, \text{mid}) \\ \mathbf{A}(:, \text{mid}+i:k_{\text{sub}}:J) \end{bmatrix}^T, \quad (9)$$

where, mid represents the column index for time 0, and $i = 1, \dots, k_{\text{sub}}$. Due to this subdivision, each $\mathbf{A}_{\text{sub}}(i)$ has a lower coherence and a k_{sub} -times smaller column dimension compared to \mathbf{A} , thereby addressing both the challenges previously described.

- 3) The measurement vector \vec{y} may now be represented as $\vec{y} = \mathbf{A}_{\text{sub}}(i)\vec{x}_{\text{sub}}(i)$, $i = 1, \dots, k_{\text{sub}}$. These k sub-problems can be directly solved using a greedy CS method or basis-pursuit minimization.
- 4) The columns of each sub-matrix $\mathbf{A}_{\text{sub}}(i)$ capture a different part of the depth grid. Since the detected scene is sparse, only a few of the $\vec{x}_{\text{sub}}(i)$ s will have sharp peaks representing the on-grid targets for the specific sub-matrices. $k_{\text{sub}}/2$ of the $\vec{x}_{\text{sub}}(i)$ s having the highest standard deviation values are selected, since high standard deviation (or variance) corresponds to the CS results with the sharpest peaks.

- 5) The fine-grid locations corresponding to these peaks are used to construct the final support S .
- 6) The final estimate is given by solving a normal CS problem, as expressed in (8), such that

$$\vec{x}_{\text{final}} = \text{CS}(\mathbf{A}_S, \vec{y}), \quad (10)$$

where \mathbf{A}_S consists of the columns of \mathbf{A} identified by the support set S .

Algorithm 1: Subdivision-based Nested CS

Data: sensing matrix \mathbf{A} , measurement vector \vec{y}

Result: \vec{x}_{final}

```

for  $I = 1$  to  $k_{\text{sub}}$  do
     $\Omega^{(I)} = \begin{bmatrix} 1:k_{\text{sub}}:\text{mid}-I \\ \text{mid} \\ \text{mid}+I:k_{\text{sub}}:J \end{bmatrix}^T$ ;
     $\vec{x}_{\text{sub}}^{(I)} = \text{CS}(\mathbf{A}_{\text{sub}}^{(I)}, \vec{y})$ ;
     $\sigma^{(I)} = \text{std}(\vec{x}_{\text{sub}}^{(I)})$ ;
end
 $[\sigma, \text{idx}] = \max_k(\sigma, k_{\text{sub}}/2)$ ;
for  $I = 1$  to  $k_{\text{sub}}/2$  do
     $\text{loc}_{\text{sub}} = \text{findpeaks}(\vec{x}_{\text{sub}}(\text{idx}^{(I)}))$ ;
     $\text{loc}^{(I)} = \Omega^{(\text{idx}^{(I)})}(\text{loc}_{\text{sub}})$ ;
     $S = \cup_I \text{loc}^{(I)}$ ;
end
 $\vec{x}_{\text{final}} = \text{CS}(\mathbf{A}_S, \vec{y})$ ;

```

V. RESULTS

A. Simulation Results

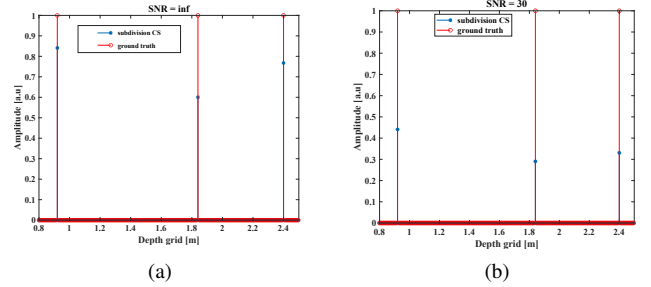


Fig. 1: Synthetic Scene with SNR= ∞ and SNR=30.

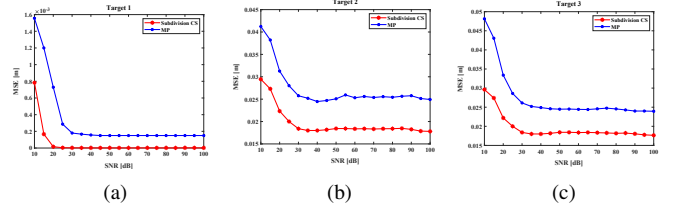


Fig. 2: SNR vs MSE plots for the 3 synthetic targets.

The performance of the algorithm was tested on synthetic measurements from 3 point targets at depths of 0.92, 1.84 and 2.4 m respectively. Following the parameters of the real measurement setup, 53 frequency samples were considered ranging

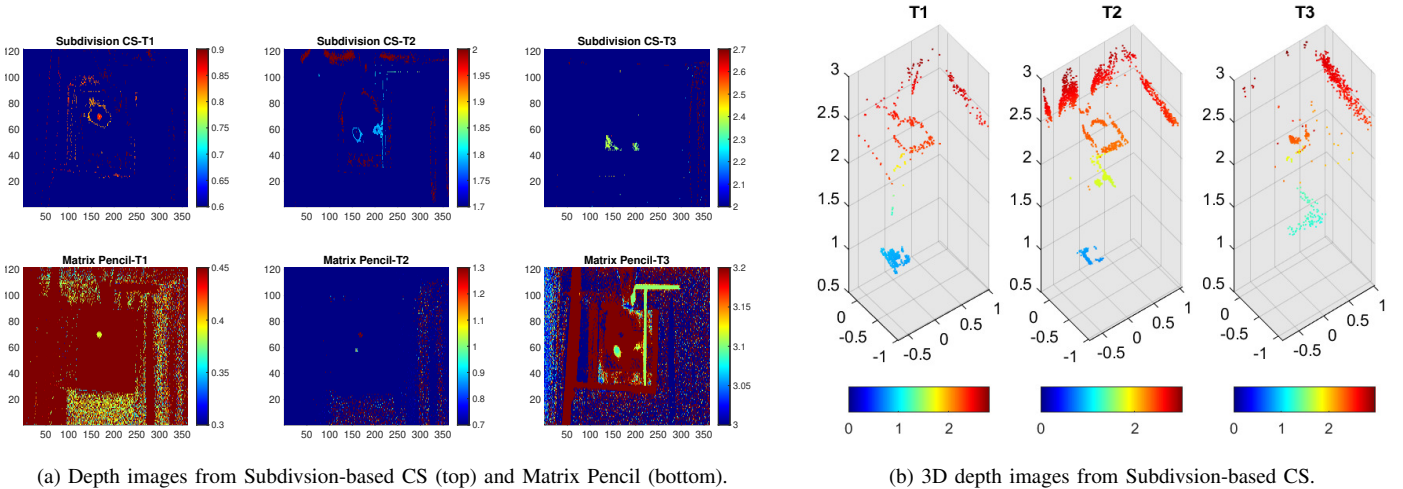


Fig. 3: Depth Image results from Subdivision-based CS and Matrix Pencil methods.

from 0 to 179.214 MHz, with steps of 3.514 MHz. Since a centered measurement vector is considered, as described in Section IV-A, the frequency grid ranges from -179.214 MHz to 179.214 MHz. The measurements corresponding to the negative frequencies are complex conjugates of the ones at the positive frequencies. Therefore, the centered measurement vector \vec{y} consists of 103 measurements. For all simulations, a depth of 3m was considered with a grid resolution of 8mm.

Fig. 1 shows the 3 peaks on the depth grid and the reconstruction results from subdivision-based nested CS for $\text{SNR} = \infty$ and $\text{SNR} = 30$. As expected, the amplitudes of the detected peaks reduce for the noisy case, however, the positions are detected accurately. Using the Iterative Soft-Thresholding Algorithm (ISTA) as the CS reconstruction method, every instance of the subdivision algorithm has a computation time of 0.004 sec, while the standard ISTA requires 0.13 sec. This improvement greatly affects the computation time, specially when the a large number of ToF pixels are considered.

B. Performance Analysis

The matrix-pencil method [16] has been used as the basis for performance comparison. It was observed that the matrix-pencil method performed better for lower values of the pencil parameter L , $\frac{J}{2} \leq L \leq J$. This agrees with the conditions for a noise-robust pencil method in [16]. Fig. 2 shows the MSE vs SNR plots for the depth images of the 3 targets obtained using the subdivision-based nested CS method (in red) and the matrix-pencil method (in blue), for an average of 5 noisy measurements. The error reduces with increase in SNR for both methods and becomes constant at around 20 dB. For all 3 cases, the results from the CS method has a lower error as compared to the MPI method. This is expected, since the synthetic scene is quite sparse. The small irregularities in the low SNR region arise due to a small number of measurements used for the average.

TABLE I: Target position estimates from the Subdivision-based CS method and the matrix pencil method.

Methods	Targets		
	T1 (m)	T2 (m)	T3 (m)
Ground Truth	0.87	1.87	2.408
Matrix Pencil	0.38	1	3.1-3.18
Subdivision-based CS	0.86	1.937	2.35-2.4

To analyze the performance of the proposed algorithm for different levels of sparsity and different numbers of measurements in the presence of noise, phase transition diagrams are constructed. In Fig 4, $\frac{m}{K}$ denotes the ratio of the number of available measurements to the number of range cells, and $\frac{s}{m}$ denotes the ratio of sparsity to the number of measurements. For each plot, the ratios $\frac{m}{K}$ and $\frac{s}{m}$ are varied from 0.05 to 1 in steps of 0.05 and for every point $(\frac{m}{K}, \frac{s}{m})$, 20 iterations of the proposed method are executed. The success or failure of each iteration is determined by calculating the averaged mean-squared error between the ground truth and the reconstruction results. An error of 0.001 or less is considered a success. The averaged rate of success is then shown in the corresponding position in the phase transition diagram. Since the sensing matrix is always horizontal ($K \gg m$), the number of columns (K) considered for the phase transition diagrams must correspond to the maximum number of available measurements (m). Figs show the phase transition plots for the nested CS method for $\text{SNR} = \infty, 80, 40, \text{ and } 20$ dB.

C. Results on Multi-Target Measurement Data

This section discusses the results obtained using real ToF data. The experimental setup, shown in Fig. 5, consists of 3 targets—a glossy translucent target (T1) at 0.87 m, a diffusive translucent target (T2) at 1.87 m, and an opaque white placard target (T3) at 2.408 m depth.

Fig. 3(a) shows the depth images obtained using the subdivision-based CS (top) and the matrix pencil method (bottom) on real measurements. Fig. 3(b) presents the 3D

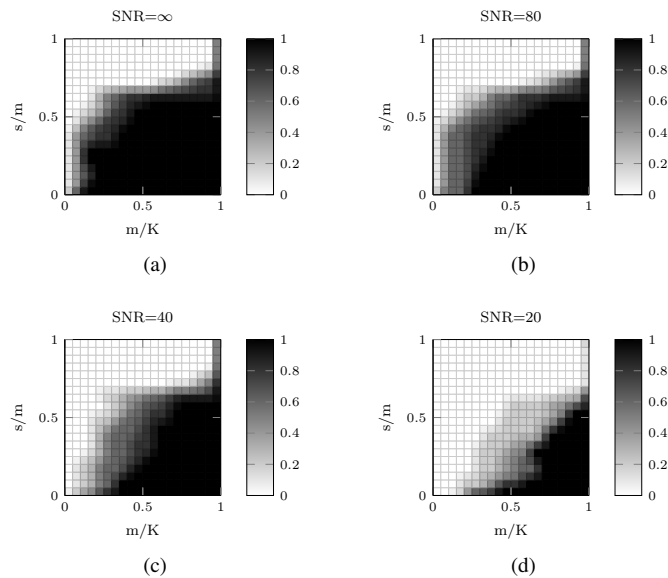


Fig. 4: Phase transition diagrams. Black corresponds to 100% empirical success probability. White corresponds to 0% success probability.

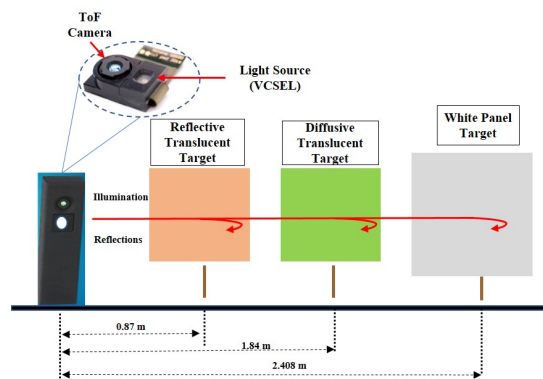


Fig. 5: Experimental Setup

plots of the depth images obtained using subdivision-based CS. Table I shows the estimated depths of each target from both methods. Since the position in depth is decided by the location of the amplitude peaks, the sparsity-inducing methods perform better as they are capable of identifying prominent data points in noisy environments. The matrix pencil method shows much higher sensitivity to noise and fails with real data. Specially in the case of T3, the same amplitude was detected for a large number of inaccurate points in the depth grid, leading to an incorrect estimate. Overall, the estimation accuracy of the CS based method is much better than that of the matrix pencil method adopted in prior works to solve the MPI problem.

VI. CONCLUSION

A new subdivision-based nested CS algorithm was proposed with an aim to improve on the disadvantages of existing CS methods applied to ToF MPI mitigation. Simulation results

were presented along with a performance analysis. A 3-target setup was used to test the performance of the algorithm for both simulated and real-world measurements. The proposed algorithm is flexible and can be further explored using different types of CS methods. A more detailed comparative study based on different CS methods can be performed in future. The effect of the pencil parameter on the performance of the matrix pencil method can also be further explored.

REFERENCES

- [1] Hsu, S., Acharya, S., Rafii, A., New, R. (2006). Performance of a Time-of-Flight Range Camera for Intelligent Vehicle Safety Applications. In: Valldorf, J., Gessner, W. (eds) *Advanced Microsystems for Automotive Applications 2006*. VDI-Buch, Springer, Berlin, Heidelberg.
- [2] U. B. Himmelsbach, T. M. Wendt and M. Lai, "Towards Safe Speed and Separation Monitoring in Human-Robot Collaboration with 3D-Time-of-Flight Cameras," 2018 Second IEEE International Conference on Robotic Computing (IRC), Laguna Hills, CA, USA, 2018, pp. 197-200, doi: 10.1109/IRC.2018.00042.
- [3] Y. Cui, S. Schuon, D. Chan, S. Thrun and C. Theobalt, "3D shape scanning with a time-of-flight camera," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 1173-1180, doi: 10.1109/CVPR.2010.5540082.
- [4] C. Bamji et al., "A Review of Indirect Time-of-Flight Technologies," in *IEEE Transactions on Electron Devices*, vol. 69, no. 6, pp. 2779-2793, June 2022, doi: 10.1109/TED.2022.3145762.
- [5] Y. Zhao, X. Wang, Y. Zhang, Y. Fang, and B. Su, "Polarization-based approach for multipath interference mitigation in time-of-flight imaging," *Appl. Opt.* 61, 7206-7217 (2022).
- [6] Naik, Nikhil, et al. "A light transport model for mitigating multipath interference in time-of-flight sensors." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
- [7] A. Kirmani, A. Benedetti and P. A. Chou, "SPUMIC: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods," 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 2013, pp. 1-6, doi: 10.1109/ICME.2013.6607553.
- [8] A. Islam, M. A. Hossain and Yeong Min Jang, "Interference mitigation technique for Time-of-Flight (ToF) camera," 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, Austria, 2016, pp. 134-139, doi: 10.1109/ICUFN.2016.7537001.
- [9] S. S. Patil and V. S. Inamdar, "Mitigating Multipath Interference for Time of Flight Range Sensors via Deep Convolutional Networks," 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 2021, pp. 277-281, doi: 10.1109/ICCC54389.2021.9674237.
- [10] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Opt. Lett.* 39, 1705-1708 (2014).
- [11] Conde, Miguel Heredia, et al. "Fast Multipath Estimation for PMD Sensors."
- [12] S.S. Patil, P.M. Bhade, V.S. Inamdar Depth recovery in time of flight range sensors via compressed sensing algorithm. *Int J Intell Robot Appl* 4, 243-251 (2020).
- [13] M. F. Duarte and Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications," in *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053-4085, Sept. 2011, doi: 10.1109/TSP.2011.2161982.
- [14] Foucart, S., Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. ISBN: 9780817649487 0817649484
- [15] S. Guha, A. Bathelt, M. H. Conde and J. Ender, "Radar Band Fusion Using Frame-Based Compressed Sensing," in *IEEE Journal of Selected Topics in Signal Processing*, 2022, doi: 10.1109/JSTSP.2022.3220403.
- [16] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814-824, May 1990, doi: 10.1109/29.56027.
- [17] Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*. 2004 Nov;57(11):1413-57.