Spatio-temporal Super-resolution for CS-based ToF 3D Imaging

Alvaro Lopez Paredes*, Miguel Heredia Conde*, Thoriq Ibrahim†, Anh Ngoc Pham†, Keiichiro Kagawa‡
*Center for Sensor Systems, University of Siegen, Germany
†Graduate School of Integrated Science and Technology, Shizuoka University, Japan
‡Research Institute of Electronics, Shizuoka University, Japan
Corresponding author: alvaro.lparedes@uni-siegen.de

***Abstract*—Time-of-Flight cameras, especially the ones featuring sensors built upon macro-pixel architectures, are generally characterized by lower resolutions than other imaging modalities. In this work, we propose a practical single-frame super-resolution scheme to surpass this limitation by exploiting the architecture of a laboratory-designed macro-pixel ultra-high-speed image sensor. In addition, we propose a two-step sparsity-aware greedy algorithm for the recovery of the signal in the time (depth) domain. The main contribution of our algorithm is the introduction of a preliminary screening step to define a set of feasible support candidates in which we replace the demodulation functions, the assembly of which yields the sensing matrix, by the theoretical binary codes they are built upon. We empirically demonstrate that our algorithm improves the recovery performance with respect to Order Recursive Matching Pursuit for sparsity ≤ 2.**

***Index Terms*—Coded imaging, Compressive Sensing, macro-pixel, super-resolution, Time-of-Flight.**

I. Introduction

Time-of-Flight (ToF) cameras are generally characterized by lower resolutions than other imaging modalities, such as RGB (color) cameras. This gap is even bigger when the depth sensor is built upon a macro-pixel structure, as the one we are considering in this paper [1], [2]. In this paper, we aim to surpass these limitations and propose:

1) A **practical single-frame spatial super-resolution (SR) framework** which permits us to reliably work in the sub-pixel range.
2) An **original Compressive Sensing (CS) oriented algorithm** which allows for the recovery of the depth and intensity of each sub-pixel, reaching temporal (depth) SR, and the reduction of the depth estimation error with respect to Order Recursive Matching Pursuit (ORMP) [3] and prior work [1].

II. Theoretical Background

A ToF camera produces a 3D model of the surrounding environment by exploiting the opacity of most objects or, analogously, the *sparsity* of the signals representing them. In cameras based on indirect-ToF, the sensor-to-target return-trip time, equivalent to the distance between them, is obtained from samples of the correlation of the reflected signal with $m \geq 1$ demodulation functions $\vec{a}_i \in \mathbb{R}^n$ implemented in the ToF camera. In Pulse-based (PB)-ToF [4], [5], the emitted signal consists of a train of short light pulses, and $\vec{a}_i$ is initially designed as a custom (0,1)-binary sequence $\vec{a}_{0,i}$ with $\vec{a}_{0,i} \in \mathbb{R}^{n_0}$ subsequently discretized in n_{steps}, such that $n = n_0 \times n_{\text{steps}}$, and filtered accounting for the real response of the pixel. This filtering operation can be expressed as the convolution of two shift-invariant functions, such as the code and the Instrument Response Function (IRF) [4], [5]. The assembly of $[\vec{a}_i]_{i=1}^m$ yields a sensing matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ [4], and the scalar products of the scene response function $\vec{x}$ with $[\vec{a}_i]_{i=1}^m$ yield the measurement vector $\vec{y} \in \mathbb{R}^m$. As $m \ll n$, we encounter an under-determined system of equations, that translates into a constrained ℓ_0-minimization problem (1) when the inherent *sparsity* (or *compressibility*) of $\vec{x}$ is considered.

$$\hat{\vec{x}} = \arg\min_{\vec{x}} \|\vec{x}\|_0, \quad \text{s.t.} \quad \vec{y} = \boldsymbol{A} \cdot \vec{x} \tag{1}$$

In order to reconstruct $\vec{x}$ from $\vec{y}$, classical greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [6] or ORMP [3], generate a discrete probability distribution as a result of the scalar product of the residual of $\vec{y}$ and the normalized columns of $\boldsymbol{A}$, and, then, estimate the support of the signal upon it. Therefore, the design of $\boldsymbol{A}$ [7]–[10] plays a fundamental role on the reconstruction success. Since every column of $\boldsymbol{A}$ represents a distinct fraction of the time (depth) domain, the optimal $\boldsymbol{A}$ would be the one which attained the lowest inter-column coherence $\mu = \max\left(\frac{|\vec{a}^{i\intercal} \cdot \vec{a}^j|}{\|\vec{a}^i\|_2 \|\vec{a}^j\|_2}\right)$ with $1 \leq i \leq n$, $1 \leq j \leq n$ and $i \neq j$. In this sense, Tropp [6] demonstrated that exact retrieval can be guaranteed if $\mu < \frac{1}{2s-1}$, being s the *sparsity*.

III. Methodology

A. Spatial Super-resolution: From Macro-pixel to Sub-pixel Range using the Sub-pixel Intensity Maps.

The resolution of a depth camera is often augmented on a multi-frame framework making use of a set of images from other imaging modalities whose sensors are characterized by a larger number of pixels [11], on a single-frame framework based on deep learning approaches [12], [13], or considering

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. Development of the image sensor was partially supported by Grant-in-Aid for Scientific Research (S) Numbers 17H06102 and 18H05240. Development of the image sensor was partially supported by Grant-in-Aid for Scientific Research (S) Numbers 17H06102 and 18H05240.

CS techniques together with spatial light modulators (SLMs) [14]–[16] such as digital micro-mirror devices (DMD). For instance, in [14] DMD masks are implemented together with Block Compressive Sensing (BCS) to allow for spatial SR. In addition, they divided the sensor read-out into various blocks to avoid data overload during transmission, and designed the read-out mode to make it compatible with BCS. Differently, we do not consider CS techniques in the spatial domain, which would require additional hardware, but exploit the particular architecture of the macro-pixel, i. e., the arrangement of the sub-pixels within the macro-pixel and the fact that each sub-pixel comprises $m_{\mathrm{SP}} = 4$ taps or integration channels [1]. We increase the resolution up to the sub-pixel range, by performing a demosaicing guided by the high resolution intensity map $\boldsymbol{I}$. Therefore, the increase of the spatial resolution is fixed ($\times 4$) and given by the size of the sub-pixels and the macro-pixel. Moreover, our algorithm does not yield an excessive increase of computational load with execution times $\sim \mathcal{O}\left(10^{-1}\,\frac{\mathrm{ms}}{\mathrm{pix}}\right)$. Thus, our methodology is a valuable tool to increase the spatial resolution of depth sensors built upon a macro-pixel architecture to reach sub-pixel SR.

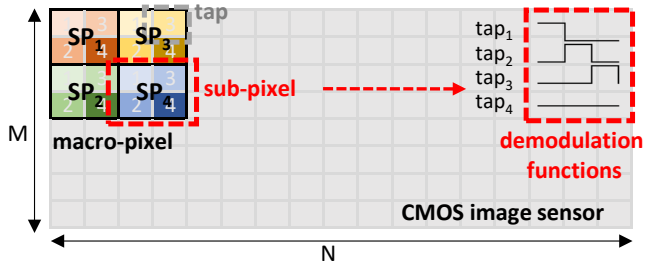

Fig. 1. ToF sensor architecture: each macro-pixel consists of four sub-pixels which, in turn, contain four integration channels or taps.

As shown in Fig. 1, our sensor consists of an array of $M \times N$ macro-pixels, being $M = 94$ and $N = 106$. Each macro-pixel comprises $n_{\mathrm{SP}} = 4$ sub-pixels ($\mathrm{SP_i}$ with $i = 1, 2, 3$, and 4), which, in turn, contain $m_{\mathrm{SP}} = 4$ integration channels or taps. This permits the generation of $m \leq 16$ demodulation functions per macro-pixel. The spatial SR framework (from macro- to sub-pixel resolution) is summarized in Algorithm 1 and consists of two steps:

- **Calculation of the high-resolution intensity map used in the resolution transfer:** We generate one Low-resolution (LR) intensity map $\boldsymbol{I}_{\mathrm{SP_t}}^{\mathrm{LR}}$ for each of the sub-pixels by calculating $\|\vec{y}_{\mathrm{SP_t}}^{\mathrm{LR}^{(j,k)}}\|_2$ with $1 \leq t \leq n_{\mathrm{SP}}$, $1 \leq j \leq M$ and $1 \leq k \leq N$, respectively. Then, we re-map them as per the location of each sub-pixel within the macro-pixel, and combine them $\boldsymbol{I} = \mathcal{M}\left(\boldsymbol{I}_{\mathrm{SP_1}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_2}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_3}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_4}}^{\mathrm{LR}}\right)$, with $\boldsymbol{I} \in \mathbb{R}^{2M \times 2N}$.
- **Generation of m SR raw images via bilateral filtering:** We perform an up-scaling of the raw images $\mathcal{A}\left(\boldsymbol{Y}_i^{\mathrm{LR}}\right)$ with $1 \leq i \leq m$ using bi-cubic interpolation, yielding $\tilde{\boldsymbol{Y}}_i \in \mathbb{R}^{2M \times 2N}$. Then, we apply a bilateral filter over them [17], the weights of which are calculated as $w^{(t_y,t_x)} = e^{-\left(\frac{(t_y-j)^2+(t_x-k)^2}{2\sigma_K^2} - \frac{(I^{(t_y,t_x)}-I^{(j,k)})^2}{2\sigma_I^2}\right)}$, with $(t_y, t_x) \in \Omega^{(j,k)}$, $1 \leq j \leq 2M$, and $1 \leq k \leq 2N$. As a result, $\boldsymbol{Y}_i \in \mathbb{R}^{2M \times 2N}$, with $1 \leq i \leq m$, raw frames are generated.

Algorithm 1: Spatial super-resolution scheme.

Data: $\boldsymbol{Y}_l^{\mathrm{LR}}$ with $1 \leq l \leq m$, $\Omega^{(j,k)}$, σ_K, σ_I
Results: $\boldsymbol{Y}_l$ with $1 \leq l \leq m$
for $j = 1 : M$ **do**
  for $k = 1 : N$ **do**
    for $t = 1 : n_{\mathrm{SP}}$ **do**
      $\hat{i}_{\mathrm{SP_t}}^{\mathrm{LR}^{(j,k)}} = \|\vec{y}_{\mathrm{SP_t}}^{\mathrm{LR}^{(j,k)}}\|_2$
    end
  end
end
$\boldsymbol{I} = \mathcal{M}\left(\boldsymbol{I}_{\mathrm{SP_1}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_2}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_3}}^{\mathrm{LR}}, \boldsymbol{I}_{\mathrm{SP_4}}^{\mathrm{LR}}\right)$
for $l = 1 : m$ **do**
  $\tilde{\boldsymbol{Y}}_l = \mathcal{A}\left(\boldsymbol{Y}_l^{\mathrm{LR}}\right)$
  for $j = 1 : 2M$ **do**
    for $k = 1 : 2N$ **do**
      $y_l^{(j,k)} = \frac{\sum_{(t_y,t_x)\in\Omega^{(j,k)}} \tilde{y}_l^{(t_y,t_x)} \cdot w^{(t_y,t_x)}}{\sum_{(t_y,t_x)\in\Omega^{(j,k)}} w^{(t_y,t_x)}}$
    end
  end
end

B. Time Super-resolution: Exploiting the Prior Knowledge of the Sensing Matrix.

In [18], a deep multi-modal generative reconstruction model from direct-ToF single photon avalanche diode (SPAD) data was proposed to capture the depth-intensity cross-correlations, and achieve time SR. In addition, the memory footprint is reduced by projecting the raw data onto a randomly-generated matrix. Differently, in our work $\vec{y}$ is already constructed upon the projection of $\vec{x}$ onto $\left[\vec{a}^j\right]_{j=1}^n$ yielding a Compression Ratio (CR) $\mathrm{CR} = \frac{m}{n} = 0.0625$. However, the only use of $\boldsymbol{A}$ during the support estimation may lead to a catastrophic failure, when the index of the identified support element largely differs from the correct one, due to the presence of highly-correlated columns ($\mu \to 1$) far apart from each other, due to the impact of the pixel response function. We add a preliminary screening step accounting for the knowledge of the theoretical codes the demodulation functions are generated upon, the assembly of which yields $\boldsymbol{A_0}$. This improves the recovery performance. Also, our scheme only requires one single iteration to retrieve any s-sparse signal, yielding computational times of the same order of magnitude as ORMP[1] for $s > 1$. The SR in the time (depth) domain, described in Algorithm 2, consists of two steps:

- **Initial screening using custom (0,1)-binary codes:** We perform an scalar product of $\vec{y}^{(j,k)} = \left[y_i^{(j,k)}\right]_{i=1}^m$, with $1 \leq j \leq 2M$ and $1 \leq k \leq 2N$, with respect to the

[1]The obtained recovery times for $s = 2$ in our experiments are $\bar{t}_{\mathrm{ORMP}} = 0.22\,\mathrm{ms}$ and $\bar{t}_{\mathrm{ours}} = 0.66\,\mathrm{ms}$.

normalized columns of $\boldsymbol{A}_0 \in \mathbb{R}^{m \times n_0}$ and generate a discrete probability function $\vec{g}_0^{(j,k)} = \hat{\boldsymbol{A}}_0^{\top} \cdot \vec{y}^{(j,k)}$. We apply an initial thresholding $\mathcal{H}_{\kappa \cdot s}\left(\vec{g}_0^{(j,k)}\right)$ with $\kappa = 2$. Then, we estimate $K \geq s$ disjoint sets of feasible candidates, $\left[\Omega_{0,l}^{(j,k)}\right]_{l=1}^{K}$ being s the sparsity of the signal being retrieved.

- **Signal reconstruction using real pixel responses:** We generate K discrete probability functions $\vec{g}_l^{(j,k)} = \hat{\boldsymbol{A}}_{\Omega_{0,l}^{(j,k)}}^{\top} \cdot \vec{y}^{(j,k)}$ with $\hat{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ being $n = n_0 \cdot n_{\text{steps}}$. Then, we calculate $\vec{g}_{\max}^{(j,k)} = \left[\max_{1 \leq t \leq |\vec{g}_l^{(j,k)}|}\left(g_l^{t^{(j,k)}}\right)\right]_{l=1}^{K}$ composed by the maxima of the discrete probability functions of each group and the corresponding support $\Omega_{\max}^{(j,k)} = \cup_{l=1}^{K}\left\{\omega_{0,l}^{1^{(j,k)}} + \underset{1 \leq t \leq |\vec{g}_l^{(j,k)}|}{\arg\max}\left(g_l^{t^{(j,k)}}\right)\right\}$. We retrieve the vector $\tilde{\vec{r}}^{(j,k)}$ which is given by the elements of $\Omega_{\max}^{(j,k)}$ corresponding to the s largest entries of $\vec{g}_{\max}^{(j,k)}$ and multiply it by the grid size Δr to obtain the corresponding distance. Finally, we sort $\vec{\tilde{r}}^{(j,k)}$ and generate $\tilde{\boldsymbol{R}}_q$ distance maps with $q = 1 : s$.

In addition, we apply a **variational denoiser** [19] $\mathcal{D}\left(\tilde{\boldsymbol{R}}_i\right)$ to yield s $\boldsymbol{R}_q$ distance[2] maps.

Algorithm 2: 2-step coarse-to-fine greedy retrieval.

Data: $s, \kappa, n_{\text{steps}}, \boldsymbol{A}_0, \boldsymbol{A}, y_i$ with $1 \leq i \leq m$
Results: $\tilde{z}_q$ with $q = 1 : s$
Initialize: $K = 0$
$\vec{g}_0 = \mathcal{H}_{\kappa \cdot s}\left(\hat{\boldsymbol{A}}_0^{\top} \cdot \vec{y}\right)$
$\tilde{\Omega}_0 = \text{supp}\left(\vec{g}_0\right)$
for $t = 1 : |\tilde{\Omega}_0|$ **do**
  if $(\tilde{\omega}_0^{t+1} - \tilde{\omega}_0^{t}) > 1$ **then**
    $K = K + 1$
  end
  $\beta^t = K$
end
for $l = 1 : K$ **do**
  $\Gamma_l = \{t : \beta^t = l, \text{ s.t. } 1 \leq t \leq |\tilde{\Omega}_0|\}$
  $\Omega_{0,l} = \{1 + n_{\text{steps}} \cdot (\min\left(\tilde{\Omega}_{0,\Gamma_l}\right) - 1) :$
  $(n_{\text{steps}} \cdot \max\left(\tilde{\Omega}_{0,\Gamma_l}\right))\}$; $\tilde{\Omega}_{0,\Gamma_l} = \{\tilde{\Omega}_0[i]\}_{i \in \Gamma_l}$
  $\vec{g}_l = \hat{\boldsymbol{A}}_{\Omega_{0,l}}^{\top} \cdot \vec{y}$
  $g_{\max,l} = \max_{1 \leq t \leq |\vec{g}_l|}(g_l^t)$
  $\Omega_{\max,l} = \omega_{0,l}^{1} + \underset{1 \leq t \leq |\vec{g}_l|}{\arg\max}\left(g_l^t\right)$
end
$\tilde{\Omega} = \text{supp}\left(\mathcal{H}_s\left(\vec{g}_{\max}\right)\right)$
$\vec{\tilde{r}} = \text{sort}\left(\left[\omega_{\max,\tilde{\omega}_q} \cdot \Delta r\right]_{q=1}^{s}\right)$

[2] $z^{(j,k)} = r^{(j,k)} \sin\phi^{(j,k)} \sin\theta^{(j,k)}$, being $\phi^{(j,k)}$ and $\theta^{(j,k)}$, the observation angles of pixel (j,k) from the cartesian x-axis (lateral) and y-axis (vertical), respectively.

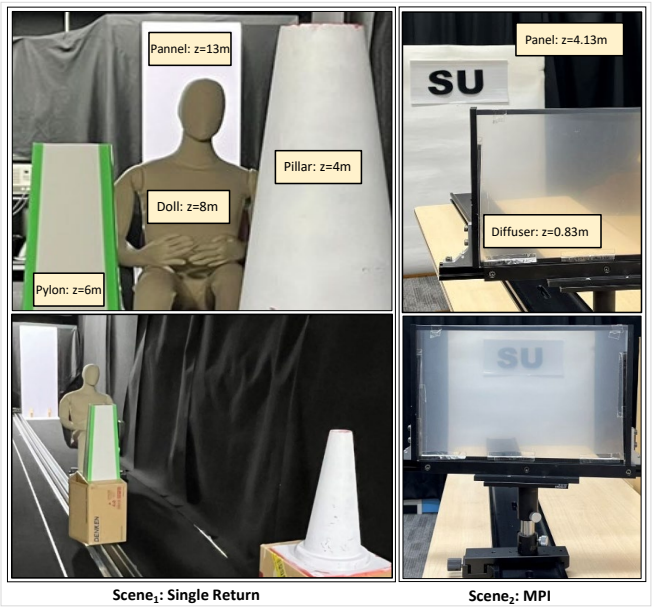

Fig. 2. Experimental setup for the evaluation of the spatio-temporal super-resolution scheme: scene_1 (left column) for single-return assessment contains four targets placed at various distances up to 13 m, whilst scene_2 (right column) for the evaluation of multi-path interference consists of a translucid diffuser and an opaque panel separated by 3.5 m.

C. Pre-processing Steps: Preventing Discrepancies between the Theoretical and Real Sensing Matrices.

We identify and amend various possible causes of discrepancy between the (0,1)-binary codes and the demodulation functions which may lead to a detriment on the performance of our algorithm. Firstly, there may exist a misalignment between the columns of the sensing matrix and the corresponding ones in the theoretical codes due to the **non-instantaneous status change in the pixel control signals.** This can be corrected by calculating the slope of the phase difference between both signals in the frequency domain [20], [21], and performing a complex rotation of the codes. Secondly, there may be **differences in modulation capacity of the sub-pixels** (sets of rows of the sensing matrix) which can be corrected by performing an equalization of $\vec{a}_i$ and $\vec{y}^{(j,k)}$, pre-multiplying both by a diagonal matrix $\boldsymbol{D}$ with $d_i^i = \|\vec{a}_i\|^{-1}$ with $1 \leq i \leq m$. This becomes of special relevance for Sliced OMP [4] in very noisy environments and low-light conditions, in which a preliminary set of measurements representing disjoint depth sub-domains, is used to restrict the signal support. Thus, the differences in modulation capacity need to be reduced to a minimum to allow for a correct determination of the signal support. Finally, **high-frequency noise** may be present on the measured demodulation functions. This can be removed if a low-pass filter matching the pixel bandwidth is applied to $\vec{a}_i$.

IV. Experimental Evaluation

A. Hardware Constrains on Demodulation Functions

The configuration of our camera allows for the acquisition of up to $m = 16$ measurements per pixel on a single shot.

TABLE I
EVALUATION OF THE SPATIO-TEMPORAL SUPER-RESOLUTION SCHEME FOR scene_2 (MULTI-PATH INTERFERENCE)

Methodology	Target	Histogram Max.	Mean(z)	σ_z	Δz	$\mathrm{RMSE}_{\Delta z}$
Up-scaling + 2-step	Panel Diffuser	4.220 m 0.815 m	4.236 m 0.811 m	0.277 m 0.115 m	3.405 m (+0.105 m)	0.324 m
Spatial SR + ORMP	Panel Diffuser	3.850 m 0.795 m	3.834 m 0.775 m	0.8258 m 0.260 m	3.055 m (-0.245 m)	0.813 m
Spatial SR + 2-step	Panel Diffuser	4.215 m 0.825 m	4.240 m 0.810 m	0.297 m 0.112 m	3.390 m (+0.090 m)	0.342 m
Prior work [1]	-	-	-	-	3.465 m (+0.165 m)	0.310 m

This prevents the appearance of motion artifacts due to relative displacement between acquisitions [22], [23]. However, the design of the demodulation functions must follow a number of rules derived from the layout of the sensor and hardware capabilities. The amount of charge transferred to the taps may become uncertain if more than one tap is on simultaneously for a multi-tap charge modulator. Also, the $m_{\mathrm{deg}} = 4$ taps do not have a dedicated drain and there may exist saturation and leaks to other taps if none of the taps of the sub-pixels are on. These two considerations translate to a number of non-zero elements per column, $n_{\mathrm{deg}} = 4$, distributed in such a way that no more than one non-zero element coexists in each of the sub-pixels. We consider the codes and the corresponding pixel demodulation functions in [1], as for both $\mu < 1$.

B. Experimental Setup

We empirically evaluate the reconstruction methodology using the setup presented in Fig. 2. We make use of the complementary metal oxide semiconductor (CMOS) ultra-high speed (UHS) ToF sensor [1] together with a laser emitter of wavelength $\lambda = 660\,\mathrm{nm}$ and a measured Full Width at Half Maximum $\mathrm{FWHM} = 2.55\,\mathrm{ns}$. Since the maximum range in our setup, given by the illumination system, is $r_{\max} = 16\,\mathrm{m}$, we can use up to $256\,\mathrm{bits}$ in $8-$bits increments yielding $n_0 = 32$. We discretize each of the n_0 elements in $n_{\mathrm{steps}} = 10$ sub-divisions to ensure $\mu < 1$, yielding $n \leq 320$ samples, i. e., a grid size $\Delta r = 0.05\,\mathrm{m}$. The raw images employed in this analysis are averaged over $n_{\mathrm{real}} = 100$ realizations. The scene_1 consists of several opaque targets placed at different distances with respect to our camera, being the maximum range $r_{\max} = 16\,\mathrm{m}$. For the evaluation of Multi-Path Interference (MPI), we use the data from scene_2 [1] and perform a shift of the origin of the coordinate system [1] as per the actual distances between the targets and the sensor. The scene_2 consists of one opaque panel at $z_{\mathrm{panel}} = 4.13\,\mathrm{m}$ plus a diffuser at $z_{\mathrm{diffuser}} = 0.83\,\mathrm{m}$.

C. Experimental Results

We evaluate the performance of the spatial SR scheme with respect to bi-cubic interpolation for $s = 1$ by presenting the depth and intensity maps using the proposed two-step recovery algorithm in the depth domain for scene_1. Table II presents the average depth reconstruction error, which is given by the difference between the average retrieved depth and the corresponding theoretical depth for each of the targets. Although the results presented are limited by the experimental setup constrains and the unavailability of 3D ground truth (GT) for scene_1 (s = 1), they allow for a fair comparison with prior work [1]. We observe an increase of the level of detail in the recovered targets, especially in dark areas (see inset in Fig. 3). Also, we find that the introduction of the bilateral filter helps to reduce the overall level of noise, the variability with respect to [1], and the number of spurious outliers in dark areas.

We choose a variation of this algorithm, ORMP, as a base case to allow for comparison with previous works which make use of time gating or coded demodulation [24], [25] which use OMP as a baseline for their validations. We observe an improvement of the depth accuracy on the former one. Table I compares the performance of the proposed scheme for the recovery of $2-$sparse signals (scene_2). We obtain a slight improvement of estimation of the relative distance when introducing our spatial SR scheme. We find that our technique correctly detects the depths of both targets, presents smaller depth recovery errors than ORMP and than reported in prior work [1], and improves the estimation of the relative distance (Δz) between both targets, given by the difference of the histogram maxima of each depth map, although the variability of the results is marginally increased. This variability could be decreased if μ is further optimized.

TABLE II
EVALUATION OF THE SPATIO-TEMPORAL SUPER-RESOLUTION SCHEME FOR scene_1 (SINGLE RETURN PER OBSERVED DIRECTION)

Absolute depth estimation error				
Method	Panel	Pillar	Pylon	Doll
Up-scaling+2-step	0.073 m	0.067 m	0.010 m	0.003 m
Spatial SR + ORMP	0.177 m	0.036 m	0.047 m	0.089 m
Spatial SR +2-step	0.074 m	0.066 m	0.010 m	0.002 m
Prior work	0.048 m	0.126 m	0.071 m	0.058 m

V. SUMMARY AND CONCLUSIONS

In this work, we have described a practical scheme for the pre-processing of the raw images of a single-shot ToF camera to achieve spatial sub-pixel super-resolution. We have shown that our methodology outperforms the results using bi-cubic interpolation for $s = 1$, especially in poorly illuminated areas. Also, we have introduced an algorithm to recover a 3D scene for $s \leq 2$, which enhances the depth resolution of the ToF sensor and overcomes ORMP and [1], with a similar computational cost. Our prospective research includes a detailed evaluation of two-step coarse-to-fine greedy retrieval together with the implementation of adaptive or optimal

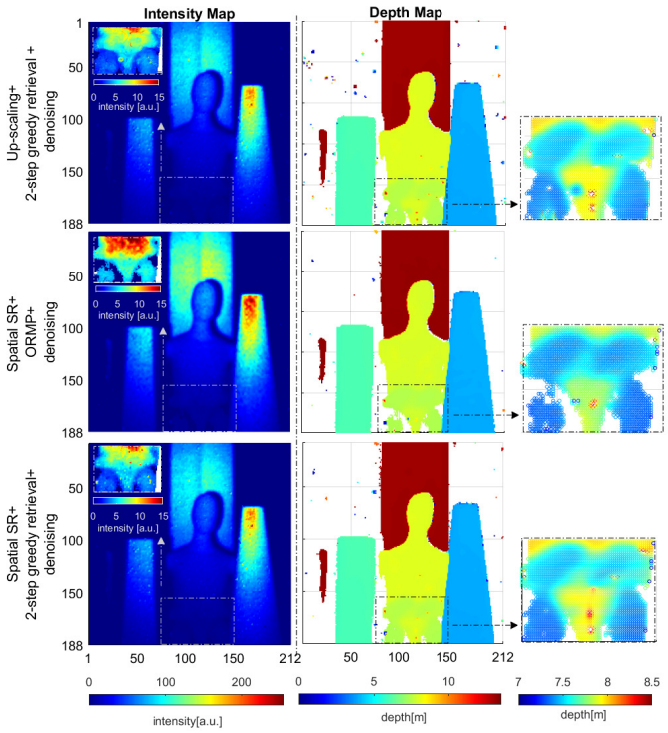

Fig. 3. Evaluation of spatio-temporal super-resolution scheme for scene_1: depth and intensity maps recoverd via bi-cubic interpolation + two-step recovery (top), spatial super-resolution + ORMP (center), and spatial super-resolution + two-step recovery (bottom).

demodulation coding schemes [4] to improve the recovery performance, especially under low-light conditions and MPI, and on techniques for depth denoising based on deep learning to reduce the execution times.

REFERENCES

[1] K. Kagawa, M. Horio, A. N. Pham, T. Ibrahim, S.-i. Okihara, T. Furuhashi, T. Takasawa, K. Yasutomi, S. Kawahito, and H. Nagahara, "A dual-mode 303-megaframes-per-second charge-domain time-compressive computational CMOS image sensor," *Sensors*, vol. 22, no. 5, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/5/1953

[2] M. Horio, Y. Feng, T. Kokado, T. Takasawa, K. Yasutomi, S. Kawahito, T. Komuro, H. Nagahara, and K. Kagawa, "Resolving multi-path interference in compressive Time-of-Flight depth imaging with a multi-tap macro-pixel computational cmos image sensor," *Sensors*, vol. 22, no. 7, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/7/2442

[3] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, 1995.

[4] A. Lopez Paredes, M. Heredia Conde, and O. Loffeld, "Sparsity-aware 3D ToF sensing," *IEEE Sensors Journal*, vol. 23, no. 4, pp. 1–1, 2023.

[5] A. Bhandari, M. Heredia Conde, and O. Loffeld, "One-bit time-resolved imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1630–1641, 2020.

[6] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[7] M. Gupta, A. Velten, S. K. Nayar, and E. Breitbach, "What are optimal coding functions for Time-of-Flight imaging?" *ACM Trans. Graph.*, vol. 37, no. 2, Feb. 2018. [Online]. Available: https://doi.org/10.1145/3152155

[8] M. Heredia Conde, K. Hartmann, and O. Loffeld, "Simple adaptive progressive edge-growth construction of LDPC codes for close(r)-to-optimal sensing in pulsed ToF," in *2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, 2016, pp. 80–84.

[9] X.-Y. Hu, E. Eleftheriou, and D.-M. Arnold, "Progressive edge-growth Tanner graphs," in *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No.01CH37270)*, vol. 2, 2001, pp. 995–1001 vol.2.

[10] G. Sriruthchataboon, A. Bajpai, L. Wuttisittikulkij, and P. Kovintavewat, "PEG-like algorithm for LDPC codes," in *2014 International Electrical Engineering Congress (iEECON)*, 2014, pp. 1–4.

[11] D. Chan, H. Buisman, C. Theobalt, and S. Thrun, "A noise-aware filter for real-time depth upsampling," *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 10 2008.

[12] Z.-H. Niu, X.-P. Lin, A.-N. Yu, Y.-H. Zhou, and Y.-B. Yang, "Lightweight and accurate single image super-resolution with channel segregation network," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1630–1634.

[13] A. Villar-Corrales, F. Schirrmacher, and C. Riess, "Deep learning architectural designs for super-resolution of noisy images," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1635–1639.

[14] W. Zhang, P. Song, X. Wang, Z. Zheng, Y. Bai, and H. Geng, "Fast lightweight framework for time-of-flight super-resolution based on block compressed sensing," *Opt. Express*, vol. 30, no. 9, pp. 15 096–15 112, Apr 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-9-15096

[15] F. Li, H. Chen, A. Pediredla, C. Yeh, K. He, A. Veeraraghavan, and O. Cossairt, "CS-ToF: High-resolution compressive time-of-flight imaging," *Opt. Express*, vol. 25, no. 25, pp. 31 096–31 110, Dec 2017. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-25-25-31096

[16] C. Damian, F. Garoi, C. Udrea, and D. Coltuc, "The evaluation of single-pixel camera resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2517–2523, 2020.

[17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 839–846.

[18] V. Poisson, V. T. Nguyen, W. Guicquero, and G. Sicard, "Luminance-depth reconstruction from compressed time-of-flight histograms," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 148–161, 2022.

[19] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'09)*, 04 2009.

[20] M. Heredia Conde, "Raw data processing for practical Time-of-Flight super-resolution," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1650–1654.

[21] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–14, 2006.

[22] M. Heredia Conde, K. Kagawa, T. Kokado, S. Kawahito, and O. Loffeld, "Single-shot real-time multiple-path Time-of-Flight depth imaging for multi-aperture and macro-pixel sensors," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1469–1473.

[23] F. Li, P. Ruiz, O. Cossairt, and A. K. Katsaggelos, "Multi-frame super-resolution for time-of-flight imaging," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2327–2331.

[24] 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.*, vol. 39, no. 6, pp. 1705–1708, Mar 2014. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-39-6-1705

[25] A. Bhandari, A. M. Wallace, and R. Raskar, "Super-resolved time-of-flight sensing via fri sampling theory," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE Press, 2016, p. 4009–4013. [Online]. Available: https://doi.org/10.1109/ICASSP.2016.7472430