On the Formulation of Coded Demodulation and 3D Reconstruction in Rotating PB-ToF sensors

Alvaro Lopez Paredes, Miguel Heredia Conde
Center for Sensor Systems (ZESS), University of Siegen, Germany
alvaro.lparedes@uni-siegen.de, heredia@zess.uni-siegen.de

***Abstract*—The use of Time-of-Flight cameras has been recently popularized due to the relatively low-power consumption and manageable size. However, the optical system imposes a restriction in the field of view of the camera and motion artifacts appear in unsteady environments. In this paper, we present a novel sensing system that enables flexible trade-offs between the field of view and the lateral resolution, no longer fixed by the array geometry, in close to real-time operation. Our approach is based on the introduction of a controlled rotation during sensing. As a result, we propose an original sparsity-aware greedy algorithm for 3D reconstruction. The novel computational sensor is validated via numerical simulations.**

***Index Terms*—Coded demodulation, Compressive Sensing, Depth priors, 3D imaging, Time-of-Flight**

I. Introduction

In this paper, we examine three important topics for the current and future development of indirect Pulse-based Time-of-Flight (PB-ToF) cameras. Firstly, the increase of the angular range at bounded hardware and computational complexity. Secondly, the enhancement of the lateral resolution. Finally, the transformation of an adverse circumstance, such as the relative motion between the camera and the scene, into a useful tool during sensing.

Our methodology relies on four fundamental aspects. Firstly, we make use of Compressive Sensing (CS) techniques to reduce the number of measurements required for 3D reconstruction, and allow for real-time operation. Secondly, we design a near-to-optimal coding scheme [1] which yields temporal super-resolution, i. e., the recovery of the target's depth in a grid much finer than the number of elements of the codes that the demodulation functions are built upon. Thirdly, we exploit some a priori knowledge of the signals being recovered, i. e., the relationship between the depth at which the target is found and the perceived intensity. Finally, we make use of the rotation of the sensing system, and the derived sampling redundancy, to uncouple the lateral resolution from the pixel count.

Specifically, the two main contributions of this paper are:

1) An original 360° PB-ToF 3D sensing system leveraging coded demodulation.
2) A novel algorithm to reach temporal super-resolution whilst uncoupling the lateral resolution from the number of pixels in the introduced rotating PB-ToF camera.

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 (MENELAOS^{NT}).

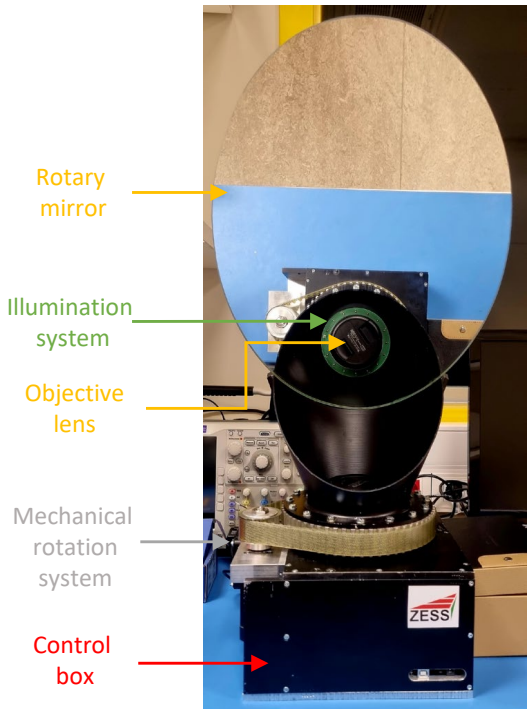

Fig. 1. Components of the camera prototype developed at ZESS.

II. General Desciption of the Camera System

As shown in Fig. 1, our prototype is designed to allow for rotation whilst minimizing the number of moving parts during operation. The camera comprises four sub-systems. The sensing system consists of a PB-ToF sensor array, where each pixel can be seen as a correlator between a delayed pulse, the delay of which encodes the depth of the target, and a number of demodulation functions [1]. The illumination system comprises a number of vertical-cavity surface-emitting lasers (VCSELs). The optical system is composed of a large-aperture lens capable of operating in very-low-light conditions and reach medium-to-long ranges, and a mirror rotating around the vertical axis and placed at $45°$ with respect to it. The mirror is used to increase the angular range, redirecting the light paths from the illumination system to the scene, and from the scene to the sensor. The mechanical rotation system consists of two externally toothed gears, and a truncated cone.

III. Methodology

A. ToF Camera Geometrical Modelling

As shown in Fig. 2, we make use of the observation direction $\vec{n}^{(j,k)}$ for each pixel to retrieve the distance to the target $r_P^{(j,k)}$, with $-M/2 \leq j \leq M/2$ and $-N/2 \leq k \leq N/2$, being M and N the number of rows and columns of the sensor array, respectively. The time-dependant geometrical model of the sensing system can be analytically modelled as follows: The mirror plane π_0 is given by $P_0 = (0, 0, d_{\pi_0})$ and the normal $\vec{n}^{\pi_0}$ (1), being ω_1 the rotational velocity of the mirror around z_1-axis, and d_{π_0} the height at which z_1-axis intersects the mirror.

$$\vec{n}^{\pi_0}(t) = \boldsymbol{R}_{\mathrm{z}}(\omega_1 t)\, \boldsymbol{R}_{\mathrm{x}}(45\,^\circ)\, \vec{z}_1 \tag{1}$$

The observation direction for the pixel (j, k) is given by $\vec{n}^{(j,k)}$ (2), being β_z, and β_x the angles formed by the y_1-axis and the projections of $\vec{n}^{(j,k)}$ onto the $O_{x_1 y_1}$ and $O_{y_1 z_1}$ planes, respectively.

$$\vec{n}^{(j,k)} = \boldsymbol{R}_{\mathrm{z}}(\beta_z)\, \boldsymbol{R}_{\mathrm{x}}(\beta_x)\, \vec{z}_1 \tag{2}$$

The intersection point between the line of sight and the mirror plane is given by $P = \alpha \vec{n}^{(j,k)}$, with α such that $\alpha \vec{n}^{(j,k)} \in \pi_0$. Then, we calculate the distance with respect to the camera reference system origin $\Delta\vec{r}_{12}^{(j,k)} = O_1 - P^{(j,k)}$, and, finally, the re-directed observation direction, given by the coordinates of $P^{(j,k)}$ and the direction vector $\vec{n}'^{(j,k)}$ (3).

$$\vec{n}'^{(j,k)}(t) = \boldsymbol{R}_{\mathrm{z}}(\beta_z + \omega_1 t)\, \boldsymbol{R}_{\mathrm{x}}(90\,^\circ + \beta_x)\, \vec{z}_1 \tag{3}$$

The position vector of the observed point with respect to the pixel (j, k), supposing $O_0 \equiv O_1$ is given in (4).

$$\vec{r}^{(j,k)}(t) = \cancel{\Delta\vec{r}_{0,1}} + \Delta\vec{r}_{1,2}^{(j,k)}(t) + \lambda \cdot \vec{n}'^{(j,k)}(t) \tag{4}$$

The coordinates of the virtual focal point are given in (5). Note that, differently from [1], [2] the virtual focal point O_2 is time-dependant.

$$\vec{O_1 O_2}(t) = d_{\pi_0} \cdot \boldsymbol{R}_{\mathrm{z}}(\omega_1 t)\, (0, -1, 1)^{\intercal} \tag{5}$$

B. ToF Sensing Scheme

We consider two different sensing approaches depending on wether ω_1 is high enough to induce significant changes in the observation direction during the exposure time t_{exp}.

1) Semi-continuous rotation: In this regimen of operation, the camera stops at each angular pose to take up to m measurements. In this case, the observation direction[1] $\vec{n}' = \vec{n}'(t = t_0)$ remains constant during t_{exp}. The emitted signal $p_{\vec{r}}(t)$, given by a train of light pulses periodically repeated at f_{r} during t_{exp}, interacts with the scene represented by the Scene Response Function (SRF) defined in (6), where $\delta(t)$ and $\{\Gamma_{\vec{r}}[i],\ \tau_{\vec{r}}[i]\}_{i=0}^{s-1}$ are the Dirac delta function and the reflectivities and time delays introduced by s return paths of light [3], respectively.

[1]For readability, hereinafter, we refer to the pixel (j, k) and omit the super-script in the formulation.

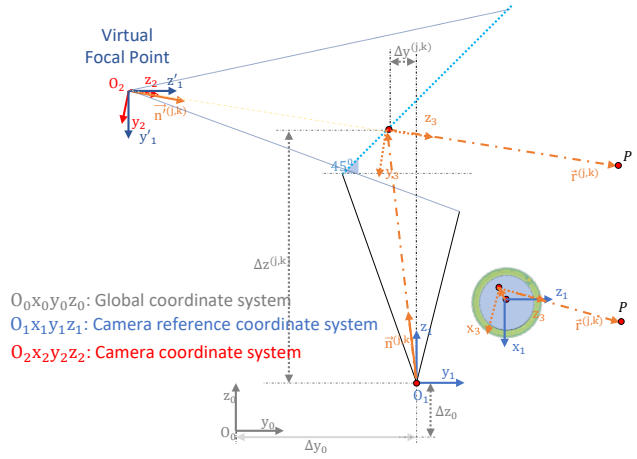

Fig. 2. Schematic of the observation geometry of the proposed ToF camera.

$$h_{\vec{r}}(t, t') = \sum_{l=0}^{s-1} \Gamma_{\vec{r}}[l]\, \delta(t - t' - \tau_{\vec{r}}[l]) \tag{6}$$

Then, the reflected signal can be formulated as the convolution of both signals $r_{\vec{r}}(t) = (p_{\vec{r}} * h_{\vec{r}})(t)$. The pixel correlates $r_{\vec{r}}$ with m controllable shift-invariant functions $[a_{0,i}]_{i=1}^{m}(t)$ (7).

$$y_{i,\vec{r}}(t) = ((p_{\vec{r}} * h_{\vec{r}}) \otimes a_{0,i})(t), \quad 1 \leq i \leq m \tag{7}$$

Ideally, the demodulation functions are a periodic repetition of binary codes of n elements, such that $a_{0,i}^{l} = a_{0,i}^{l+n}$ or, equivalently, in the continuous-time domain $a_{0,i}(t) = a_{0,i}(t + t_{\mathbf{r}}),\quad 0 < t \leq t_{\mathrm{exp}} - t_{\mathbf{r}}$, being t_{r} the code duration. Each of the n elements are further discretized in n_{steps} sub-steps yielding n_{samples} sub-divisions for each code. Equation (7) translates into (8), by exploiting cyclic convolution properties [3].

$$y_{i,\vec{r}}(t) = ((p_{\vec{r}} \otimes a_{0,i}) * h_{\vec{r}})(t), \quad 1 \leq i \leq m \tag{8}$$

As $\vec{n}'$ remains constant during t_{exp}, (8) yields (9) in the discrete case, where $\vec{y}_{\vec{r}} = [y_{i,\vec{r}}(t = 0)]_{i=1}^{m}$ is the measurement vector, $\boldsymbol{A}$ is the sensing (or measurement) matrix with $\boldsymbol{A} = \left[\left[a_i^l\right]_{i=1}^{m}\right]_{l=1}^{n_{\mathrm{samples}}} := (p_{\vec{r}} \otimes a_{0,i})(t_l)$, and $\vec{x}_{\vec{r}} := [h_{\vec{r}}(t_l)]_{l=1}^{n_{\mathrm{samples}}}$ is the signal to be reconstructed.

$$\vec{y}_{\vec{r}} = \boldsymbol{A} \cdot \vec{x}_{\vec{r}} \tag{9}$$

2) Continuous rotation: Differently, if ω_1 is high enough to induce a relevant change of $\vec{n}'(t)$ with $0 \leq t < t_{\mathrm{exp}}$ (3), the time-domain cross-correlation $\vec{a}_i$ will effectively sense over a 1D curve in the 3D spatial domain $\left[\vec{a}_i^{\mathrm{3D}}\right]_{i=1}^{m}$ (10), as illustrated in Fig. 3. In order to guarantee $\mu < 1$ and, therefore, uniqueness of the solution of the inverse problem,

$t_\mathrm{r} \geq t_\mathrm{exp}$, and the number and bandwidth of the elements shall be adjusted accordingly.

$$\vec{a}_i^{\,3\mathrm{D}}\left(\vec{r}\right) = \begin{cases} a_i\left(t\right) & \text{for} \quad \vec{r} = \dfrac{c(t - \lfloor \frac{t}{t_\mathrm{r}} \rfloor)}{2} \cdot \vec{n}'\left(t\right) \\ 0 & \text{elsewhere} \end{cases} \tag{10}$$

We translate the formulation from the continuous-time domain, into a global discrete grid expressed in spherical coordinates $D := \{\Gamma_k\}_{k=1}^{n_D}$ where $\Gamma_k = [\theta_k, \phi_k, \rho_k]$. Firstly, we define a set of n_tot equidistributed times $\Upsilon_\mathrm{t} = \{t_l\}_{l=1}^{n_\mathrm{tot}}$ with $t_l \in [0, t_\mathrm{exp})$, and calculate the corresponding cross-correlations $\Upsilon_\mathrm{a} = \{a_i\left(t_l\right)\}_{l=1}^{n_\mathrm{tot}}$, and position vectors $\Upsilon_\mathrm{q} = \{\vec{q}_l\}_{l=1}^{n_\mathrm{tot}}$ such that $\vec{q}_l = \vec{r}(t_l)$. Secondly, we estimate the cross-correlation $a_i^{3\mathrm{D}}(\vec{\rho}_k)$ (11) for each $\vec{\rho}_k \in D$, being $\mathcal{I}$ an interpolation operator, which represents a row of $\boldsymbol{A}^{3\mathrm{D}}$ and explains how each of the scattered points of the grid contributes to the pixel measurement. Thus, the pixel measurements are $y_i = a_i^{3\mathrm{D}}\left(\vec{\rho}_k\right) \cdot \vec{x}_{\vec{\rho}_k}$.

$$\vec{a}_i^{\,3\mathrm{D}}\left(\vec{\rho}_k\right) = \begin{cases} \mathcal{I}\left(\vec{\rho}_k, \Upsilon_\mathrm{q}, \Upsilon_\mathrm{a}\right) & \text{if} \quad |\vec{q} - \vec{\rho}_k| < \varepsilon \\ 0 & \text{elsewhere} \end{cases} \tag{11}$$

$\boldsymbol{A}^{3\mathrm{D}}$ shall now account for all m measurements and all pixels $\left[\left[\left[a_i^{(j,k)}\right]_{i=1}^{m}\right]_{j=1}^{M}\right]_{k=1}^{N}$ (12). The resulting matrix is structured and features very-low density. The recovery of the sparse targets in D can now be solved jointly for all the pixels of the sensor array.

$$\vec{y}_{\vec{\rho}} = \boldsymbol{A}_{\vec{\rho}}^{3\mathrm{D}} \cdot \vec{x}_{\vec{\rho}} \tag{12}$$

If $\boldsymbol{A}$ exhibits an eminently block diagonal structure, the limited measurement entanglement allows for working on a *per-pixel* basis. As in the previous case, we generate Υ_t, Υ_a, and Υ_q, and calculate the cross-correlations as $\vec{a}_i^{\,3\mathrm{D}} = \left[a_i^l\left(\vec{q}_l\right)\right]_{l=1}^{n_\mathrm{tot}}$, and the resulting sensing matrix $\boldsymbol{A}_j^{3\mathrm{D}}$ (for pixel j). This methodology is computationally advantageous, as it allows to sequentially recover the signals for each single pixel, which reduces the size and complexity of the problem. However, it may lead to sampling inhomogeneities due to the relative motion between the mirror and sensor.

C. Signal Recovery and Image Reconstruction

Since $m \ll n_\mathrm{samples}$, (9) and (12) are under-determined. We exploit the sparsity of $\vec{x}$, yielding a constrained ℓ_0-minimization problem (13), i. e., to find the sparsest solution which complies with (9) or (12).

$$\hat{\vec{x}}_{\vec{r}} = \underset{\vec{x}_{\vec{r}}}{\arg\min} |\vec{x}|_0, \quad \text{s.t.:} \quad \vec{y}_{\vec{r}} = \boldsymbol{A} \cdot \vec{x}_{\vec{r}} \tag{13}$$

Generally, in classical greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [4], a discrete probability density function $\vec{g}$ is generated as the scalar product of the normalized columns of $\boldsymbol{A}$ and the residual of the measurement vector $\vec{y}_{\vec{r}}$. Then, the target location is identified as the first peak of this

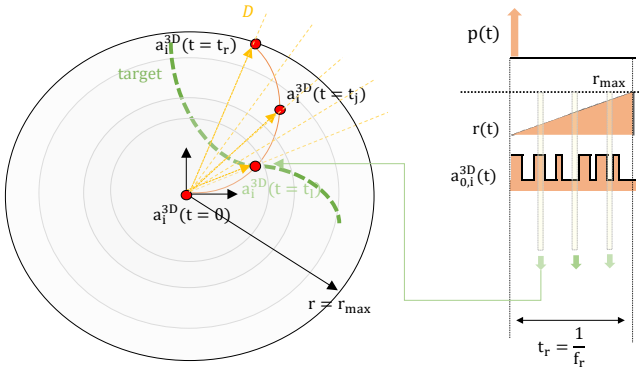

Fig. 3. Sensing scheme in continuous rotation mode.

distribution for $s = 1$. This assumption may not be correct in some scenarios, such as for low SNR signals, i. e., under low-lighting conditions or multi-path interference (MPI).

The use of prior modeling based on the relationship between the depth and intensity maps have been applied in tasks such as depth super-resolution exploiting High Resolution (HR) intensity maps [5]. As described in Algorithm 1, we propose to incorporate an empirical prior 2D distribution over depth an intensity $\vec{g}_\mathrm{DI}$ during the calculation of the discrete probability density function $\hat{\vec{g}} = \vec{g} \odot \vec{g}_\mathrm{DI}$, which accounts for unlikely depth-intensity pairs and, therefore, helps to reduce large reconstruction errors. We make use of the transient profiles and depth maps from the state-of-the-art dataset [6] as ground truth (GT). This dataset consists of $n_\mathrm{scenes} = 25$ scenes with a resolution of $120\,\mathrm{px} \times 160\,\mathrm{px}$. We select 21 of them to calculate an empirical 2D discrete probability density function, i. e., 3.84×10^5 transient profiles, and use the remaining four scenes for testing and validation of the proposed recovery algorithm. We perform a pre-processing of the depth maps by removing any existing flying pixels, as per Neighborhood Distance (ND) methodology [7]. We extract and merge the depth-intensity histograms of the 21 scenes, being the intensity approximated as the ℓ_2-norm of each transient profile which consists of $n_\mathrm{samples} = 2000$ samples, and estimate $\boldsymbol{G}_\mathrm{DI}^\mathrm{raw}$, such that $\vec{g}_\mathrm{DI}^\mathrm{raw}(\tilde{I}) = \left[P\left(l | I = \tilde{I}\right)\right]_{l=1}^{n_\mathrm{samples}}$. As shown in Fig. 4, we perform 1D smoothing via least-squares followed by non-dimensionalization, yielding $\vec{g}_\mathrm{DI}(\tilde{I})$. Then, we define the cumulative distribution function, such that $\left[g_{\mathrm{DI\,cum}}^l\left(\tilde{I}\right)\right]_{l=1}^{n_\mathrm{samples}} = \left[\sum_{s<l} g_\mathrm{DI}^s\left(\tilde{I}\right)\right]_{l=1}^{n_\mathrm{samples}}$, and finally, the indicator function $\left[\tilde{g}_\mathrm{DI}^l\left(\tilde{I}\right)\right]_{l=1}^{n_\mathrm{samples}}$ (14) being $\varepsilon \ll 1$.

$$\tilde{g}_\mathrm{DI}^l\left(\tilde{I}\right) = \begin{cases} 1 & \text{if} \quad g_{\mathrm{DI\,cum}}^l\left(\tilde{I}\right) < 1 - \varepsilon \\ 0 & \text{elsewhere} \end{cases} \tag{14}$$

D. An Approach to uniquely determine per-pixel depths at high angular rates.

The non-ambiguous range (r_max) is given by the maximum number of elements which guarantees the uniqueness of $\left[\vec{a}_0^{\,j}\right]_{j=1}^{n}$ ($\mu < 1$). In medium ranges $r_\mathrm{max} \leq \mathcal{O}\left(10^2\,\mathrm{m}\right)$,

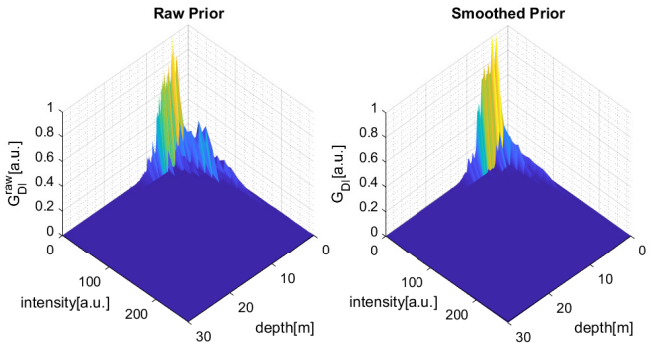

Fig. 4. Empirically estimated 2D discrete probability density function.

Algorithm 1: Continuous Greedy retrieval

Data: $\boldsymbol{A}$, $\vec{y}^{(k)}$, ε
Result: Φ
Initialize: $s=0$; $\Upsilon_{\mathrm{c}1}=\emptyset$; $\hat{\Upsilon}_{\mathrm{c}2}=\emptyset$; $\Gamma_{\mathrm{c}1}=\emptyset$; $\Gamma_{\mathrm{c}2}=\emptyset$
for $k=1:MN$ **do**
  $\beta^{(k,s)}=\emptyset$
  $\vec{x}^{(k,s)}=\vec{0}$
  $\vec{\varepsilon}^{(k,s)}=\vec{y}^{(k)}$
  $\Upsilon_{\mathrm{c}}^{(k,s)}=\emptyset$
  while $\left(|\beta^{(k,s)}|<s_{\max}\right)$ and $\left(\|\vec{\varepsilon}^{(k,s)}\|_2<\varepsilon\right)$ **do**
    $s=s+1$
    for $k=1:MN$ **do**
      $\tilde{I}^{(k,s)}=\|\vec{\varepsilon}_{\vec{r}}^{(k,s-1)}\|_2$
      $\vec{g}^{(k,s)}=\left[\frac{(\vec{a}^l)^{\intercal}\vec{\varepsilon}_{\vec{r}}^{(k,s-1)}}{\|\vec{a}^l\|_2}\right]_{l=1}^{n_{\mathrm{samples}}}$
      $\vec{\tilde{g}}^{(k,s)}=\vec{g}^{(k,s)}\odot\vec{\tilde{g}}_{\mathrm{DI}}\left(\tilde{I}^{(k,s)}\right)$
      $l_{\max}=\underset{1\le l\le n_{\mathrm{samples}}}{\arg\max}\left(\hat{g}_l^{(k,s)}\right)$
      $\beta^{(k,s)}=\beta^{(k,s-1)}\cup l_{\max}$
      $\vec{x}^{(k,s)}=\boldsymbol{A}_{\boldsymbol{\beta^{(k,s)}}}{}^{\dagger}\cdot\vec{y}^{(k)}$
      $\vec{\varepsilon}^{(k,s)}=\vec{y}^{(k)}-\boldsymbol{A}_{\boldsymbol{\beta^{(k,s)}}}\cdot\vec{x}^{(k,s)}$
      $\Upsilon_{\mathrm{c}}^{(k,s)}=\left\{\vec{r}_i\left(t(l_{\max})+i\cdot t_{\mathrm{r}}\right)\right\}_{i=0}^{\left\lfloor\frac{t_{\exp}}{t_{\mathrm{r}}}\right\rfloor-1}$
    end
    if $\|\varepsilon^{(k,1)}\|_2<\varepsilon$ **then**
      $\Gamma_{c1}=\Gamma_{c1}\cup\{k\}$
      $\Upsilon_{c1}=\Upsilon_{c1}\cup\left\{\Upsilon_c^{(k,s)}\right\}$
    else
      $\Gamma_{c2}=\Gamma_{c2}\cup\{k\}$
      $\hat{\Upsilon}_{c2}=\hat{\Upsilon}_{c2}\cup\left\{\Upsilon_c^{(k,s)}\right\}$
    end
  end
end
$\Upsilon_{\mathrm{c}2}=\cup_{k=1}^{|\hat{\Upsilon}_{c2}|}\left\{\hat{\Upsilon}_{\mathrm{c}2}^k\cap\left\{\Upsilon_{\mathrm{c}1}^l\right\}_{l=1}^{|\Upsilon_{c1}|}\right\}$
$\Phi_{\mathrm{c}1}=\cup_{k=1}^{|\Gamma_{c1}|}\left(\Gamma_{c1}^k,\Upsilon_{\mathrm{c}1}^k\right)$
$\Phi_{\mathrm{c}2}=\cup_{k=1}^{|\Gamma_{c2}|}\left(\Gamma_{c2}^k,\Upsilon_{\mathrm{c}2}^k\right)$
$\Phi=\Phi_{c1}\cup\Phi_{c2}$

and $t_{\mathrm{r}}\ll t_{\exp}$, as $t_{\exp}\sim\mathcal{O}(0.1\,\mathrm{ms})$. Thus, an unique solution may not be guaranteed in a *per-pixel* basis, as the code may successively be repeated $n_{\mathrm{rep}}=\left\lfloor\frac{t_{\exp}}{t_{\mathrm{r}}}\right\rfloor$ times during $t_{\exp}$, yielding an ill-posed problem. A general assumption is to consider that the observed direction remains invariant for low and moderate angular velocities $\omega_1\sim\mathcal{O}(1\,\mathrm{fps})$, as the angular range travelled during $t_{\exp}$ is such that yields negligible changes in azimuth for the range considered ($\Delta r_\theta\sim\mathcal{O}(\mathrm{mm})$ at $r\sim\mathcal{O}(10\,\mathrm{m})$). However, this assumption may lead to an erroneous determination of the distance at the object boundaries [8]. As presented in Algorithm 1, we propose a cost-effective approach to determine the depths observed by each single pixel during $t_{\exp}$. If the ℓ_2-norm of the residual of the measurement vector of a pixel does not exceed the noise level ε for $s=1$, all the observed depths during $t_{\exp}$ lies on an iso-radial region, and the single depth of the target over a limited angular range can be retrieved. Otherwise, the problem can be tackled in most cases as a reflective multi-path Interfence (MPI), such that the SRF can be explained by few Dirac deltas. In this case, we account for the intersection with the support of the SRF observed by the adjacent pixels (redundant sampled points), the trajectories of which are fully determined during $t_{\exp}$.

The proposed scheme is as follows:

1) We determine for each single pixel $[k]_{k=1}^{MN}$ a set of candidate 3D target points $\Upsilon_{\mathrm{c}}^{(k,s)}=\left\{\vec{r}_i(t=t(l_{\max})+i\cdot t_{\mathrm{r}})\right\}_{i=0}^{\left\lfloor\frac{t_{\exp}}{t_{\mathrm{r}}}\right\rfloor-1}$.
2) We classify the pixel k and the corresponding target points $\Upsilon_{\mathrm{c}}^{(k,s)}$ in $\Gamma_{\mathrm{c}1}$ and $\Upsilon_{\mathrm{c}1}$, respectively, if $\|\varepsilon^{(k,1)}\|_2<\varepsilon$ or, otherwise, in $\Gamma_{\mathrm{c}2}$ and $\hat{\Upsilon}_{\mathrm{c}2}$.
3) We determine $\left\{\Upsilon_{c2}^k\right\}_{k=1}^{|\hat{\Upsilon}_{c2}|}=\left\{\hat{\Upsilon}_{\mathrm{c}2}^k\cap\left\{\Upsilon_{\mathrm{c}1}^l\right\}_{l=1}^{|\Upsilon_{c1}|}\right\}$.
4) We obtain the observed points by all pixels during $t_{\exp}$ Φ as the union of $\Phi_{\mathrm{c}1}=\cup_{k=1}^{|\Gamma_{c1}|}\left(\Gamma_{c1}^k,\Upsilon_{\mathrm{c}1}^k\right)$, and $\Phi_{\mathrm{c}2}=\cup_{k=1}^{|\Gamma_{c2}|}\left(\Gamma_{c2}^k,\Upsilon_{\mathrm{c}2}^k\right)$.

Our approach exploits the underlying low-dimension of the union of sub-spaces where the 3D shapes live, at the time it avoids the need for training a 3D dictionary (costly task) or relying on parametric models.

IV. Numerical Simulations

A. Preliminary Considerations and Simulation Setup

We perform a numerical validation of the proposed methodology for the continuous rotation during $t_{\exp}=0.1\,\mathrm{ms}$ by making use of two testing scenes from [6]. We consider $n'_{\mathrm{rep}}=20$ equispaced sequences ($n'_{\mathrm{rep}}\times n_{\mathrm{samples}}$ observation directions) for each single pixel over $t_{\exp}$ for the generation of the ground truth (GT). We generate $m=10$ custom codes via Gradient Combinatorial [1], each of them consists of $n=128$ elements. Then, we further discretize the elements in $n_{\mathrm{steps}}=6$ sub-steps, yielding $n_{\mathrm{samples}}=768$ samples. With regards to the cross-correlation function, a Gaussian filter of standard deviation $\sigma=2\,\mathrm{ns}$ $(0.6\,\mathrm{m})$ is applied to the custom sequences to obtain a realistic representation of the measurement functions [9]. Also, we add Additive White

TABLE I
OPERATIONAL PARAMETERS OF OUR TOF CAMERA

Operational parameter	Acronym	Value
Modulation frequency	f_{m}	448 MHz
Repetition frequency	f_{r}	3.5 MHz
Horizontal field of view	FOV	30 °
Vertical distance mirror-sensor	d_{π_0}	0.3 m
Angular rate	ω_1	0 rpm − 600 rpm
Exposure time	t_{exp}	0.1 ms
Number of measurements	m	10
Grid resolution	Δr	0.055 m
Time for 360° reconstruction	$t_{360°}$	1 s − 2 s

Gaussian Noise (AWGN) with various Signal-to-Noise Ratio (SNR) levels $\mathrm{SNR} \in [-20\,\mathrm{dB}, 120\,\mathrm{dB}]$ to corrupt the m measurements. As the field of view (FOV) of the dataset is greater than the FOV of our camera, we define a region of interest in the scene, by performing a translation of the origin of the camera reference system. Table I summarizes the operating parameters of our camera.

B. Numerical Results

Fig. 5 presents the depth reconstruction errors in terms of Root Mean Square Error (RMSE) over $n_{\mathrm{real}} = 16$ realizations and the percentage of total number of recovered points with respect to GT, for various angular rates $\omega_1 \leq 600\,\mathrm{rpm}$. We perform a comparison of our methodology with the results obtained via OMP (only one iso-radial distance per pixel during t_{exp}) to allow for comparison with other works in the field [10], [11]. The figures corresponding to depth reconstruction error show four differentiated regions: a first area dominated by noise and characterized by large reconstruction errors, followed by a plateau with $\mathrm{RMSE} \sim \mathcal{O}(\mathrm{cm})$, and a decay until a residual error $\mathrm{RMSE} \leq 2\,\mathrm{cm}$ for $\mathrm{SNR} \geq 70\,\mathrm{dB}$ in both scenes. This residual error is due to the presence of target's depths which lie in adjacent highly-correlated columns in $\boldsymbol{A}$ over t_{exp}. The error would decrease if μ was further reduced (for instance, by introducing offsets between each of the demodulation functions [2] or increasing f_{m}). We observe that this residual error decreases when reducing ω_1. We find that that our methodology yields smaller reconstruction errors than OMP for the same angular rate. Also, we observe that the implementation of the prior yields a net improvement of the accuracy for very noisy signals, and that its impact progressively diminishes up to a noise threshold $\mathrm{SNR}_0 = 10\,\mathrm{dB}$ for $\mathrm{bathroom} - \mathrm{cycles} - 2$ and $\mathrm{SNR}_0 = 0\,\mathrm{dB}$ for $\mathrm{hot} - \mathrm{living}$, from which no further improvement is obtained. Finally, we show that we are able to uniquely determine $\geq 95\%$ of the points observed for all pixels for $\omega_1 \leq 600\,\mathrm{rpm}$ in both scenes.

V. CONCLUSIONS AND PROSPECTIVE WORK

In this paper, we have presented an original 360° PB-ToF 3D camera that enables flexible trade-offs between the FOV and the lateral resolution. Moreover, we have proposed an original methodology which exploits the continuous rotation of the camera to accurately reconstruct the depths observed by each single pixel during t_{exp}. We show the adequacy

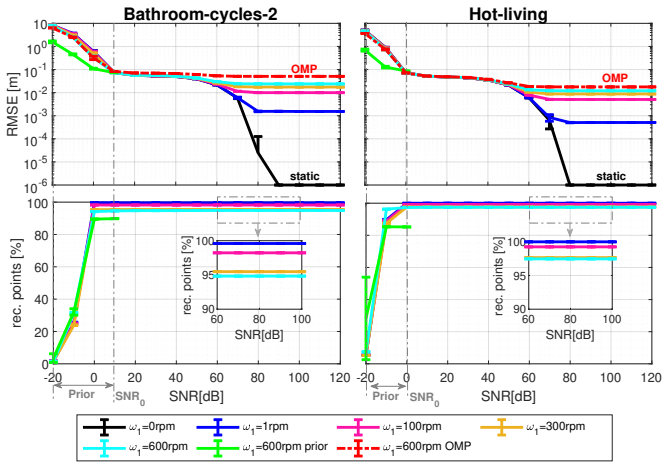

Fig. 5. Depth reconstruction errors (top) and percentage of recovered points (bottom) via our methodology and OMP for two scenes from [6].

of our sensing scheme, and that the inclusion of a depth prior improves the retrieval accuracy for very noisy signals. Prospective work includes the experimental evaluation of the proposed schemes, the generation of benchmark datasets for indoors and outdoors, and the exploration of strategies based on 3D dictionaries to exploit the sparsity in the 3D space.

REFERENCES

[1] 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.

[2] A. Lopez Paredes, M. Heredia Conde, and O. Loffeld, "CS-ToF sensing by means of greedy bi-lateral fusion and near-to-optimal low-density codes," in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 1996–2000.

[3] 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.

[4] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, pp. 2231 – 2242, 11 2004.

[5] C. D. Herrera, J. Kannala, P. Sturm, and J. Heikkila, "A learned joint depth and intensity prior using Markov Random Fields," in *2013 International Conference on 3D Vision - 3DV 2013*, 2013, pp. 17–24.

[6] F. Gutierrez-Barragan, H. Chen, M. Gupta, A. Velten, and J. Gu, "iToF2dToF: A robust and flexible representation for data-driven time-of-flight imaging," *arXiv preprint arXiv:2103.07087*, 2021.

[7] A. Sabov and J. Krüger, "Identification and correction of flying pixels in range camera data," in *Proceedings of the 24th Spring Conference on Computer Graphics*, ser. SCCG '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 135–142. [Online]. Available: https://doi.org/10.1145/1921264.1921293

[8] H. Sarbolandi, M. Plack, and A. Kolb, "Pulse based time-of-flight range sensing," *Sensors*, vol. 18, no. 6, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/6/1679

[9] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, "Near-infrared, depth, material: Towards a trimodal time-of-flight camera," in *2020 IEEE SENSORS*, 2020, pp. 1–4.

[10] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar, "Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles," *ACM Transactions on Graphics (TOG)*, vol. 32, 11 2013.

[11] 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