Fast Time-Domain Super-Resolution for Single-Shot Multi-Path ToF Imaging

Peyman F. Shahandashti*, P. López*, V.M. Brea*, D. García-Lesta*, Miguel Heredia Conde*†
*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Spain
†Center for Sensor Systems (ZESS), University of Siegen, Germany
E-mail: peyman.fayyaz@usc.es

***Abstract*—This paper presents a fast and simple multi-path depth retrieval method based on the Inverse Discrete Fourier Transform (IDFT) for Indirect Time of Flight (IToF) imaging capable of achieving super-resolution in the time domain from a limited number of samples in the frequency domain. To this aim, an *ad hoc* hardware architecture providing quasi-optimal Fourier measurements by means of a single-shot multifrequency ToF sensor concept is exploited. The proposed method has been validated for one and two bounces, showing that the system is able to recover the distance to both targets in one shot with 16 frequencies (2 -32 MHz).**

***Index Terms*—CMOS image sensor, depth image sensor, macro-pixel, multi-frequency, multi-path, single-shot, time-of-flight.**

I. Introduction

Among all the 3D image sensors, the ToF technique has arguably attracted the most commercial and scientific interest in the last years [1]–[3]. An interesting example is the rapid growth in the usage of IToF in mobile devices for various applications, including face recognition, auto-focusing, and augmented reality. Although a significant advancement in the resolution, measurement range, and precision of the IToF has been made, some challenges still have to be addressed. Multifrequency measurements can be used to solve some of these challenges, for instance, in phase ambiguity problems and Multi Path Interference (MPI) [4], [5]. MPI is a cross-disciplinary problem affecting systems working with different ranges of wavelengths, including radar, wireless communication, and optical systems. To overcome these problems, a high frame rate and real-time calculation of the raw data are required, which are computationally expensive and consume high power.

In our prior work, we presented a chip that can generate the type of measurements that parametric frequency estimation methods, such as the matrix pencil, need to solve the MPI problem in a closed-form manner. Unfortunately, applying this method sequentially to all pixels becomes time-consuming for large pixel arrays. Here, we present a simple alternative that can very efficiently solve the problem using the same type of measurements. We introduce a simple signal processing method based on IDFT that results in time-domain super-resolution from single-shot multifrequency raw images utilizing a dedicated hardware architecture. The multifrequency measurements are used to estimate the Fourier coefficients of the scene response function, which can be ideally modeled as a train of Dirac delta functions [4].

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. Coauthors of this work are funded by Spanish Ministry of Science, Innovation and Universities under grant RTI2018-097088-B-C32, from Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria Accreditation 2019-2022 ED431G-2019/04 and Reference Competitive Group Accreditation 2021-2024 ED431C2021/048, co-funded by (ERDF/FEDER programme).

The paper is organized as follows. Section II presents the hardware architecture that generates the Fourier measurements from single-shot multifrequency images, while the time-domain super-resolution method is described in Section III. The main results and comparison with a baseline approach are shown in Section IV.

II. Design of multi-frequency ToF system

The single-shot multifrequency IToF approach is based on the use of macro-pixels consisting of several subpixels. As shown in Fig. 1(a), each designed subpixel includes a $4\,\mu\text{m} \times 4\,\mu\text{m}$ pinned photodiode (PPD), two transfer gates (TX) to demodulate oppositely in a two-tap lock-in pixel structure, diode-connected anti-blooming transistors, global shutter transistors to synchronize the readout process in the pixel array, reset transistors, source followers (SF) as voltage buffers, and row select transistors to connect the output to the next stages sequentially [6]. The process starts with resetting the FD nodes to V_{res}. Then, during the integration time, the generated electrons will be accumulated in FD nodes, reducing their voltage depending on the received illumination intensity. Finally, the voltage of FDs will be read by turning on the row select transistors [7], [8].

In order to cope with multiple return paths, multiple demodulation frequency measurements are required. In this work, we consider, without loss of generality, a single macro-pixel consisting of an array of 4 by 4 subpixels, as shown in Fig. 1(b). The subpixels can demodulate with the same frequency (conventional IToF) or with different frequencies (single-shot multifrequency approach). In the multifrequency ToF system, each subpixel demodulates with frequencies that are multiples of a base frequency, f_1, namely $f_n = \{nf_1\}_{n=1}^{n=16}$, while the taps have a 180° phase shift. Connecting macro-pixels can shape the entire pixel array.

A measurement setup of the ToF 3D sensor is sketched in Fig. 1(c). It comprises a modulated light source, a target to reflect the light, and a ToF pixel array to demodulate

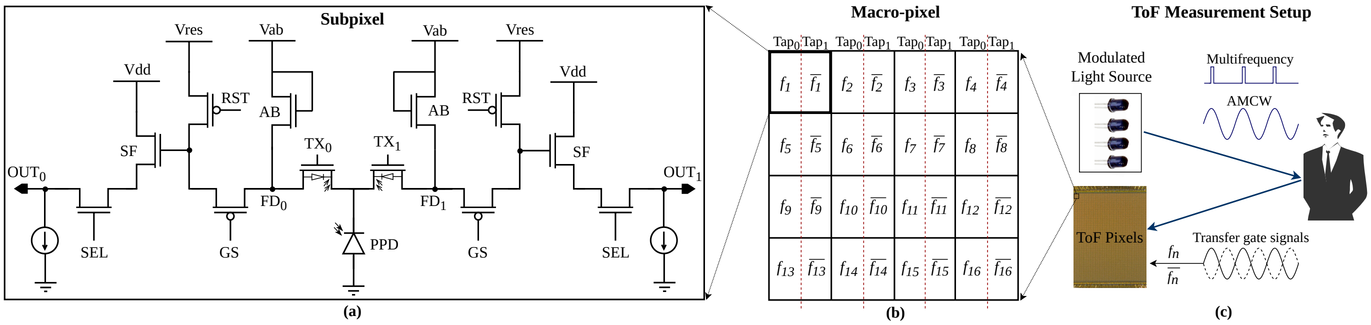

Fig. 1: Multi-frequency ToF system showing the schematic of the designed pixel (a), the macro-pixel consisting of 16 subpixels (b) that are controlled with $f_n = \{nf_1\}_{n=1}^{n=16}$, and modulation schemes in classical AMCW ToF and the proposed multifrequency measurement approach (c).

light echo. Most conventional ToF systems operate with sine wave modulation, called Amplitude Modulated Continuous Wave (AMCW). Generally, these systems are all homodyne, meaning that the pixels demodulate at the same frequency the light source is modulated. However, in the proposed single-shot multifrequency demodulation method, the illumination signal is a periodically-repeated ultrashort pulse with a low duty cycle, close to a Dirac delta function. Furthermore, as opposed to conventional AMCW ToF, the demodulation frequency is different for each subpixel. Under this approach, undistorted Fourier measurements corresponding to different frequencies can be generated in a single shot.

III. Depth retrieval method with multifrequency measurements

In the direct ToF technique, we can calculate a distance by measuring the time it takes for a light pulse to travel from an object to a sensor. Alternatively, in IToF, the phase delay of the reflected signal from a periodic signal source can be measured for each pixel in the sensor array, allowing a distance map of the scene to be calculated as,

$$d = \frac{c\phi}{4\pi f} = c\frac{t_d}{2}, \quad t_d = \frac{\phi}{2\pi f}, \tag{1}$$

where d is the object distance, c is the speed of light, ϕ is the phase shift, f is the modulation frequency, and t_d is the time delay that corresponds to the measured phase shift.

In the mono-frequency AMCW baseline approach, four samples of the cross-correlation function at different phase shifts are usually measured, and the phase shift is estimated using Four Bucket Method. In a two-tap structure, these measurements can be carried out in two frames, and the phase shift can be calculated through,

$$\phi = \arctan\left(\frac{m_3 - m_1}{m_0 - m_2}\right), \tag{2}$$

where m_0 to m_3 are the outputs of the pixel (OUT_0 and OUT_1) when applying modulating signals with the phase shifts of 0°, 90°, 180°, and 270°. Unfortunately, this method only works on mono-frequency measurements of a single bounce. It does not contemplate the multi-bounce case because in single frequency measurement, the bandwidth in the frequency domain is zero. Therefore, only one target is distinguishable with one frequency.

Different methodologies exist to resolve multi-target measurements [1]. Here, distance retrieval using IDFT is proposed. Estimating sparse targets from a set of multi-frequency measurements can be expressed as a spectral estimation problem. The fundamental concept of this method is that the most basic spectral estimation tool is the Fourier transform, so by leveraging it, the multifrequency estimation can be solved in the simplest and fastest way. The idea is to transform from the frequency domain, where the measurements from the 16 subpixels are located, to the time domain, where the scene response function can be retrieved, and this can be attempted utilizing an inverse Fourier transform. The IDFT is the sum of weighted complex sinusoids where the weights are given by the signal that should be transformed,

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j2\pi kn/N}, \tag{3}$$

where n is a running index, and N is the number of steps in both domains. The number of measurements in the frequency domain is $(2 \times P) + 1$, where P is the number of frequency measurements, $P = 16$ in this realization, the factor 2 accounts for positive and negative frequencies, and one is for the zero frequency. The measurements $X(k)$, for the frequency k, can be formed as $X(k) = (m_3(k) - m_1(k)) + j(m_0(k) - m_2(k))$, where $m_0(k)$ to $m_3(k)$ are exactly the same as values in (2), but for 16 different frequencies. The $X(k)$ for negative frequencies can be formed by conjugating the measurements for positive frequencies.

To reach a much finer number of discrete steps in the time domain than the typical IDFT calculation, the idea is to increase the number of N in the exponent of (3) (normalization factor). Therefore, more points in the target domain (time domain in this case) can be generated. By doing this, the number of samples in the time domain will be $N' = K \times N$, where K is the “super-resolution factor,” i. e., the ratio between the number of measurements in the time domain and the num-

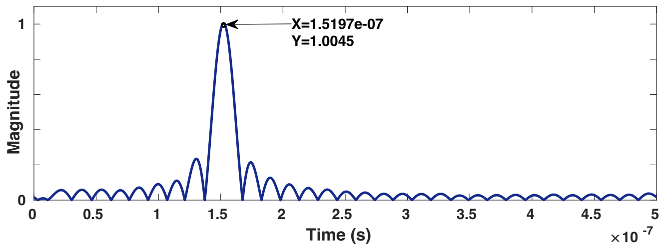

Fig. 2: Magnitude of the IDFT for the single-bounce test.

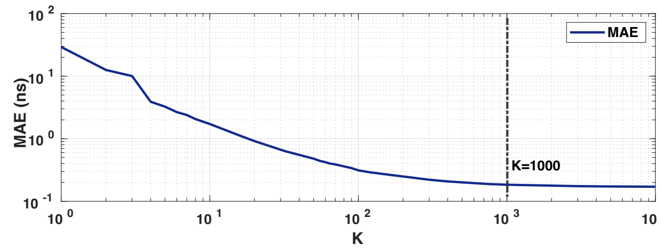

Fig. 3: MAE of the estimated time delay versus factor K.

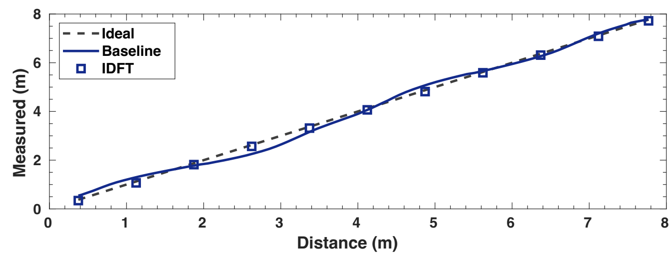

Fig. 4: The retrieved distance with baseline and IDFT methods.

ber of measurements in the frequency domain. The optimum value of K is the value beyond which the accuracy no longer improves. At this point, the result is no longer limited by the grid resolution, and it is futile to refine the temporal grid any further. The magnitude of the $x(n)$ in (3) will show a single peak in the time domain (in the case of a single bounce), and this single peak will be located precisely where the Dirac delta function was in the scene response function. Since time and frequency are interrelated in a Fourier transform, the time delay value corresponding to the location of the detected peak in the time domain can be calculated. The time axis can be entirely determined by the inverse of the base frequency step, so the time step in the target domain for a given K can be deduced. The distance can be directly recovered from (1) by estimating the time delay.

IV. Results and Discussion

In order to evaluate the performance of the IDFT method with the multifrequency demodulation ToF hardware architecture, two sets of experiments were performed corresponding to single-bounce and double-bounce scenarios. The simulations were performed with the setup shown in Fig. 1, a macro-pixel including 16 subpixels. The demodulation frequency of subpixels increases with a 2 MHz step, starting from 2 MHz, up to 32 MHz for the last subpixel.

In the first test, an impulse illumination signal with a repetition frequency of $f_1 = 2\,\text{MHz}$ and pulse-width of 5 ns (1% duty cycle) is considered, and OUT_0 and OUT_1 are measured in two consecutive frames for the 16 subpixels. Then, the IDFT is performed on the phasors obtained for the measured frequencies, and the phase and magnitude values in the time domain are extracted. Fig. 2 shows the magnitude of the IDFT obtained when the received illumination signal is set to have a 150 ns time delay. As explained in Section III, the highest peak in the magnitude diagram of the time domain corresponds to the time delay, from which we can retrieve the distance. As it can be seen, the method can recover the time delay with reasonable accuracy.

To find the optimum value in terms of accuracy for the super-resolution factor, we have evaluated the errors for different values of K. Fig. 3 illustrates the Mean Absolute Error (MAE) of the retrieved time delays with respect to K. The MAE is calculated over 100 samples for each K value. As expected, the error diminishes with increasing K. However, higher values of K require more processing and consume more time. A reasonable trade-off is selecting $K = 1000$ in this case, at which accuracy does not experience further significant improvement.

The baseline approach has two differences with respect to the IDFT method. First, the illumination signal is a sine wave (AMCW in Fig. 1) with the same frequency at which all the subpixels demodulate ($f_n = \{f_1\}_{n=1}^{n=16}$), which is 20 MHz in this case (among the frequency range of the multifrequency setup). The baseline method only measures the phase shifts in a single-frequency in one shot. Second, the distance retrieval method is based on the "Four Bucket Method," explained in Section III. Fig. 4 compares the estimated distances between the baseline and the IDFT approach in the designed ToF system. From Fig. 4, it is clear that leveraging measurements at different frequencies allows for annihilating the wiggling effect, as one would expect, as harmonic frequencies contribute to the estimation, instead of disturbing it like under the classical mono-frequency hypothesis. The multifrequency macro-pixel architecture trades spatial resolution for frame rate, enabling single-shot operation.

The same simulations are performed in the second test for the two-bounces case, where the pixels capture two signals simultaneously. The simulation setup is the same as in the first test, except that the illumination signal is changed to a superposition of two bounces. Additionally, the amplitude of the second illumination signal is considered $1/8$ of the first illumination signal for realism. Fig. 5 illustrates the magnitude of the IDFT obtained from the 16 measurements when the time delay of the first and second targets are set to 45 ns and 110 ns, respectively. The first and second peaks in the magnitude diagram in the time domain are equivalent to the time delays of the first and second bounces. The method allows for single-shot two-bounce distance retrieval, as opposed to conventional mono-frequency approaches, which can only deal

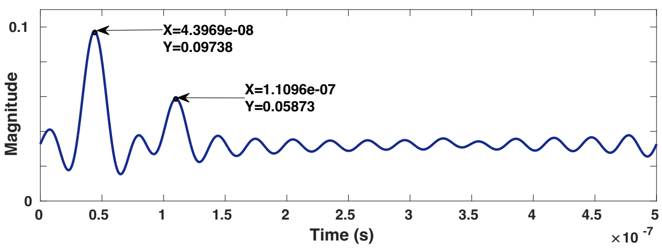

Fig. 5: Magnitude of the IDFT for the two bounces test.

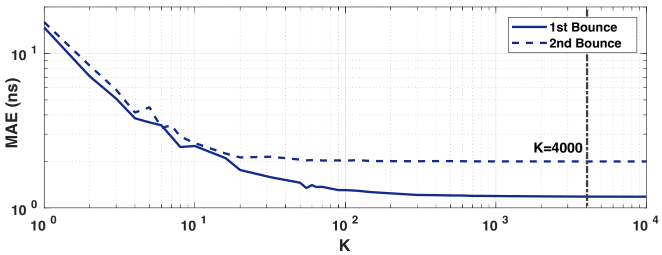

Fig. 6: MAE of the estimated time delay with respect to K for the two bounces test.

with one-bounce situations.

As in the first test, the effect of the K factor on the error is studied for the two-bounce case in Fig. 6. This time the errors of the time delays for the first and second bounces are considered. The time delay is measured for 10 samples in each value, and the MAE is calculated. Unsurprisingly, similar to the single-bounce, the error is reduced by increasing the factor K. Here a value of $K = 4000$ is selected as a reasonable trade-off for the IDFT operation.

Fig. 7 demonstrates the functionality of the IDFT method in recovering distance in the two-bounce case. The simulations are performed assuming that the first target is fixed at 1 m and displacing the other closer until they no longer can be distinguished, when a single main peak instead of two peaks will appear in the magnitude of IDFT. The minimum distance that can be resolved is 4.7 m in 2 MHz. As can be seen, both distances are estimated when the second target is located in the range of 5 m to 11.5 m.

Table I highlights a comparison between computation time and the accuracy of the IDFT and matrix pencil methods. The computation time of the super-resolved IDFT method scales linearly with the number of paths, since only detecting the peaks scales in the method, while the Fourier transform computation does not experience a time increase. Also, the distance could be retrieved faster than the matrix pencil, even for very large K, up to 500, and deliver comparable accuracy.

V. Conclusion

This work presents a fast, straightforward method for single-shot multifrequency ToF imaging that efficiently solves the MPI problem. The proposed hardware architecture allows us to apply 16 different frequencies to a macro-pixel that implements single-shot multifrequency demodulation. In a standard

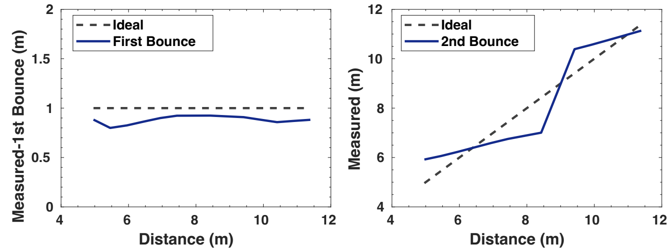

Fig. 7: The retrieved distance of the first (left) and second (right) targets with IDFT method.

TABLE I: A comparison of computation time and distance RMSE values between IDFT and matrix pencil methods.

Method		Computation Time (ms)			Distance RMSE (cm)		
	# Paths	K=10	K=100	K=1000	K=10	K=100	K=1000
IDFT	1	0.890	4.16	46.61	30	4.5	2.7
	2	1.67	6.24	73.24	31.7	15.4	13.5
Matrix Pencil	1		16.16			1.7	
	2		48.61			13	

time-gating approach, a resolution of $1/N$ times the domain size can be obtained from N measurements. A conventional IDFT would equally provide a resolution of $1/N$ of the depth range, which is way too low, e.g., 4.7 m for a 2 MHz base frequency. Instead, in this method, by using the idea of super-resolution from multifrequency measurements, it is feasible to go down to a resolution of $1/(K \times N)$, where K is the desired super-resolution factor, which is only limited by the SNR of the measurements and model mismatches. Compared to the classical method, this approach can resolve multiple paths with custom trade-off between computation time and resolution through the parameter K, enabling real-time multi-path depth imaging.

References

[1] A. Bhandari and R. Raskar, “Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-d imaging,” *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 45–58, 2016.

[2] M. Heredia Conde, *Compressive sensing for the photonic mixer device.* Springer, 2017.

[3] A. Bhandari, M. Heredia Conde, and O. Loffeld, “One-Bit Time-Resolved Imaging,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 1630–1641, 2020.

[4] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, “Fast multipath estimation for PMD sensors,” in *5th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar, and Remote Sensing (CoSeRa 2018)*, 2018.

[5] F. Chen, R. Ying, J. Xue, F. Wen, and P. Liu, “A configurable and real-time multi-frequency 3d image signal processor for indirect time-of-flight sensors,” *IEEE Sensors Journal*, vol. 22, no. 8, pp. 7834–7845, 2022.

[6] D. Stoppa and et al., “A Range Image Sensor Based on 10-μm Lock-In Pixels in 0.18-μ m CMOS Imaging Technology,” *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, 2010.

[7] P. F. Shahandashti and et al., “Proposal of a Single-Shot Multi-Frame Multi-Frequency CMOS ToF Sensor,” in *2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*. IEEE, 2021, pp. 1–4.

[8] P. F. Shahandashti and et al, “A 2-Tap Macro-Pixel-Based Indirect ToF CMOS Image Sensor for Multi-Frequency Demodulation,” in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2022, pp. 1–5.