

Fourier Analysis on Transient Imaging with a Multifrequency Time-of-Flight Camera

Jingyu Lin^{*†}, Yebin Liu^{*}, Matthias B. Hullin[‡], and Qionghai Dai^{*}

^{*}Beijing Key Laboratory of Multi-dimensional & Multi-scale Computational Photography (MMCP),
Tsinghua National Laboratory for Information Science and Technology,
Department of Automation, Tsinghua University, Beijing 100084, China

[†]College of Electrical Engineering, Guangxi University, Nanning 530004, China

[‡]Institute of Computer Science II, University of Bonn, 53113 Bonn, Germany

Abstract—A transient image is the optical impulse response of a scene which visualizes light propagation during an ultra-short time interval. In this paper we discover that the data captured by a multifrequency time-of-flight (ToF) camera is the Fourier transform of a transient image, and identify the sources of systematic error. Based on the discovery we propose a novel framework of frequency-domain transient imaging, as well as algorithms to remove systematic error. The whole process of our approach is of much lower computational cost, especially lower memory usage, than Heide et al.’s approach using the same device. We evaluate our approach on both synthetic and real-datasets.

Keywords-transient imaging; Fourier analysis; time-of-flight camera; multifrequency

I. INTRODUCTION

Transient imaging is a family of techniques that aim to capture non-stationary distributions of light. Starting with early efforts in optics literature [1], this direction of research more recently captured attention from researchers in the field of computer vision and graphics. In this work, we follow the definition in [2] and denote by a *transient image* a time-image sequence $i(x, y, t)$ that represents an optical impulse response of a scene at a high enough temporal resolution to observe light “in flight” before its distribution in the scene achieves a global equilibrium. This concept breaks with the longstanding (and in most cases, reasonable) assumption in graphics and vision that the speed of light is infinite. Recent research has given rise to many exciting applications of ultrafast time-resolved measurements, including scene reflectance capture [3], looking around corners [4], [5] and bare sensor imaging [6]. The state of the art in terms of imaging quality is currently held by high-end systems consisting of a femtosecond laser and a streak camera [7]. These systems directly sample the time dimension and achieve a temporal resolution of about 2 picoseconds per frame. However, they are prohibitively expensive for many laboratories, fragile, complex to operate, slow and extremely sensitive to ambient light.

To lower the barrier of transient imaging, Heide et al. [8], [9] demonstrated a compact acquisition system on a much

smaller budget without ultrafast light sources and detectors, and successfully reconstruct transient images by a computational technique. Their system is a modified inexpensive time-of-flight (ToF) camera. A transient image of a scene is reconstructed from a collection of images of the scene captured by their system operating on hundreds of different modulation frequencies. Their reconstruction is formulated as a linear inverse problem solved by numerical optimization, which makes the ill-condition and noise problem difficult to be analyzed.

In this paper, we discover that ToF camera based transient imaging is essentially a frequency-domain sampling technique, compared with the time-domain sampling technique using a femtosecond laser and a streak camera [7]. Taking advantage of the frequency-domain sampling, we introduce the Fourier analysis for a better understanding and reconstruction of the transient image. Based on our Fourier analysis, we identify the sources of the systematic errors, including the non-sinusoidal and frequency-variant modulation signal, and the limited working frequency range. To resolve these systematic errors, a frequency-domain transient image reconstruction approach works in a pixel-wise manner is proposed.

The proposed reconstruction approach has been tested on both synthetic and real data and yields good results. Compared with the approach in [8], our approach takes full advantage of the intrinsic characteristic of the multifrequency ToF system and is fast and memory efficient. The source code of this work is open to public. We believe that this work will contribute to a better understanding of ToF transient imaging systems in both acquisition and reconstruction.

II. RELATED WORK

A ToF camera is a scannerless range imaging system that resolves the depths of the entire scene simultaneously with each laser or light pulse, as opposed to scanning LIDAR systems using point-by-point laser beam. Such a realtime depth sensor simplifies many computer vision tasks and enables convenient solutions [10] for shape reconstruction [11], motion capture [12], gesture recognition [13], etc.

To enable high-quality depth sensing for the above applications, one of the main challenges is the multipath interference (MPI) problem [14]. MPI refers to false depth measurement due to optical superposition of multiple light paths—global illumination in space and time. MPI inversion and the reconstruction of transient images are closely related but have different goals. The latter aims at resolving the amount of light bouncing back from the surface as a function of time, while the former tries to remove all the indirect light components bouncing back from the depth surface and extract only the direct light component (which typically also corresponds to the shortest possible path).

Fuchs et al. [15] and Jimenez et al. [16] resolve diffuse MPI with Lambertian scene assumption using only one modulation frequency and solving with high computational iterative optimization. Dorrington et al. [17] models two-path interference arising from specular surfaces with two-frequency measurements using numerical optimization. Godbaz et al. [18] uses 3 or 4 modulation frequencies and proposes a close-form solution while Kirmani et al. [19] uses 5 frequencies. Both of them mitigate MPI under specular scene assumption. With multifrequency sampling, MPI under more general scene assumption is recently investigated by Freedman et al. [20] and Bhandari et al. [21]. They achieve real-time resolving of MPI in these works. However, both of them use sparse reflection regularizers which limits their application scenario. The sparsity assumption has also recently been adopted in transient image reconstruction [22]. Our technique does not enforce such sparsity constraints but requires more frequencies.

Our work is also related with frequency-domain computing of one-dimensional signals under multifrequency sampling frameworks. In remote sensing, Simpson et al. [23] introduce an Amplitude Modulated Continuous Wave (AM-CW) technique with discrete stepped frequency in Lidar system and use inverse Fourier transform to recover the scattering function of the environment across time. In fluorescence lifetime imaging microscopy (FLIM) [24], time-resolved fluorescence light can be captured and reconstructed by either time domain techniques [25] or frequency-domain techniques [26]. The relationship between time- and frequency-domain FLIM is similar to time-domain transient imaging using a streak camera [7] and our proposed frequency-domain technique using a ToF camera.

III. FOURIER ANALYSIS ON TRANSIENT IMAGING

In this section we show that the transient image and the data acquired by a multifrequency ToF camera are related through the Fourier transform. Based on this discovery, we propose a framework of frequency-domain transient imaging. Throughout this paper, we use lower-case letters for time-domain signals and the corresponding upper-case letters for their Fourier counterparts.

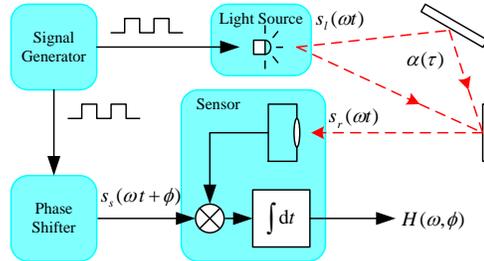


Figure 1. Operating principle of the transient imaging system.

A. Multifrequency ToF Camera

The referred transient imaging system is a ToF camera consisting of a light source and an optical sensor. Both units are modulated by the same periodic signal generator that can be operated across a wide frequency range.

Fig. 1 illustrates the working principle of the camera. The output of the light source is a periodic wave denoted by $s_l(\omega t)$, where $\omega = 2\pi f$ and f is the modulation frequency. Light is transmitted to the scene and bounces back to the sensor along different paths. When reaching the sensor the light signal along a single ray path is delayed by τ and attenuated by α . The delay time τ and the attenuation coefficient α are uniquely determined by the ray path. The photons along all ray paths are superimposed at the sensor end, resulting in the light signal received by the sensor represented as

$$s_r(\omega t) = E_0 + \int_0^\infty \alpha(\tau) s_l(\omega(t - \tau)) d\tau, \quad (1)$$

where E_0 is environmental illumination. The second term on the right hand side of (1) is the convolution of the input $s_l(\omega t)$ with a scene-related term $\alpha(t)$, namely the impulse response of the scene at a point. The imaging system simultaneously acquires this response across a 2D sensor plane. The temporal sequences of all the points constitute a sequence of images, $\alpha_{(x,y)}(t)$, known as a *transient image* $i(x, y, t)$. In this paper, we do not consider any relation between adjacent pixels. Therefore, we omit the pixel coordinates (x, y) in all symbols throughout this paper, and $\alpha(t)$ is the desired transient image.

At the sensor the received light signal $s_r(\omega t)$ carrying the transient information of the scene is integrated over an exposure time of NT , where N is an integer and $T = 1/f$ is the period of the modulation signal. The sensor gain is modulated by a zero-mean signal $s_s(\omega t + \phi)$, where ϕ is a programmable phase offset with respect to the light source signal $s_l(\omega t)$. Thus the acquired image is

$$H(\omega, \phi) = \int_0^{NT} s_r(\omega t) s_s(\omega t + \phi) dt. \quad (2)$$

Substituting (1) into (2) and noting that the integration of $E_0 s_s(\omega t + \phi)$ over a period is zero, we have (see [8])

$$H(\omega, \phi) = \int_0^\infty \alpha(\tau) c(\tau, \omega, \phi) d\tau, \quad (3)$$

$$c(\tau, \omega, \phi) = \int_0^{NT} s_l(\omega(t - \tau)) s_s(\omega t + \phi) dt, \quad (4)$$

where $c(\tau, \omega, \phi)$ is the scene-independent *correlation function* between $s_l(\omega t)$ and $s_s(\omega t)$, and the working frequency ω and the phase offset ϕ are both programmable parameters. The discrete version of the correlation function is a *correlation matrix*. Since the exact shapes of $s_l(\omega t)$ and $s_s(\omega t)$ of a real system are unknown, the correlation function cannot be computed from (4) and should be obtained by a calibration process.

Given the correlation matrix $c(\tau, \omega, \phi)$ and the acquired image collection $H(\omega, \phi)$ by a multifrequency ToF camera working at a group of different frequencies and phases, the transient image $\alpha(\tau)$ can be reconstructed by solving the discrete version of (3). Heide et al. [8] optimize on $\alpha(\tau)$ by imposing spatial and temporal priors and even surface model constraints and solving a massive linear optimization problem. Kadambi et al. [22] design temporal illumination codes to make the correlation matrix invertible under the assumption that scene response is sparse, and then deconvolve the transient image on the acquired temporal sequence. Both approaches operate in time domain.

B. Fourier Analysis on Ideal Case

To better present our idea, we start from an ideal case that both the light source and the sensor gain are sine waves. It can be derived from (4) that their correlation function is also a sine wave

$$c(\tau, \omega, \phi) = A \cos(\omega\tau + \phi), \quad (5)$$

where A is a constant amplitude. We define the *complex correlation function* and the *complex image collection* as

$$\tilde{c}(\tau, \omega) = c(\tau, \omega, 0) + i \cdot c(\tau, \omega, \pi/2) \quad (6)$$

$$\text{and } \tilde{H}(\omega) = H(\omega, 0) + i \cdot H(\omega, \pi/2), \quad (7)$$

respectively. From (3) and (5)-(7) we have

$$\begin{aligned} \tilde{H}(\omega) &= \int_0^\infty \alpha(\tau) \tilde{c}(\tau, \omega) d\tau \\ &= A \int_0^\infty \alpha(\tau) (\cos(\omega\tau) - i \sin(\omega\tau)) d\tau \\ &= A \int_0^\infty \alpha(\tau) e^{-i\omega\tau} d\tau \\ &= A \cdot \mathcal{F}[\alpha(\tau)]. \end{aligned} \quad (8)$$

Here, we reach the key conclusion from (8) that the acquired complex image collection $\tilde{H}(\omega)$ is the Fourier transform of a transient image $\alpha(\tau)$. If $\tilde{H}(\omega)$ is acquired with working frequency across the whole frequency spectrum of $\alpha(\tau)$, exact $\alpha(\tau)$ can be reconstructed by implementing the inverse Fourier transform on $\tilde{H}(\omega)$. Compared with an ultrafast camera as in [7] which samples the optical response of a scene in the time domain, a multifrequency ToF camera samples the optical response of a scene in the frequency domain.

C. Extended Analysis on Non-ideal Case

For a non-ideal ToF camera we consider potential defects as follows: 1) the modulation signal is periodic but not sinusoidal; 2) the waveform of the modulation signal varies with frequency; 3) the modulation frequency is only available in a limited range from a low frequency f_L to a high frequency f_H . In this subsection we investigate systematic error in transient image reconstruction introduced by these problems.

Since the light source and the sensor gain are periodic functions, the correlation function is also a periodic function of the same frequency, such that it can be expanded into a Fourier series, i.e.,

$$c(\tau, \omega, \phi) = \sum_{n=\pm 1}^{\pm \infty} \tilde{A}_n(\omega) e^{-i(n\omega\tau + n\phi)}, \quad (9)$$

where $\tilde{A}_n(\omega)$ are complex coefficients of the Fourier series and $\sum_{n=\pm 1}^{\pm \infty} = \sum_{n=-\infty}^{-1} + \sum_{n=1}^{+\infty}$. The DC component in (9) is zero since the sensor gain $s_s(\omega t)$ is a zero-mean function. Then the complex correlation function (6) becomes

$$\tilde{c}(\tau, \omega) = \sum_{n=\pm 1}^{\pm \infty} \tilde{B}_n(\omega) e^{-in\omega\tau}, \quad (10)$$

$$\tilde{B}_n(\omega) = \tilde{A}_n(\omega) (1 + i e^{-in\pi/2}), \quad (11)$$

and the complex image collection is

$$\begin{aligned} \tilde{H}(\omega) &= \int_0^\infty \alpha(\tau) \tilde{c}(\tau, \omega) d\tau \\ &= \sum_{n=\pm 1}^{\pm \infty} \tilde{B}_n(\omega) \int_0^\infty \alpha(\tau) e^{-in\omega\tau} d\tau \\ &= \sum_{n=\pm 1}^{\pm \infty} \frac{\tilde{B}_n(\omega)}{n} \int_0^\infty \alpha\left(\frac{\tau}{n}\right) e^{-i\omega\tau} d\tau. \end{aligned} \quad (12)$$

Since the waveform of the modulation signal varies with frequency, the coefficient \tilde{B}_n in $\tilde{H}(\omega)$ vary with ω . Note that the integration part on the right hand side of (12) is the Fourier transform of $\alpha(\tau/n)$, the inverse Fourier transform on $\tilde{H}(\omega)$ can be obtained by the convolution property of the Fourier transform:

$$\tilde{h}(\tau) = \mathcal{F}^{-1}[\tilde{H}(\omega)] = \sum_{n=\pm 1}^{\pm \infty} \alpha\left(\frac{\tau}{n}\right) * b_n(\tau), \quad (13)$$

where $*$ denotes convolution operator and

$$b_n(\tau) = \mathcal{F}^{-1}[\tilde{B}_n(\omega)/n] \quad (14)$$

is a window function. Since $\tilde{B}_{-1}(\omega) = 0$, (13) can be rewritten as

$$\tilde{h}(\tau) = \alpha(\tau) * b_1(\tau) + \sum_{n=\pm 2}^{\pm \infty} \alpha\left(\frac{\tau}{n}\right) * b_n(\tau). \quad (15)$$

At the right hand side of (15), the term that contains the transient image $\alpha(\tau)$ is called *fundamental component* ($n = 1$), and the other terms contain dilated versions $\alpha(\tau/n)$ are called *dilated components* ($|n| > 1$). To extract the transient

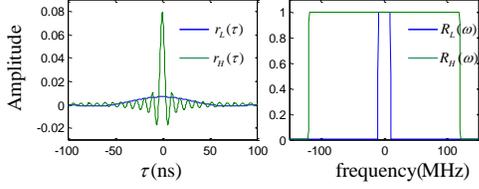


Figure 2. Rectangular frequency window functions $r_L(\tau)$ and $r_H(\tau)$.

image $\alpha(\tau)$ from $\tilde{h}(\tau)$, one needs to remove the dilated components and then invert the window function $b_1(\tau)$.

Moreover, since the working frequency ω can only be available in a limited range $[f_L, f_H]$, the frequency spectra of $\tilde{h}(\tau)$ beyond $[f_L, f_H]$ are missing. This is equivalent to impose a windowing function on $\tilde{h}(\tau)$, i.e.

$$\tilde{h}_c(\tau) = \tilde{h}(\tau) * [r_H(\tau) - r_L(\tau)], \quad (16)$$

where $r_H(\tau)$ and $r_L(\tau)$ are rectangular frequency windows cutting off at f_H and f_L , respectively. Fig. 2 shows the shapes of the two window functions in the time domain and in the frequency domain.

In summary, the imaging model represented by (12), (15) and (16) identifies three causes of systematic errors in multifrequency ToF transient imaging:

- Non-sinusoidal shapes of modulation signal introduce harmonic components in the correlation function (9) and consequently cause a transient image $\alpha(\tau)$ mixed with its dilated versions $\alpha(\tau/n)$ in (15).
- Frequency-varying shapes of the modulation signal impose a frequency window function $b_1(\tau)$ on the transient image $\alpha(\tau)$ such that the amplitude and the phase of frequency spectrum of $\alpha(\tau)$ are distorted in (15).
- Limited working frequency range of the modulation signal results in frequency spectrum truncation of the transient image $\alpha(\tau)$ in (16).

D. Framework of Fourier Domain Transient Imaging

Based on the above analysis, we propose a framework of transient imaging in the frequency domain, as illustrated in Fig. 3. First we use a multifrequency ToF camera to acquire image set $H(\omega, \phi)$ at a group of frequencies and phase offsets. The working frequency is specified in available frequency range of the camera. Then the image set $H(\omega, \phi)$ is rectified to remove the effect of the frequency window function based on the knowledge of the camera correlation function precomputed through a camera calibration process. Finally, we reconstruct the transient image by taking the inverse Fourier transform followed with the removal of the dilated components and the compensation for the low frequency spectrum.

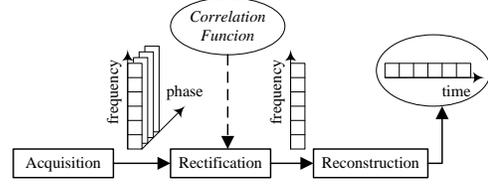


Figure 3. Framework of frequency domain transient imaging. The correlation function is obtained through a calibration process.

IV. DATA RECTIFICATION

We rectify the distorted amplitude and phase of the acquired data by computing

$$\tilde{R}(\omega) = \tilde{H}(\omega) / \tilde{B}_1(\omega), \omega \in [2\pi f_L, 2\pi f_H]. \quad (17)$$

It can be derived from (12) that, by ignoring the dilated components we have

$$\tilde{R}(\omega) = \int_0^\infty \alpha(\tau) e^{-i\omega\tau} d\tau. \quad (18)$$

Then taking the inverse Fourier transform on $\tilde{R}(\omega)$ yields the transient image $\alpha(\tau)$ mixed with the dilated component.

In (17), the parameter $\tilde{B}_1(\omega)$ can be computed by $\tilde{B}_1(\omega) = 2\tilde{A}_1(\omega)$ (see (11)), where $\tilde{A}_1(\omega)$ is the correlation function obtained through the following calibration process.

Camera Calibration. We install the light source and the camera closely, and face them towards a diffuse white board. We capture calibration data at selected flying times τ and frequencies ω and use the following *parameterized correlation function*

$$A_0(\omega) + \sum_{n=1}^{n_0} A_n(\omega) \exp\{-i[n\omega(\tau + \tau_0) - \phi_n(\omega)]\} \quad (19)$$

to fit the captured calibration data. Here,

$$\tilde{A}_n(\omega) = A_n(\omega) \exp[\phi_n(\omega)]. \quad (20)$$

We use a Fourier series with a small number of components (for example $n_0 = 4$), since in general cases the fundamental component in the Fourier series of a periodic function is dominant. Moreover, in (13) and (14) the dilated components are suppressed by $1/n$, such that the high order harmonic components have little effect on the reconstruction results.

This model has four parameters: τ_0 , $A_0(\omega)$, $A_n(\omega)$, and $\phi_n(\omega)$. τ_0 is the flying time of the shortest ray path. It corresponds to a reference location used to determine object depth. $A_0(\omega)$ is a complex function denoting unfiltered DC component, and has no further use. $A_n(\omega)$ and $\phi_n(\omega)$ are real functions denoting amplitudes and phase offsets of corresponding components, respectively.

We employ the gradient descent method to fit the calibration data from [8]. Fig. 4 shows that our model is well fit using the proposed method. The amplitudes of the fundamental component and the harmonic components are compared in Fig. 5. It is obvious that the fundamental component is predominant over the others, and its amplitude declines when

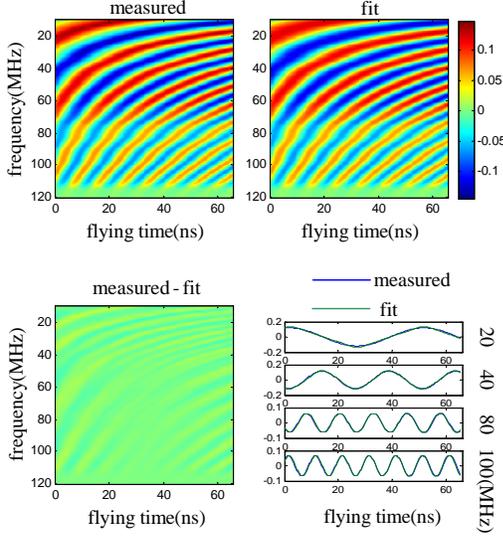


Figure 4. Correlation matrices. Top left: measured data from [8]. Top right: the fitting result from our parameterized complex correlation function after fitting process. Bottom left: error between measured data and our fitting result. Bottom right: comparison of measured data and fitting result at some frequencies.

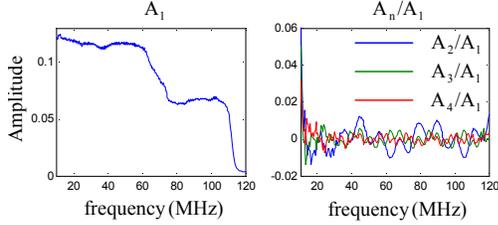


Figure 5. Amplitude of the fundamental component and the harmonic components of a correlation function. Left: the fundamental component A_1 declines with the frequency ω . Right: The ratios between the harmonic components and A_1 are very small fractions.

the frequency increases. Although the calibration step is time-consuming, it needs to be performed only once for an imaging system.

V. TRANSIENT IMAGE RECONSTRUCTION

The core of transient image reconstruction is the inverse Fourier transform on the rectified data. Since only samples in limited working frequency range are obtained, the inverse discrete Fourier transform (DFT) is

$$\beta(\tau) = \frac{1}{M} \sum_{\omega=\pm 2\pi f_L}^{\pm 2\pi f_H} \tilde{R}(\omega) e^{i\omega\tau}, \quad (21)$$

where $M = 1/f_s/\tau_s$ is the number of points in FFT, f_s frequency step, τ_s the flying time resolution, and $\tau > \tau_0$.

It can be derived from (16) and (17) that what we obtain in (21) is

$$\beta(\tau) = \tilde{h}(\tau) * b_1^{-1}(\tau) * [r_H(\tau) - r_L(\tau)], \quad (22)$$

where b_1^{-1} is the inverse of b_1 . Then by substituting (15) into (22) we have

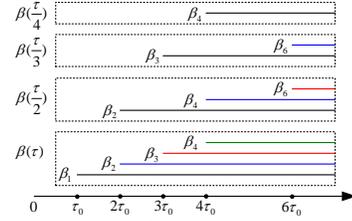


Figure 6. Demonstration of fundamental component extraction. After dilation, $\beta(\tau/2)$ contains components of $\beta_{2n}(\tau)$ and $\beta(\tau/m)$ contains components of $\beta_{mn}(\tau)$. Note that $\beta_m(\tau) = 0$ for $\tau < m\tau_0$. We can remove $\beta_m(\tau)$, $\tau < 2m\tau_0$ from $\beta(\tau)$ by subtracting $\beta(\tau/m)$ from $\beta(\tau)$.

$$\beta(\tau) = \beta_1(\tau) + \sum_{n=\pm 2}^{\pm \infty} \beta_n(\tau) * [b_n(\tau) * b_1^{-1}(\tau)], \quad (23)$$

where we define

$$\beta_n(\tau) = \alpha(\tau/n) * [r_H(\tau) - r_L(\tau)]. \quad (24)$$

To obtain $\alpha(\tau)$, we first obtain $\beta_1(\tau)$ by removing the dilated components and then recover missing frequency spectra.

A. Dilated Component Removal

We remove the dilated components based on the observation: if the shortest flying time of a scene is τ_0 , then $\alpha(\tau/n) = 0$ for $\tau \in [0, n\tau_0)$, such that $\beta(\tau)$, $\tau \in [0, 2\tau_0)$ contains only the fundamental component and nothing else.

As shown in Fig. 6, if we dilate $\beta(\tau)$ to $\beta(\tau/m)$, $m=1,2,\dots$ its components are also dilated, such that we can remove $\beta_m(\tau)$, $\tau < 2m\tau_0$ from $\beta(\tau)$ by subtracting $\beta(\tau/m)$ from $\beta(\tau)$. The remaining components can be removed iteratively in a similar way. Algorithm 1 summarizes this process.

Algorithm 1: Transient image reconstruction by Fourier analysis (refer to Fig. 6)

Input: Rectified data $\tilde{R}(\omega)$, working frequency range $[f_L, f_H]$ and step f_s , flying time resolution τ_s , parameters of complex correlation function τ_0 , $A_n(\omega)$, and $\phi_n(\omega)$, number of components n_0 .

Output: $\beta_1(\tau)$.

Steps:

- 1: Compute $b_n(\tau)$ by (14).
 - 2: Compute $\beta(\tau)$, $\tau > \tau_0$ by (21).
 - 3: Interpolate $\beta(\tau/m)$, $\tau \in [m\tau_0, 2m\tau_0]$.
 - 4: $\beta(\tau) = \beta(\tau) - \sum_{m=2}^{n_0} \beta(\tau/m) * [b_m(\tau) * b_1^{-1}(\tau)]$.
 - 5: Update $\beta(\tau)$ in a similar way as 3 to 4.
 - 6: $\beta_1(\tau) = \beta(\tau)$.
-

B. Frequency Spectra Recovery

Missing high frequency spectrum results in transient image being blurred in the time dimension. The best way to prevent high frequency spectrum from being truncated is increasing the highest working frequency of the ToF camera and acquiring data at higher frequencies.

As for the low frequency spectrum, we propose *peak boosting* algorithm to recover missing low frequency spectrum in $[0, \omega_1]$. By ignoring r_H , (24) can be simplified to

$$\beta_1(\tau) = \alpha(\tau) - \alpha(\tau) * r_L(\tau). \quad (25)$$

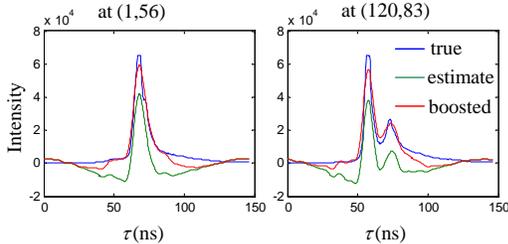


Figure 7. Demonstration of effect of our peak boosting algorithm. Estimate of the ground truth sinks down due to missing low frequency spectrum. Our peak boosting algorithm locally boost it close to the ground truth.

The second term is the missing low frequency spectrum part, which needs to be recovered from $\beta_1(\tau)$ and added to $\beta_1(\tau)$. Fig. 2 shows that the convolution of a peak with $r_L(\tau)$ is flat and centered at the peak. Therefore, the missing part $\alpha(\tau) * r_L(\tau)$ makes each peak in $\alpha(\tau)$ sink down locally. Our idea is finding a peak in $\beta_1(\tau)$, convoluting the peak with $r_L(\tau)$, and adding to $\beta_1(\tau)$. In this way, the peak and its neighborhood will be boosted. By repeating the steps, the missing low frequency spectrum can be recovered. Algorithm 2 summarizes the process.

More explanation about Algorithm 2 is given as follows. Since $\alpha > 0$, we have $\beta_1 < \alpha$, consequently $\beta_{pk} < \alpha$ since β_{pk} is the positive part of β_1 (θ is a small value used to prevent noise from boosting). In Step 3 adding $\beta_{pk} * r_L$ to yield β_b such that β_b approximates α but never over boosted since $\beta_{pk} < \alpha$.

Fig. 7 demonstrates the effect of our peak boosting algorithm on time profiles of two pixels of a synthetic scene. The estimate of a transient image, $\beta_1(\tau)$, sinks down due to missing low frequency spectrum. After peak boosting, $\beta_b(\tau)$ is close to the ground truth $\alpha(\tau)$.

Algorithm 2: Peak boosting

Input: time profile $\beta_1(\tau)$, lowest working frequency ω_1 , peak threshold θ .
Output: low frequency spectrum recovered transient image $\beta_b(\tau)$.
Steps:
1: Initialize $\beta_b(\tau) = \beta_1(\tau)$.
2: Find out peaks $\beta_{pk}(\tau) = \beta_b(\tau)[\beta_b(\tau) > \theta]$.
3: Update $\beta_b(\tau) = \beta_1(\tau) + \beta_{pk} * r_L(\tau)$.
4: Repeat **2** to **3** until $\beta_b(\tau) > 0$.

VI. EXPERIMENTAL RESULTS

We evaluated our approach using both synthetic data for ground truth comparisons, and real data sets downloaded from [8]. The source code of our reconstruction algorithm and all the reconstructed transient image videos are available in the supplemental material. In the following, we show some of the results with performance discussions.

A. Synthetic data

First we test on two data sets synthesized from the same ground truth transient image $\alpha(\tau)$ captured by Velten et al.

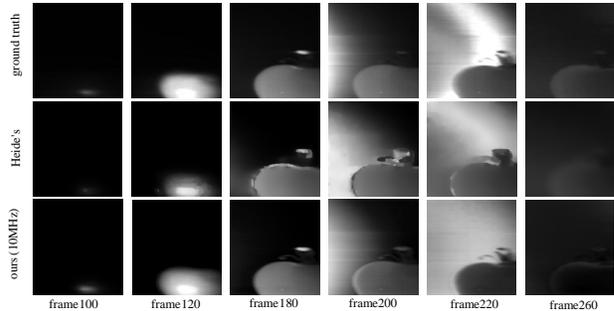


Figure 8. Reconstructed transient images vs. ground truth. Our result is able to vary almost simultaneously with the ground truth (the profile of the tomato in frame 100, 120, 180, 200) and it has no artifact (frame 180, 200). Our result exhibits no highlight in frame 220 and the apple is dark in frame 260.

[7]. This ground truth is also used by Heide et al. [8] for evaluation. The captured image collections $H(\omega, \phi)$ are generated from (3) with the calibrated correlation matrix from [8]. The frequency range of one of the image collections is from 10MHz to 120MHz (called the *10MHz data set*), and the other from 3MHz to 120MHz (called the *3MHz data set*). Gaussian noise $N(0, 0.005^2)$ is added to each image frame.

Fig. 8 shows the transient images reconstructed from the 10MHz data set. Compared with Heide et al.'s result [8], our result is better in matching with the temporal varying tendency of the ground truth (the profile of the tomato in frame 100, 120, 180, 200). Moreover, artifact does not appear in our results while it is quite obvious in Heide et al.'s (frame 180, 200). One flaw in our result is that it does not exhibit the highlight appeared in the ground truth (frame 220) due to the missing high frequency spectrum. Another flaw is that the tomato in frame 260 is darker, since the missing low frequency spectrum is not fully recovered. This problem can be solved by acquiring images with lower frequencies, as shown in our result from the 3MHz data set (see supplementary material).

Fig. 9 further compares the simulation results by showing the time profiles of some pixels. The plots in the first row show that our result cannot reach the true top of a sharp pulse, and the plots in the second row show that our result can not follow an exponential response. However, compared with Heide et al.'s result, our result is able to obtain the shape of the response and does not miss a peak (indicated by arrows in Fig. 9).

Fig. 10 shows that, although our result from the 10MHz data set does not follow exponential responses due to loss of too much low frequency spectrum, our result from the 3MHz data set is able to follow exponential responses. That means our approach is able to recover exponential responses without post processing if the low working frequencies are sampled.

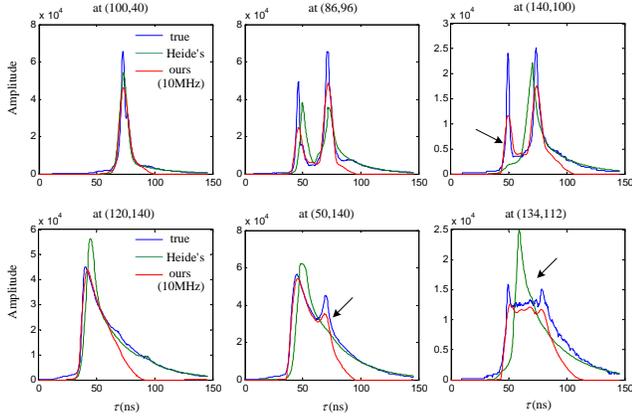


Figure 9. Time sequences of some pixels from the simulation results in Fig. 8. Our result is able to capture the shape of the response and does not miss a peak (compared with Heide et al.’s result indicated by arrows). The flaws of our result are that it does not reach the true top of a sharp pulse (first row) and does not follow exponential responses (second row). The reasons are loss of high frequency spectrum ($>120\text{MHz}$) and loss of low frequency spectrum ($0\text{-}10\text{MHz}$), respectively.

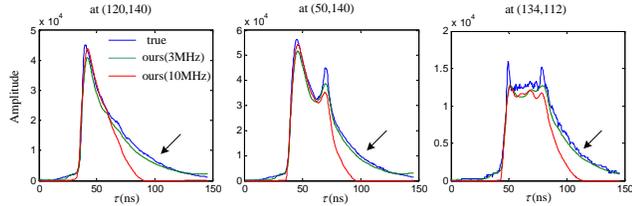


Figure 10. Comparison of our results from the 3MHz data set and the 10MHz data set. Although our result from 10MHz data set does not follow exponential responses due to loss of too much low frequency spectrum, our result from the 3MHz data set is able to follow exponential responses without more actions (indicated by arrows).

B. Real scenes

These data sets are captured by a real imaging system with frequencies ranging from 10MHz to 120MHz with 0.5MHz step increments and phases $\phi = 0, \pi/2$. The calibrated correlation matrix of the imaging system is used to train a parameterized complex correlation function and the result has been shown in Fig. 4 and Fig. 5. The time resolution of the reconstructed transient image is $\tau_s = 0.33\text{ns}$.

We reconstructed all of the five public data sets and the results of two of the data sets are illustrated in Fig. 11. The top row is a simple scene of a corner while the bottom row is a complicated scene with several objects and mirrors. Both of the reconstructed transient images matched to our imagination of how light propagate in these scenes.

Our results are close to the i-step results in [8], while the u-step results in [8] contain errors. Fig. 12 shows the difference in the time sequences of pixels of the mirror scene. The results in [8] by i-step are similar to our results before peak boosting. The u-step in [8] uses a mixture of Gaussians and exponentials to model the scene response such that exponential response can be recovered, as shown in

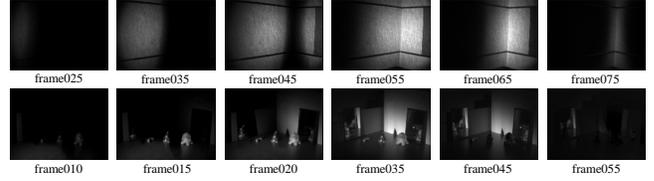


Figure 11. Our reconstructed transient images of two real scenes.

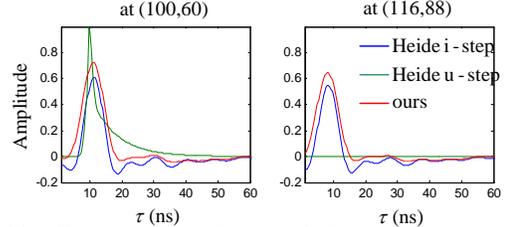


Figure 12. Time sequences of two pixels from the mirror scene in Fig. 11. Heide et al.’s result by i-step is like our result before peak boosting. Heide et al.’s result by u-step uses a mixture of Gaussians and exponentials to model the scene response (left green plot). However, this model may not work (right green plot).

the left of Fig. 12. However, the right of Fig. 12 shows that this model may not work. That is why the u-step results in [8] are worse (refer to the supplemental material). Our approach is able to achieve good results without relying on any prior of a scene or assumption on scene response.

Moreover, the whole process of our approach (not including calibration) is of much lower computational complexity both in space and time than [8]. The main computational intensive operation is the DFT in the reconstruction step. Its space complexity is $O(M)$, and its time complexity is $O(M \log M)$ per pixel, where M is the number of points in FFT. Dilated component removal and peak boosting work on time sequence and usually need to run only once or twice. Their space complexity is $O(n)$, and their time complexity is $O(n)$ per pixel, where n is the number of flying time samples. Comparably, the approach in [8] is a global optimization algorithm whose spatial complexity is at least of order kn , where k is the number of the whole data set. We test on a PC with a CPU of Intel i7-2600 3.40G and 8G RAM. Our approach takes an average running time of 17.2s, while the approach in [8] with a single i-step takes 133.1s, and the approach in [8] with u-step take several hours.

VII. CONCLUSION

In this work, we use Fourier analysis to investigate the principle of transient imaging with a multifrequency ToF camera. Our study reveals the intrinsic relationship between a captured image collection and the quality of the reconstructed transient image. We discuss problems such as measurement noise, harmonic component disturbance and missing low spectrum in a real imaging system, and propose a frequency-domain reconstruction approach to solve these problems. We believe that our Fourier analysis provides not only new insight to the transient imaging problem, but also

potential new theories and techniques for other ToF imaging problems like the multipath ToF problem and the looking through diffusing media problem.

ACKNOWLEDGMENT

This work was supported by the Project of NSFC (No. 61327902, 61035002 & 61120106003), the National Program for Significant Scientific Instruments Development of China (2013YQ140517), and the Max Planck Center for Visual Computing and Communication (German Federal Ministry of Education and Research, 01IM10001).

REFERENCES

- [1] N. Abramson, "Light-in-flight recording by holography," in *Los Angeles Technical Symposium*. International Society for Optics and Photonics, 1980, pp. 140–143.
- [2] R. Raskar and J. Davis, "5d time-light transport matrix: What can we reason about scene properties," *Int.Memo*, 2008.
- [3] D. Wu, M. O'Toole, A. Velten, A. Agrawal, and R. Raskar, "Decomposing global light transport using time of flight imaging," in *CVPR*, 2012, pp. 366–373.
- [4] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using ultrafast transient imaging," *IJCV*, vol. 95, no. 1, pp. 13–28, 2011.
- [5] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature Communications*, vol. 3, p. 745, 2012.
- [6] D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, M. O'Toole, N. Naik, Q. Dai, K. Kutulakos, and R. Raskar, "Frequency analysis of transient light transport with applications in bare sensor imaging," in *ECCV*, 2012, pp. 542–555.
- [7] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar, "Femto-photography: capturing and visualizing the propagation of light," *ACM Trans. Graph.*, vol. 32, no. 4, p. 44, 2013.
- [8] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, p. 45, 2013.
- [9] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in *CVPR*, 2014.
- [10] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight sensors in computer graphics," in *Proc. Eurographics (State-of-the-Art Report)*, 2009, pp. 119–134.
- [11] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt, "3d shape scanning with a time-of-flight camera," in *CVPR*, 2010, pp. 1173–1180.
- [12] V. Ganapathi, C. Plogemann, D. Koller, and S. Thrun, "Real time motion capture using a single time-of-flight camera," in *CVPR*, 2010, pp. 755–762.
- [13] E. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3, pp. 334–343, 2008.
- [14] S. Fuchs, "Multipath interference compensation in time-of-flight camera images," in *ICPR*, 2010, pp. 3583–3586.
- [15] S. Fuchs, M. Suppa, and O. Hellwich, "Compensation for multipath in ToF camera measurements supported by photometric calibration and environment integration," in *ICVS*, 2013, pp. 31–41.
- [16] D. Jiménez, D. Pizarro, M. Mazo, and S. Palazuelos, "Modeling and correction of multipath interference in time of flight cameras," *Image and Vision Computing*, vol. 32, no. 1, pp. 1–13, 2014.
- [17] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, "Separating true range measurements from multi-path and scattering interference in commercial range cameras," in *IS&T/SPIE Electronic Imaging*, 2011.
- [18] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar," in *IS&T/SPIE Electronic Imaging*, 2012.
- [19] A. Kirmani, A. Benedetti, and P. A. Chou, "Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods," in *ICME*, 2013, pp. 1–6.
- [20] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for ToF sensors," *arXiv preprint arXiv:1403.5919*, 2014.
- [21] 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," *Optics Letters*, vol. 39, no. 6, pp. 1705–1708, 2014.
- [22] 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 Trans. Graph.*, vol. 32, no. 6, p. 167, 2013.
- [23] M. L. Simpson, M.-D. Cheng, T. Q. Dam, K. E. Lenox, J. R. Price, J. M. Storey, E. A. Wachter, and W. G. Fisher, "Intensity-modulated, stepped frequency cw lidar for distributed aerosol and hard target measurements," *Applied optics*, vol. 44, no. 33, pp. 7210–7217, 2005.
- [24] E. B. van Munster and T. W. Gadella, "Fluorescence lifetime imaging microscopy (flim)," in *Microscopy Techniques*. Springer, 2005, pp. 143–175.
- [25] W. Becker, *Advanced time-correlated single photon counting techniques*. Springer, 2005, vol. 81.
- [26] P. C. Schneider and R. M. Clegg, "Rapid acquisition, analysis, and display of fluorescence lifetime-resolved images for real-time applications," *Review of Scientific Instruments*, vol. 68, no. 11, pp. 4107–4119, 1997.