Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Review:

Transient imaging with a time-of-flight camera and its applications*

Jing-yu LIN†1, Ri-hui WU2, Hong-man WANG3, Ye-bin LIU†‡4

(1*College of Electrical Engineering, Guangxi University, Nanning 530004, China*)

(2*University of Chinese Academy of Sciences, Beijing 100049, China*)

(3*School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China*)

(4*Department of Automation, Tsinghua University, Beijing 10084, China*)

†E-mail: jylin@gxu.edu.cn; liuyebin@tsinghua.edu.cn

Received Aug. 23, 2017; Revision accepted Sept. 20, 2017; Crosschecked Sept. 30, 2017

Abstract: Transient imaging is a technique in photography that records the process of light propagation before it reaches a stationary state such that events at the light speed level can be observed. In this review we introduce three main models for transient imaging with a time-of-flight (ToF) camera: correlation model, frequency-domain model, and compressive sensing model. Transient imaging applications usually involve resolving the problem of light transport and separating the light rays arriving along different paths. We discuss two of the applications: imaging objects inside scattering media and recovering both the shape and texture of an object around a corner.

Key words: Transient imaging; Time-of-flight (ToF) camera; Scattering media; Around corners
https://doi.org/10.1631/FITEE.1700556 **CLC number:** TP37

1 Introduction

Continuous advances in imaging technology have greatly improved the ways in which human eyes can observe the natural world in both scale and dimension. However, due to the limitations of classical imaging models, it is still difficult to observe ultra-high speed phenomena such as the transport of light. In such a model, a pixel is formed by integrating the photons of incident light over time at a point determined by the camera sensor. Thus, a shorter exposure time means fewer incident photons and consequently a lower signal-to-noise ratio (SNR). For events occurring at light speed, the exposure time required by such an imaging system is insufficient to capture enough photons to form images. Therefore, a new imaging model is needed to achieve breakthroughs in the speed of imaging.

In this paper, we first introduce techniques for acquiring transient images and their imaging models, and then discuss two potential applications for handling transient images, i.e., recovering both the shape and texture of an object (1) inside scattering media and (2) around corners beyond the line of sight. A transient image $i(x, y, t)$ is the record of the transient process of light propagation before it reaches a stationary state, while a conventional image $i(x, y)$ is the steady state of light distribution. When a ray of light interacts with a subject, the transient image of the subject carries spatial and surface information in the path of the light. If the ray of light reaches the subject only once, it carries the depth information of the spot it reaches. If the ray of light reflects inside the subject, it carries the internal spatial relationship information

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61561005 and 61531014), the National Key Foundation for Exploring Scientific Instrument (No. 2013YQ140517), and the Natural Science Foundation of Guangxi Province, China (No. 2015GXNSFAA139284)

ORCID: Jing-yu LIN, http://orcid.org/0000-0002-7194-3300

along its path. Thus, from transient images we can observe visual information about a subject, which is otherwise unobservable by conventional methods.

2 Time-domain transient imaging

Directly acquiring a transient image requires an ultra-fast imaging system (Velten *et al.*, 2013). To record the light signal in an ultra-short exposure time, the sensor must be ultra-sensitive. An ultra-fast streak camera can work at a sampling time of less than 1 ps, but it can acquire only one line encompassing hundreds of pixels each time. To acquire a two-dimensional (2D) plane, a scanning mirror is required to scan in the direction vertical to the pixel line, and then the pixel lines for all scanning angles are combined into a 2D image. To acquire a transient image, the imaging system requires a femtosecond laser as the light source, which emits ultra-short laser pulses. However, the photos from a laser pulse are very few such that the pixels captured by a streak camera will be too weak to read. To address this issue of insufficient incident photons, the laser needs to emit pulses for some time at each scanning angle. Then all the images from the same angle are added together to increase the SNR of the image. This method requires that the laser synchronize with the camera using a light splitter. This results in a bulky, sophisticated, and expensive system. However, it can obtain transient images at a high temporal resolution for up to several picoseconds.

Transient imaging with a time-of-flight (ToF) camera (Heide *et al.*, 2013) is more convenient and cheaper than using a streak camera. A ToF camera (Fig. 1) is a compact imaging system, which modulates both its laser and its sensor with the same high-frequency periodic signal. The operating principle is shown in Fig. 2. The camera accumulates the incident light energy over the exposure time, and thus the image acquired by the ToF camera is an integration of the reflected light and the modulated gain of the sensor. The image obtained for each pixel is formulated as follows:

$$H(\omega,\varphi)=\int i(\tau)C(\omega,\varphi,\tau)\mathrm{d}\tau, \qquad (1)$$

where $C(\omega, \varphi, \tau)$ is the correlation function between the modulated laser signal and modulated sensor gain, and $i(\tau)$ is the transient image. After measuring correlation function $C(\omega, \varphi, \tau)$ and acquiring images $H(\omega, \varphi)$ at a range of frequencies and phases, the transient images can be solved from Eq. (1).

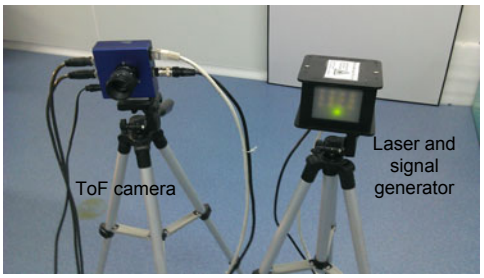

Fig. 1 A prototype time-of-flight (ToF) camera
The laser (right) and the camera (left) are modulated by the same signal generator

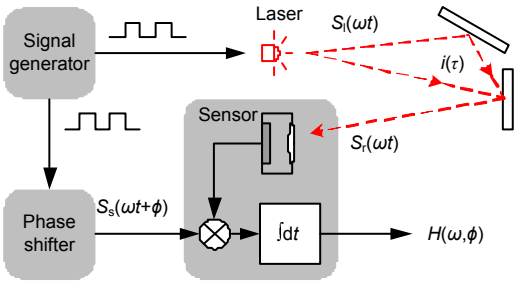

Fig. 2 Operation principle of a time-of-flight (ToF) camera

If the laser and sensor in a ToF camera are modulated by different signals, the imaging equation still holds. By designing a modulation signal such that the correlation function is invertible, the transient image can be solved from Eq. (1) in a single shot (Kadambi *et al.*, 2013). In theory, the modulation signals need to be across a wide spectrum. In practice, the distortion of the light emitted by the laser will degrade the quality of the solved transient images.

3 Frequency-domain transient imaging

The methods mentioned in Section 2 obtain a temporal sequence of images either by optical means (using a streak camera) or by solving an imaging equation (using a ToF camera). Therefore, we characterize these methods as 'time-domain transient imaging'. When the sequence of images obtained is in the frequency domain, i.e., the Fourier transform

along the time axis of a transient image, we call these methods 'frequency-domain transient imaging'. It has been proven that such images can be achieved with a ToF camera working at frequencies ranging from low to high by emitting an ideal sinusoidal laser signal and recording the optical response (Lin *et al.*, 2014). In this case, the correlation function $C(\omega, \varphi, \tau)$ in Eq. (1) must be an ideal sine function of τ, i.e.,

$$C(\omega,\varphi,\tau)=\cos(\omega\tau+\varphi). \tag{2}$$

The frequency sequence of images can be transformed into a transient image by the inverse Fourier transform.

Two issues need to be addressed for frequency-domain transient imaging. The first issue is that generating an ideal sinusoidal laser signal is difficult due to the nonlinearity of electrical components. In practice, the modulation signal is usually a square wave and the emitting light is a periodic signal with a distorted shape. One way to address the issue is by making a Fourier series decomposition. Specifically, we first acquire the correlation function at a series of equally spaced phases for each frequency ω (Fig. 3) and then decompose them into a Fourier series:

$$C(\omega,\varphi,\tau)=\sum_{n=0}^{\infty}A_n\cos(n\omega\tau+n\varphi+\varphi_n). \tag{3}$$

Next, we acquire the images at a series of equally spaced phases for each frequency and decompose them into a Fourier series. After removing phase offset φ_1 and amplitude A_1 from the fundamental component of the phase series of images, we obtain an image at each frequency ω.

The second issue is that the highest frequency of the modulation signal of a ToF camera is limited. When the working frequency is close to the upper boundary, the amplitude of the signal decreases dramatically (Fig. 4). For example, a PMD PhotonICs 19k-S3 sensor can work at a frequency less than 165 MHz. That is, while the frequency of the image series goes up to 165 MHz, the corresponding sampling time of the transient image remains about 6 ns.

One way to improve the frequency is by exploiting high-order harmonic components from the above Fourier series decomposition (Lin *et al.*, 2017). The image of the harmonic component of order n at frequency ω is equivalent to the image at frequency $n\omega$. However, the amplitudes of high-order harmonic components are usually much lower than that of the fundamental component. When the modulation signal is a square wave, the amplitudes of the components of odd orders are the inverse of the order and the amplitudes of the components of even orders are trivial. Moreover, high-frequency distortion of the electrical circuit suppresses the amplitudes of high-order components. For example, Fig. 4 shows the amplitude of each component of a prototype ToF camera. The third-order harmonic component (A_3) is nontrivial at frequencies below 90 MHz. Thus, the maximum

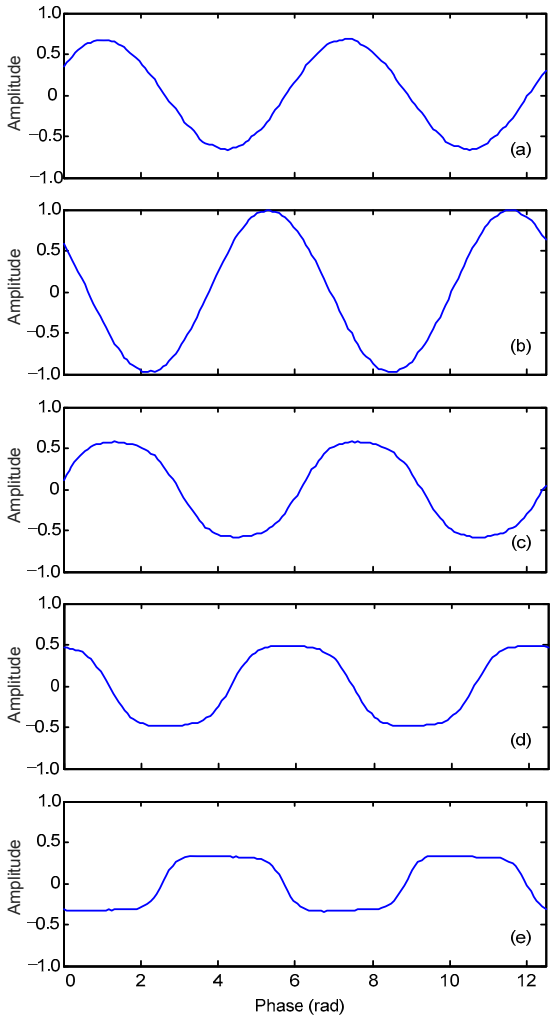

Fig. 3 The correlation function at a pixel at different frequencies: (a) 130 MHz; (b) 110 MHz; (c) 70 MHz; (d) 50 MHz; (e) 30 MHz

The waveforms at low frequencies contain more harmonic components, whereas the waveforms at high frequencies are close to sine waves

frequency of the image series obtained can get up to 90×3=270 MHz. If the distortion of the electrical circuit can be rectified, higher-order components can be achieved.

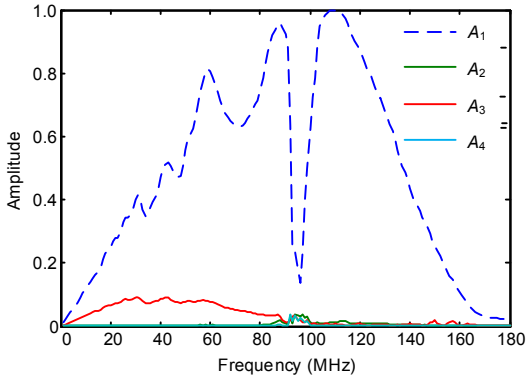

Fig. 4 Amplitude of each frequency component of the correlation function at a pixel
The third-order harmonic component (A_3) is nontrivial at frequencies below 90 MHz

4 Compressive transient imaging

Transient imaging usually requires multiple images to obtain reliable results and thus the acquisition process requires much more time than conventional imaging. This prevents it from being applied to dynamical scenes. Researchers have found that the number of images taken can be reduced greatly in the case of a scene with sparse responses. That is, the optical response is a continuous-time K-sparse signal:

$$i(t)=A_0\delta(t-t_0)+\sum_{k=1}^{K}A_k\delta(t-t_k), \tag{4}$$

where $\delta(\cdot)$ denotes the Dirac function, K is the number of optical paths, and $\{A_k,t_k\}_{k=0}^{K}$ are the amplitudes and delays. A transient image with K pulse signals can be reconstructed from a few measurements by employing an optimization method with sparsity regularization (Bhandari *et al.*, 2014; Peters *et al.*, 2015; Qiao *et al.*, 2015; Bhandari and Raskar, 2016). However, if K is a relatively large number, it will take a very long time to solve the transient image.

Instead of directly solving a transient image reconstruction problem based on sparsity regularization, here we recover the measurements from a few measurements acquired by a multi-frequency ToF camera using compressive sensing, a signal processing technique for efficiently acquiring and reconstructing a signal by finding solutions to underdetermined linear systems (https://en.wikipedia.org/wiki/Compressed_sensing). Compared to the frequency-domain method, this method reduces the number of modulation frequencies from hundreds to dozens.

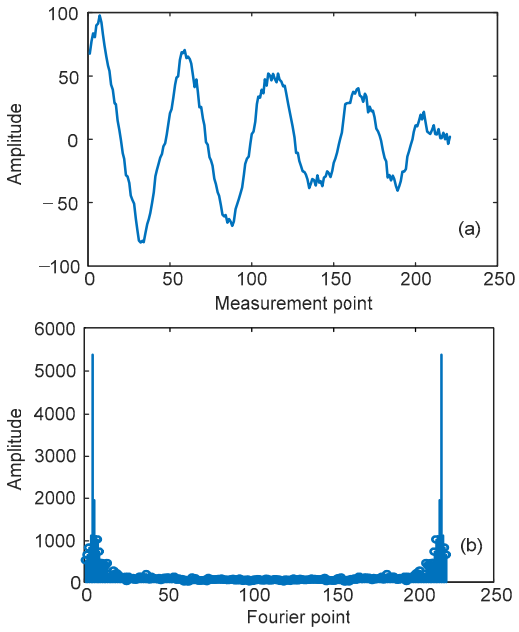

Fig. 5 The signal in the measurement domain (a) and its amplitude spectrum in the Fourier domain (b) for pixel (60, 70) with phase 0°

In this method, we take the measurements as an original signal, that is $\boldsymbol{X}\in\mathbb{R}^N$, where N is the number of modulation frequencies. In fact, the measured signal $\boldsymbol{X}$ is a frequency-domain signal (Lin *et al.*, 2014) and is not sparse in the measurement domain. Here we take it in the measurement domain. It is sparse in the Fourier transform. Take the case of pixel (60, 70) in the acquired 'Corner' image (Heide *et al.*, 2013) as an example. Fig. 5a shows the signal in the measurement domain, and Fig. 5b shows its amplitude spectrum in the Fourier transform. Most of the Fourier coefficients are small, except the first several large amplitude coefficients.

Therefore, we assume the measurement signal $\boldsymbol{X}$ to be sparse in the Fourier domain and define

$$\boldsymbol{X}=\boldsymbol{\Psi\theta}, \tag{5}$$

where $\boldsymbol{\Psi}$ is a sparse matrix, and $\boldsymbol{\theta}$ is K-sparse. Define the acquired measurement signal $\boldsymbol{Y}\in\mathbb{R}^M$ as

$$\boldsymbol{Y}=\boldsymbol{\Phi X},\quad M<<N, \tag{6}$$

where M is the number of acquired measurement signals. From Eqs. (5) and (6) we have

$$\boldsymbol{Y}=\boldsymbol{\Phi\Psi\theta}=\boldsymbol{S\theta}, \tag{7}$$

where $\boldsymbol{S}=\boldsymbol{\Phi\Psi}$. The process of recovering all $\boldsymbol{X}$'s from a $\boldsymbol{Y}$ can be implemented by solving the L_1 norm:

$$\min\left\|\tilde{\boldsymbol{\theta}}\right\|_1 \quad \text{s.t.}\ \ \boldsymbol{Y}=\boldsymbol{\Phi}\tilde{\boldsymbol{X}}=\boldsymbol{\Phi\Psi}\tilde{\boldsymbol{\theta}}=\boldsymbol{S}\tilde{\boldsymbol{\theta}}. \tag{8}$$

The length M of the acquired measurement signal is randomly selected from the N measurement signals, that is, $m_i\in[1, N]$, i=1, 2, …, M. Define the element $\varphi_{i,j}$ of measurement matrix $\boldsymbol{\Phi}_{M\times N}$ as

$$\varphi_{i,j}=\begin{cases}1, & i=1,2,\cdots,M,\ j=m_i,\\ 0, & \text{otherwise.}\end{cases} \tag{9}$$

Define sparse matrix $\boldsymbol{\Psi}$ as

$$\boldsymbol{\Psi}=\frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}} & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot 2} & \cdots & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot(N-1)}\\ 1 & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot 2} & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot 4} & \cdots & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot 2(N-1)}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}(N-1)} & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot 2(N-1)} & \cdots & \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}\cdot(N-1)(N-1)}\end{bmatrix}.$$

The signal is recovered using the orthogonal matching pursuit algorithm (Tropp and Gilbert, 2007). Algorithm 1 summarizes the recovery process.

Algorithm 1 Signal recovery process

Input: $M\times N$ sensing matrix $\boldsymbol{S}$, $M\times 1$ measured vector $\boldsymbol{Y}$, and sparsity level K of the signal.

Output: an estimate $\tilde{\boldsymbol{X}}$ in $\mathbb{R}^N$ for $\boldsymbol{X}$.

Initialize residual $\boldsymbol{r}_0=\boldsymbol{Y}$, index set $\Lambda_0=\varnothing$, matrix of the chosen atoms $T_0=\varnothing$, and iteration counter t=1.

For t=1 to $2K$ **do**

Solve the optimization problem to find index λ_t:

$$\lambda_t=\arg\max_{j=1,2,\cdots,N}\left|\left\langle \boldsymbol{r}_{t-1},\boldsymbol{s}_j\right\rangle\right|.$$

Augment index set $\Lambda_t=\Lambda_{t-1}\cup\{\lambda_t\}$ and matrix of chosen atoms $\boldsymbol{T}_t=[\boldsymbol{T}_{t-1},\boldsymbol{s}_{\lambda_t}]$.

Solve a least-squares problem to find a new signal estimate:

$$\tilde{\boldsymbol{\theta}}_t=\arg\min_{\tilde{\theta}}\left\|\boldsymbol{Y}-T_t\tilde{\boldsymbol{\theta}}\right\|_2.$$

Calculate the new residual $\boldsymbol{r}_t=\boldsymbol{Y}-T_t\tilde{\boldsymbol{\theta}}_t$.

End for

Reconstruct the Fourier spectral vector by $\tilde{\boldsymbol{\theta}}$.

Recovery $\tilde{\boldsymbol{X}}$ by the inverse Fourier transform.

Apply frequency-domain transient imaging (Lin *et al.*, 2017).

For example, the number of modulation frequencies of the real dataset 'Corner' (Heide *et al.*, 2013) is N=221. If the sparsity level K=14, we choose $M=K\log(N/K)$=39. Then, randomly select 39 frequencies from the original 221 frequencies and reconstruct the signals. Fig. 6 shows the recovered signal in comparison to the actually acquired signal for pixel (82, 67). In Fig. 6a, the measured signal is selected randomly from the actual signals. In Fig. 6b, the recovered signal is close to the ground truth.

When sparsity level K=10, and M=20, the recovered results are shown in Fig. 7. The errors are larger compared to those in the previous example due to the fewer measurements.

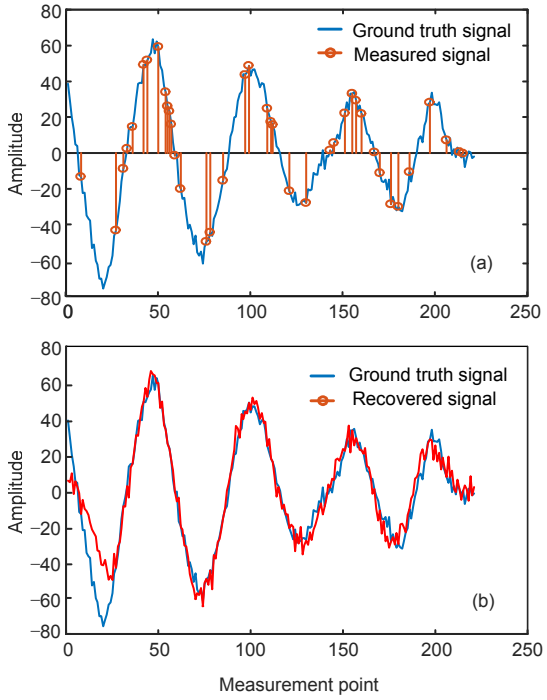

Fig. 6 Recovered signal compared to the ground truth signal at pixel (82, 67) with phase 90° (K=14, M=39)

In (a), the measured signal is selected randomly; in (b), the recovered signal is close to the ground truth (MSE= 0.30991)

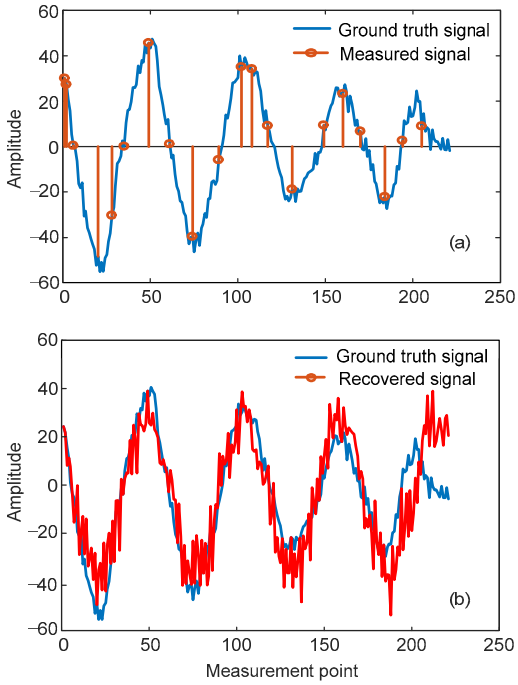

Fig. 7 Recovered signal compared to the ground truth signal at pixel (30, 50) with phase 90° (*K*=10, *M*=20)
In (a), the measured signal is selected randomly; in (b), the recovered signal has larger errors compared to that in the previous example (MSE=0.59684)

Fig. 8 compares the reconstructed transient images. The sparsity level is 10, and the number of measured frequencies is 20. These images are close even though the reconstructed ones contain more noise.

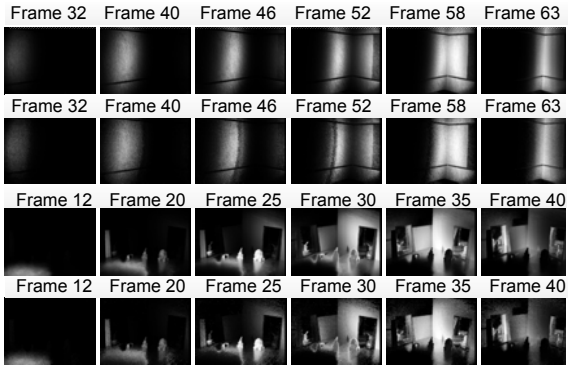

Fig. 8 Comparison of the recovered images
The first and third rows are the transient images reconstructed using 221 frequencies; the second and fourth rows are the transient images reconstructed using 20 frequencies

Experiment results show that the signal can be reconstructed from its random measurement samples if the number of samples is large enough with respect to sparsity, which is related to the structure of the observed scene. The sparser the signal is, the fewer the random measurement samples are required, and the larger error the reconstruction yields. In the future, we would like to investigate an optimal method for measurement selection.

5 Applications

In this section, we apply transient imaging to obtain both the depth map and texture of an object, when the object is inside some kind of scattering media (such as turbid water) or is obstructed and beyond the line of sight.

5.1 Imaging through scattering media

For simplicity, the behavior of light inside scattering media is usually described as single scattering with exponential decay (Nishita *et al.*, 1987). As shown in Fig. 9, the intensity received by the camera consists of single scattering rays and reflected rays:

$$E=\int_0^{d_1}\frac{I\beta \mathrm{e}^{-\beta(x+y)}}{4\pi y^2}\mathrm{d}x+\frac{I\rho \mathrm{e}^{-\beta(d_1+d_2)}}{d_2^2}, \qquad (10)$$

where I is the intensity of the light source, β is the scattering coefficient of the scattering media, the phase function is set to $\pi/4$ for homogeneous media, x and y are the lengths of the light path, and ρ is the surface albedo of the objects. We call this model a steady-state single scattering model since it is the steady-state response of the object.

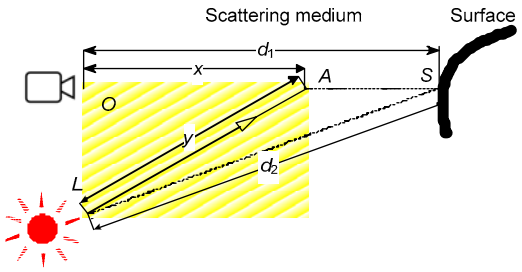

Fig. 9 Traditional single scattering model: steady-state model

We can generalize this model to a transient scenario. In Fig. 10, assume that the incident ray and reflected ray are in the same direction at each pixel

(Shim and Lee, 2016). Thus, the image formation at each pixel is independent and we can estimate the depth map and texture in a pixel-wise manner. Define a transient pixel as the impulse optical response at a pixel. The scattering component of a transient pixel is

$$S_{\mathrm{S}}(t)=\frac{I\beta \mathrm{e}^{-2\beta x(t)}}{4\pi[ct_0/2+x(t)]^2}1(t),\quad t\in[t_0,t_{\mathrm{R}}], \tag{11}$$

where t is the traveling time of a given ray, ct_0 is the path length before entering the medium, $x(t)$ is the depth of ray entering the medium, and $1(t)$ is a step function whose value is 1 when $t>0$. The reflected component of a transient pixel is

$$S_{\mathrm{R}}(t)=\frac{I\rho \mathrm{e}^{-2\beta x(t_{\mathrm{R}})}}{[ct_0/2+x(t_{\mathrm{R}})]^2}\cdot\delta(t-t_{\mathrm{R}}). \tag{12}$$

That is, the transient single scattering model is

$$\begin{aligned}S(t)=&\frac{I\beta \mathrm{e}^{-2\beta x(t)}}{4\pi[ct_0/2+x(t)]^2}1_{t\in[t_0,t_{\mathrm{R}})}\\&+\frac{I\rho \mathrm{e}^{-2\beta x(t_{\mathrm{R}})}}{[ct_0/2+x(t_{\mathrm{R}})]^2}\cdot\delta(t-t_{\mathrm{R}}),\end{aligned} \tag{13}$$

and

$$x(t)=\frac{c}{2n}(t-t_0), \tag{14}$$

where c is the speed of light, and n is the refractive index of the medium, for example, n=1.33 for a water medium.

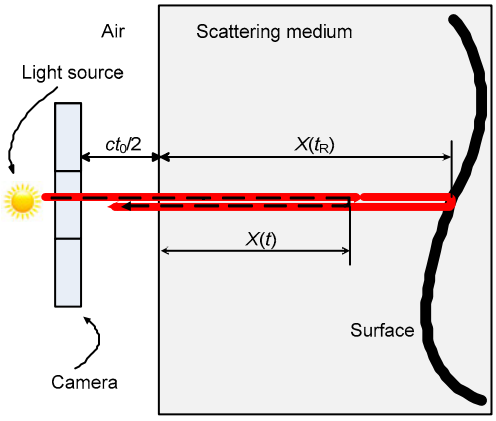

Fig. 10 Transient single scattering model

In model (13), we need to estimate depth $x(t_{\mathrm{R}})$ and texture ρ from transient pixel $S(t)$. When using a ToF camera, the temporal resolution of the reconstructed transient pixel is limited by the maximum frequency (Lin *et al.*, 2014). Therefore, the reconstructed transient pixel is

$$y(t)=S(t)*k(t), \tag{15}$$

where $k(t)$ is a smoothing kernel function.

To estimate depth $x(t_{\mathrm{R}})$, we first reconstruct a transient image $y_1(t)$ at location R_1 and then move the object to location R_2 and reconstruct a transient image $y_2(t)$. Assuming $t_{\mathrm{R1}}<t_{\mathrm{R2}}$, we have

$$\begin{aligned}y_1(t)-y_2(t)=&\frac{I\rho \mathrm{e}^{-2\beta x(t_{\mathrm{R1}})}}{[ct_0/2+x(t_{\mathrm{R1}})]^2}k(t-t_{\mathrm{R1}})\\&-\frac{I\rho \mathrm{e}^{-2\beta x(t_{\mathrm{R2}})}}{[ct_0/2+x(t_{\mathrm{R2}})]^2}k(t-t_{\mathrm{R2}})\\&+\frac{I\beta \mathrm{e}^{-2\beta x(t)}}{4\pi[ct_0/2+x(t)]^2}1_{\mathrm{R1,R2}}(t)*k(t).\end{aligned} \tag{16}$$

Note that the third item (scattering component) on the right-hand side of Eq. (16) is much smaller than the first and second items (reflected components) if the movement is small. By Taylor expansion of the right-hand side of Eq. (16) and keeping the first-order term, we have

$$\begin{aligned}y_1(t)-y_2(t)\approx&(t_{\mathrm{R1}}-t_{\mathrm{R2}})\cdot\frac{I\rho \mathrm{e}^{-2\hat{\beta}x(t_{\mathrm{R1}})}}{[ct_0/2+x(t_{\mathrm{R1}})]^2}\\&\cdot\left[\frac{c}{n}\left(\hat{\beta}+\frac{1}{ct_0/2+x(t_{\mathrm{R1}})}\right)k(t-t_{\mathrm{R1}})+k'(t-t_{\mathrm{R1}})\right].\end{aligned} \tag{17}$$

This is a unimodal function whose extreme value is at $G'(t_{\mathrm{m}})=0$ and β is estimated as

$$\hat{\beta}\approx-\frac{nk''(t_{\mathrm{m}}-t_{\mathrm{R1}})}{ck'(t_{\mathrm{m}}-t_{\mathrm{R1}})}-\frac{1}{ct_0/2+x(t_{\mathrm{R1}})}. \tag{18}$$

We can see that t_{R1}, t_{R2}, and $\hat{\beta}$ are one-to-one correspondence with a profile of $y_1(t)$–$y_2(t)$. By traversing t_{R1} and calculating the corresponding t_{R1} and $\hat{\beta}$, a group of $y_1(t)$–$y_2(t)$ can be obtained. Then t_{R1} can be estimated by fitting the profile acquired, and $x(t_{\mathrm{R1}})$ is the depth.

After having t_{R1}, texture ρ is estimated by the least-squares method as follows. The reconstructed

transient pixel $y(t)$ can be expressed as a linear combination:

$$y(t)=I f_{\mathrm{S}}(t)+I\rho f_{\mathrm{R}}(t), \tag{19}$$

where

$$f_{\mathrm{S}}(t)=\frac{\beta \mathrm{e}^{-2\beta x(t)}}{4\pi[ct_0/2+x(t)]^2}1_{t\in[t_0,t_{\mathrm{R}})}*k(t), \tag{20}$$

$$f_{\mathrm{R}}(t)=\frac{\mathrm{e}^{-2\beta x(t_{\mathrm{R}})}}{[ct_0/2+x(t_{\mathrm{R}})]^2}k(t-t_{\mathrm{R}}). \tag{21}$$

By discretizing $f_{\mathrm{S}}(t)$, $f_{\mathrm{R}}(t)$, and $y(t)$, we have

$$\boldsymbol{Y}=A\begin{bmatrix} I \\ I\rho \end{bmatrix}, \tag{22}$$

where

$$\boldsymbol{A}=\begin{bmatrix} f_{\mathrm{S}}(t_1) & f_{\mathrm{S}}(t_2) & \cdots & f_{\mathrm{S}}(t_N) \\ f_{\mathrm{R}}(t_1) & f_{\mathrm{R}}(t_2) & \cdots & f_{\mathrm{R}}(t_N) \end{bmatrix}^{\mathrm{T}}, \tag{23}$$

$$\boldsymbol{Y}=\begin{bmatrix} y(t_1) & y(t_2) & \cdots & y(t_N) \end{bmatrix}^{\mathrm{T}}. \tag{24}$$

Solving Eq. (22) yields the texture.

5.2 Imaging around corners

Looking around corners beyond the line of sight is a novel application of transient images. The MIT Media Lab (Kirmani *et al.*, 2009) proposed the original setup for the imaging system. This technique takes a diffuse wall as a mirror and recovers the shape around a corner by resolving the light paths from the wall. Another way using a multi-frequency ToF camera is to formulate the imaging equations based on the principles of the cameras and to find out their sparse solutions (Heide *et al.*, 2014). In this application, the precision of the recovered structure is basically decided by the temporal resolution of the imaging system (or the maximum frequency of the ToF cameras).

Here, we extend the existing work on reconstruction of the shapes of objects around corners (Heide *et al.*, 2014) to imaging around corners, i.e., recovering both the shape and texture around corners. Color can also be obtained if we use red, green, and blue laser diodes as the light sources. We show how to solve for the shape and texture in the following.

The setup of the imaging system is shown in Fig. 11. The coordinates are set up on the wall. The laser is located at $(\boldsymbol{x}_{\mathrm{l}}, z_{\mathrm{l}})$ and shoots at a point $\boldsymbol{x}_{\mathrm{s}}$ on the wall. The camera is located at $(\boldsymbol{x}_{\mathrm{c}}, z_{\mathrm{c}})$ and faces the wall. The pixels of the camera can be transformed by homography to the points, denoted by $\boldsymbol{x}_{\mathrm{p}}$, at the wall. These values are measured before the experiment. The transient image $i(t, \boldsymbol{x}_{\mathrm{p}})$ transformed on the wall coordinates has been acquired. The shape $z(\boldsymbol{x})$ and texture $r(\boldsymbol{x})$ of the object behind the corner, where $\boldsymbol{x}$ are the coordinates of the object projected on the wall, are unknown and need to be solved.

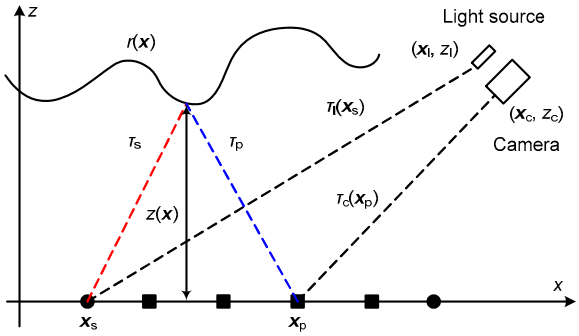

Fig. 11 Setup of the imaging system for looking around corners

The equation for transient imaging is as follows:

$$i(t,\boldsymbol{x}_{\mathrm{p}})=\int r(\boldsymbol{x})\delta[t-\tau_z(\boldsymbol{x},\boldsymbol{x}_{\mathrm{p}})]\mathrm{d}\boldsymbol{x}, \tag{25}$$

where τ_z is the time of flight from the laser to the camera, i.e.,

$$\tau_z=\tau_{\mathrm{l}}+\tau_{\mathrm{s}}(\boldsymbol{x},\boldsymbol{x}_{\mathrm{s}})+\tau_{\mathrm{p}}(\boldsymbol{x},\boldsymbol{x}_{\mathrm{s}})+\tau_{\mathrm{c}}. \tag{26}$$

The variable time delay τ_z in Eq. (25) is not easy to handle, so let us take a Fourier transform:

$$I(\omega,\boldsymbol{x}_{\mathrm{p}})=\int r(\boldsymbol{x})\exp[-\mathrm{j}\omega\tau_z(\boldsymbol{x},\boldsymbol{x}_{\mathrm{p}})]\mathrm{d}\boldsymbol{x}. \tag{27}$$

Denote

$$h_z(\omega,\boldsymbol{x},\boldsymbol{x}_{\mathrm{p}})=\exp\{-\mathrm{j}\omega\tau_z(\boldsymbol{x},\boldsymbol{x}_{\mathrm{p}})\}, \tag{28}$$

and Eq. (27) becomes

$$I(\omega,\boldsymbol{x}_{\mathrm{p}})=\int r(\boldsymbol{x})h_z(\omega,\boldsymbol{x},\boldsymbol{x}_{\mathrm{p}})\mathrm{d}\boldsymbol{x}. \tag{29}$$

It is easy to find out that Eq. (29) has the same form as the convolution imaging equation. h_z is in fact the point spread function (PSF). If the PSF is space-invariant, the image is the convolution of the source of object $r(\boldsymbol{x})$ and the PSF; otherwise, the imaging equation is in the form of Eq. (29).

Therefore, the problem of imaging around corners can be solved by classical vision methods, for example, the methods of depth from defocus. In most cases, shape $z(\boldsymbol{x})$ can be solved first and texture $r(\boldsymbol{x})$ next. However, the problem of imaging around corners is more difficult than depth from defocus in that function h_z is a sine which is spatial-infinite while a PSF is compact in both space and frequency domains, i.e., a space-frequency atom.

Except for the difficulty in solving the imaging equation, there are two more difficulties in imaging around corners. One difficulty is that the temporal resolution of transient images limits the depth resolution of the recovered shape. Therefore, a typical subject is a planar object with holes such that the shape is either a fixed value (on the plane) or the maximum value (in the hole). The other difficulty is that the light reaching the camera is too weak after the three bounces. This may be overcome by obtaining more images and summing all of them.

6 Conclusions

We have introduced three models of transient imaging with a ToF camera, which differ from the classical integration model. The first is related to a correlation function, the second is in the frequency domain, and the third is based on compressive sensing. Transient imaging applications are based on the fact that the paths of light rays can be resolved from a transient image. We discuss the methods for two applications: (1) imaging objects inside scattering media and (2) recovering both the shape and texture of an object around a corner.

At present, these acquisition techniques take too much time in dynamic scenes given that they require multiple images at different camera settings, even using compressive sensing. The acquisition time can be reduced if more information can be obtained in the shot, for example, using multiple light sources.

Another difficulty in using transient imaging is that the intensity of the light becomes too weak to extract after multiple diffuse reflections, such that applications relying on resolving multiple ray paths (like imaging around corners) yield rough results. This raises the problem of low light signal extraction in signal processing, and requires new models for diffuse reflection and advanced signal processing theories.

References

Bhandari, A., Raskar, R., 2016. Signal processing for time-of-flight imaging sensors: an introduction to inverse problems in computational 3-D imaging. *IEEE Signal Process. Mag.*, **33**(5):45-58. https://doi.org/10.1109/MSP.2016.2582218

Bhandari, A., Feigin, M., Izadi, S., *et al.*, 2014. Resolving multipath interference in Kinect: an inverse problem approach. Proc. IEEE Sensors, p.614-617. https://doi.org/10.1109/ICSENS.2014.6985073

Heide, F., Hullin, M.B., Gregson, J., *et al.*, 2013. Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph.*, **32**(4), Article 45. https://doi.org/10.1145/2461912.2461945

Heide, F., Xiao, L., Heidrich, W., *et al*., 2014. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, p.3222-3229. https://doi.org/10.1109/CVPR.2014.418

Kadambi, A., Whyte, R., Bhandari, A., *et al.*, 2013. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.*, **32**(6), Article 167. https://doi.org/10.1145/2508363.2508428

Kirmani, A., Hutchison, T., Davis, J., *et al*., 2009. Looking around the corner using transient imaging. Proc. IEEE 12th Int. Conf. on Computer Vision, p.159-166. https://doi.org/10.1109/ICCV.2009.5459160

Lin, J.Y., Liu, Y.B., Hullin, M.B., *et al.*, 2014. Fourier analysis on transient imaging with a multifrequency time-of-flight camera. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, p.3230-3237. https://doi.org/10.1109/CVPR.2014.419

Lin, J.Y., Liu, Y.B., Suo, J.L., *et al.*, 2017. Frequency-domain transient imaging. *IEEE Trans. Patt. Anal. Mach. Intell.*, **39**(5):937-950. https://doi.org/10.1109/TPAMI.2016.2560814

Nishita, T., Miyawaki, Y., Nakamae, E., 1987. A shading model for atmospheric scattering considering luminous intensity distribution of light sources. *ACM SIGGRAPH Comput. Graph.*, **21**(4):303-310. https://doi.org/10.1145/37402.37437

Peters, C., Klein, J., Hullin, M.B., *et al.*, 2015. Solving trigonometric moment problems for fast transient imaging. *ACM Trans. Graph*., **34**(6), Article 220. https://doi.org/10.1145/2816795.2818103

Qiao, H., Lin, J.Y., Liu, Y.B., *et al*., 2015. Resolving transient time profile in ToF imaging via log-sum sparse regularization. *Opt. Lett*., **40**(6):918-921. https://doi.org/10.1364/OL.40.000918

Shim, H., Lee, S., 2016. Recovering translucent objects using a single time-of-flight depth camera. *IEEE Trans. Circ. Syst. Video Technol.*, **26**(5):841-854. https://doi.org/10.1109/TCSVT.2015.2397231

Tropp, J.A., Gilbert, A.C., 2007. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, **53**(12):4655-4666. https://doi.org/10.1109/TIT.2007.909108

Velten, A., Wu, D., Jarabo, A., *et al.*, 2013. Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.*, **32**(4), Article 44. https://doi.org/10.1145/2461912.2461928