Resolving transient time profile in ToF imaging via log-sum sparse regularization

Hui Qiao,[1] Jingyu Lin,[1,2,4] Yebin Liu,[1] Matthias B. Hullin,[3] and Qionghai Dai[1,5]

[1]*Department of Automation, Tsinghua University, Beijing 100084, China*

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

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

[4]*e-mail: jylin@gxu.edu.cn*

[5]*e-mail: qhdai@tsinghua.edu.cn*

Received December 3, 2014; revised December 29, 2014; accepted December 30, 2014; posted January 9, 2015 (Doc. ID 228416); published March 4, 2015

Multi-frequency time-of-flight (ToF) cameras have been used to recover the transient time profiles of optical responses such that multipath interference can be separated. The resolution of the recovered time profiles is limited by the highest modulation frequency. Here, we demonstrate a method based on log-sum sparsity regularization to recover transient time profiles of specular reflections. We show that it improves the ability of separating pulses better than the state-of-the-art regularization methods. As an application, we demonstrate the encoding and decoding of hidden images using mirror reflections. © 2015 Optical Society of America

OCIS codes: (110.1758) Computational imaging; (100.3190) Inverse problems; (120.5700) Reflection; (150.5670) Range finding.

http://dx.doi.org/10.1364/OL.40.000918

An amplitude modulated continuous wave (AMCW) time-of-flight (ToF) imager offers the range information and shape profile of a scene by emitting a modulated light signal and measuring the phase shift of the light reflected back onto the sensor. A common challenge to such systems is multipath interference (MPI), where light reflected by multiple surfaces superimposes on a single detector and causes a false depth estimate [1]. The temporal structure of backscattered light depends on the geometry and reflectance of the scene: specular, smooth surfaces (like windows or hardwood floors) typically produce a sparse response ("specular MPI"), and diffuse surfaces produce a dense response ("diffuse MPI"). As a third common source of multipath mixing, pixels that fall onto depth discontinuities sample a varying mixture of foreground and background that is sparse both temporally and spatially ("flying pixels").

While in traditional ToF imaging, the amplitude and phase of the reflected signal can in principle be recovered from 2 measurements per pixel, multipath scattering introduces more unknowns and hence requires more measurements, as previous works have shown [2,3]. The most general case is the recovery of a full transient optical response $\alpha(t)$ per pixel, which is related to AMCW measurements through a linear relation [4] and can be expressed by a Fourier transform [5]. However, the resolution of the recovered time profile is dependent on the highest modulation frequency. To constrain the solution space of such linear inverse problems to a temporally sparse response, sparsity priors have been used, such as ℓ_1- [6] or ℓ_0- [7] norm regularizers. The development of application-specific priors in ToF and transient imaging is the subject of ongoing research, as shown, e.g., by Heide *et al.* [8], who used an exponentially modified Gaussian basis to sparsely represent even backscatter profiles that are dense in time domain.

In the case of specular MPI, it is reasonable to use sparsity assumption and employ ℓ_0 regularization, which means finding out the solution that has minimal number of nonzero values. However, a problem with ℓ_0 regularization is NP-hard [9], i.e., the exact solution to such sparse optimization requires an intractable combinatorial search. A common approach that makes such problem tractable is to minimize its convex envelope using an ℓ_1 norm as a proxy for ℓ_0 norm. However, there exists a difference between the combinatorial problem and its relaxation, namely, the dependence on magnitude: larger coefficients are penalized more heavily in the ℓ_1 norm than smaller coefficients, unlike the more democratic penalization of the ℓ_0 norm.

In this work, we propose to use log-sum sparse regularization (LSR) to address this imbalance in the recovery of time profile of specular MPI (a very common problem in video gaming, the largest ToF market [6]) from multi-frequency and multi-phase measurements. Figure 1 illustrates typical scene geometry and the associated time profile of specular MPI. We show that LSR greatly improves the temporal resolution for trails of pulses, and we demonstrate that MPI can be exploited to decompose hidden images encoded through specular reflections.

The ToF camera used in this work contains a set of laser diodes as light source and a photonic mixer device (PMD) as a ToF sensor, as shown in Fig. 1(a) [5]. The system works under a homodyne setting, where the

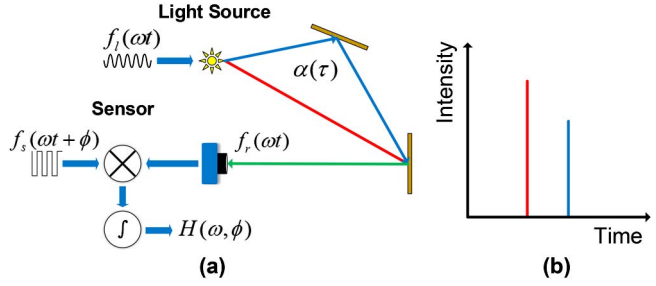

Fig. 1. (a) Working principle of the ToF camera. (b) Time profile of the received signal in specular case, comprising multiple pulses.

output of the light source $f_l(\omega t)$ and the zero-mean reference signal $f_s(\omega t+\phi)$ of the sensor share the same modulation frequency and a fixed relative phase ϕ. Light propagates through the scene to the object point and then bounces back to the sensor pixel via multiple paths. Each single light path q experiences an attenuation factor $a(q)$, and its total optical length corresponds to a delay $\mu(q)=\tau$. The transient response of a pixel $a(\tau)$ can thus be expressed as an integral over all paths that contribute to the irradiance on that pixel:

$$a(\tau)=\int_{\mu(q)=\tau} a(q)\mathrm{d}q. \tag{1}$$

At the sensor, the incident light signal $f_r(\omega t)$ is

$$f_r(\omega t)=E_0+\int_0^{\infty}\alpha(\tau)f_l(\omega(t-\tau))\mathrm{d}\tau, \tag{2}$$

where E_0 is the ambient illumination.

The PMD sensor demodulates the received light signal $f_r(\omega t)$ with the reference signal $f_s(\omega t+\phi)$ and integrates over an exposure time of NT, where N is an integer and $T=2\pi/\omega$. Thus the sensor measurement is

$$H(\omega,\phi)=\int_0^{NT} f_r(\omega t)f_s(\omega t+\phi)\mathrm{d}t. \tag{3}$$

Note that E_0 is dropped due to f_s being zero-mean. Substituting Eq. (2) into Eq. (3), we have [4]

$$H(\omega,\phi)=\int_0^{\infty}\alpha(\tau)c(\omega,\phi,\tau)\mathrm{d}\tau, \tag{4}$$

where

$$c(\omega,\phi,\tau)\overset{\text{def}}{=}\int_0^{NT} f_l(\omega(t-\tau))f_s(\omega t+\phi)\mathrm{d}t \tag{5}$$

is the correlation function between $f_l(\omega t)$ and $f_s(\omega t)$, which is a system function independent of the scene and is pre-calibrated using the method proposed in [5].

The forward formation model (4) can be discretized and written in the compact form as follows:

$$\mathbf{h}=\mathbf{C}\mathbf{i}, \tag{6}$$

where $\mathbf{h}$ is a vector composed of M ToF measurements $H(\omega,\phi)$ under different modulation frequencies and phase values, $\mathbf{i}=[a(\nu),a(2\nu),\ldots,a(L\nu)]^T$ is the discretized transient time profile at a pixel with an ultrashort time interval ν and L samples, which cover the travel time along all the paths, and $\mathbf{C}$ is a $M\times L$ correlation matrix sampling from the calibrated continuous correlation function $c(\omega,\phi,\tau)$.

In the case of specular MPI, the time profile is composed of delta functions, thus its frequency spectrum is band infinite. However, the bandwidth of the camera modulation frequencies is limited due to hardware constraints, and therefore the Eq. (6) is ill-conditioned [5]. Thus a regularization term has to be introduced. As mentioned above, ℓ_0 regularization is intractable, so we adopt nonconvex penalties in sparse signal recovery designed to more democratically penalize nonzero coefficients. Two nonconvex terms, ℓ_p norm $(0<p\le 1)$ and log-sum, are widely used, and by taking the limit $p\to 0^+$ in ℓ_p-norm regularization, we get the log-sum model [10]. It is known that when $p>0$, the closer p approaches to zero, the stronger sparse enhancement that ℓ_p-based optimization exhibits. We also comment here that when p equals zero, the ℓ_p norm exactly corresponds to the intractable discrete ℓ_0-norm problem. Therefore, log-sum exploits the limit of the ℓ_0 norm in the objective and is regarded to have more powerful sparsity enhancement capability than general ℓ_p norm.

The inverse problem of (6) with log-sum regularization is as follows:

$$\begin{aligned}&\min \|\mathbf{i}\|_{ls}\\ &\text{s.t. }\mathbf{h}=\mathbf{C}\mathbf{i}\quad\text{and}\quad \mathbf{i}\ge 0,\end{aligned} \tag{7}$$

where $\|\mathbf{i}\|_{ls}$ is the log-sum regularization defined as $\|\mathbf{i}\|_{ls}=\sum_k \log(|i_k|+\delta)$, $\delta>0$ is a small regularization constant. This optimization problem with constraints can be formulated as the problem without constraints

$$\min_{\mathbf{i}\ge\mathbf{0}}\|\mathbf{h}-\mathbf{C}\mathbf{i}\|_2^2+\lambda\|\mathbf{i}\|_{ls}, \tag{8}$$

where $\|*\|_2$ denotes the ℓ_2 norm, and λ is a positive weight balancing the data error and the sparsity. Equation (8) shows that we seek to minimize an objective function with the constraint that the number of nonzero entries is as small as possible in order to recover the transient time profile. Due to the sparsity of the transient time profile in the time domain, the optimization is robust. We can resolve MPI into its components from the nonzero values in $\mathbf{i}$. Here, we solve Eq. 8 via a majorization-minimization (MM)-type algorithm mentioned in [11,12].

For the comparison of ℓ_1 and log-sum regularization, consider a simple example where

$$\mathbf{i_0}=\begin{bmatrix}0\\1\\0\end{bmatrix},\qquad \mathbf{C}=\begin{bmatrix}2&1&1\\1&1&2\end{bmatrix}.$$

We wish to recover the sparse $\mathbf{i_0}$ from $\mathbf{h}=\mathbf{C}\mathbf{i_0}=[1\quad 1]^T$. The ℓ_1-norm regularization [6] finds an incorrect solution of $\mathbf{i}^*=[\frac{1}{3}\quad 0\quad \frac{1}{3}]^T\neq\mathbf{i_0}$, i.e., the optimal solution to the inverse problem with ℓ_1 regularization is not the answer in this case. In contrast, our log-sum regularization finds the correct solution $\mathbf{i}^*=\mathbf{i_0}$.

We quantitatively evaluate the performance of our method using synthetic data with the range of modulation frequency from 10 to 120 MHz in a step of 1 MHz and the temporal step $\nu=0.01$ ns (about 3 mm). The first synthetic data set is noiseless. Figure 2(a) shows the ground truth transient time profile that consists of three pulses at delay times of 3, 5, and 7 ns with amplitudes of 3, 2, and 1, respectively. This transient model represents a three-path MPI which is common in natural scenes. We generate the acquired measurements $\mathbf{h}$ from Eq. (6) with a calibrated correlation matrix from [5]. Figures 2(b)–2(f) present the recovered time profile by our LSR method,

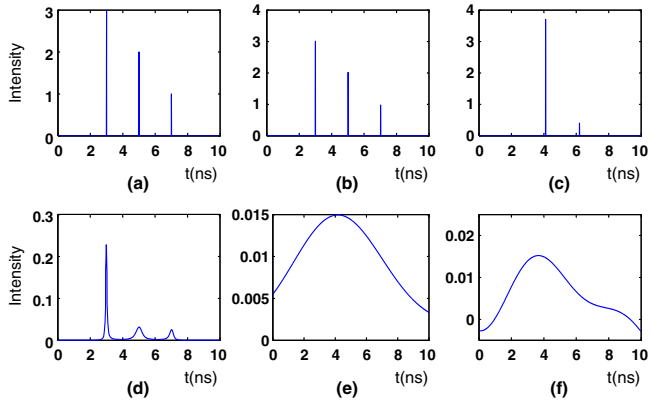

Fig. 2. Decomposition results on noiseless synthetic data. (a) The ground truth of three-path MPI. (b) Decomposition result of LSR. (c) Decomposition result of Bhandari *et al.*'s model [7]. (ℓ_0-norm regularization with 3-sparse knowledge). (d) Decomposition result of Freedman *et al.*'s model [6]. (ℓ_1-norm regularization). (e) Decomposition result of Lin *et al.*'s model [5]. (f) Decomposition result of Heide *et al.*'s model [4].

Bhandari *et al.*'s method (ℓ_0 regularization with 3-sparse knowledge) [7], Freedman *et al.*'s method (ℓ_1 regularization) [6], Lin *et al.*'s method [5], and Heide *et al.*'s method [4], respectively. Compared to the other multi-frequency methods, our LSR method resolves the transient time profile most accurately without the knowledge of the number of nonzero components.

The second synthetic data set is a two-path MPI with Gaussian noise, as shown in Fig. 3(a), used to verify the robustness and the temporal resolution of our method. It is more challenging for transient reconstruction methods

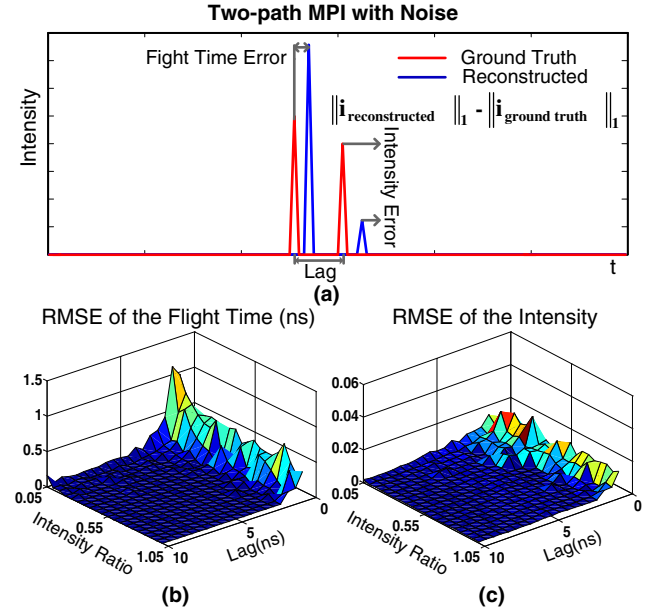

Fig. 3. Simulation of two-path MPI in the presence of noise. (a) The error between the ground truth and the recovered time profile caused by noise. There are two types of error, namely flight time error and intensity error ($\|*\|_1$ denotes the ℓ_1 norm). There is an example of ground truth with the condition of 1 ns lag and 0.8 intensity ratio. (b) and (c) RMSE of the flight time and RMSE of the intensity with our method. Run 20 times at each situation.

to recover the amplitudes and the lag times of pulses in time profiles from noisy data. In this data set, we set the lag between the two-path flight times from 0.5 to 10 ns with a step of 0.5 ns (corresponding to 15 cm in space) and the intensity ratio from 0.05 to 1 with a step of 0.05, and generate 400 groups of data in total. Gaussian noise $N(0, 0.005^2)$ is added to each element of the measurements **h**. The root-mean-squared errors (RMSEs) of the recovered time profiles by our LSR method are illustrated in Figs. 3(b) and 3(c). The results show that, when the lag is too small, the two pulses can not be separated accurately due to the limited resolution of the algorithm; when the lag is greater than 3 ns, RMSE goes down quickly, and our LSR method recovers the time profile accurately. Since the lag may be 5 ns or more in the real world, the reconstruction error is almost zero. Comparison between our LSR method and other reconstruction methods is shown in Fig. 4. The left column illustrates RMSE versus intensity ratio and the right column illustrates RMSE versus lag. Our LSR method yields the least RMSE among these methods. When the intensity difference is wide and the lag is small, it is difficult for all the methods to separate the two pulses. When the intensity ratio is greater than 0.25 and the lag is greater than 3 ns, the RMSE of our LSR method is trivial.

To verify our LSR method in real scenes, we conduct an experiment with the setup shown in Figs. 5(a) and 5(b). This experiment is also applied to show that we can discover useful information hidden in the scene by MPI. The system is based on a PMD PhotonICs 19 k-S3 160×120 sensor array in combination with a bank of three 650-nm laser diodes. Two sinusoidal outputs from an Analog Devices AD9958 function generator chip are converted to a digital square wave and used to drive light source and sensor at frequencies up to 180 MHz [4,5,13]. In this experiment, the modulation frequency is set from 5 to 165 MHz in a step of 1 MHz. The two mirrors introduce MPI into the measurements. The amplitude map and the depth map, shown in Figs. 5(c) and 5(d), correspond to the maximal amplitude and its time of the recovered transient time profile, respectively. MPI introduced by the mirrors makes the scene lit twice in an ultrashort

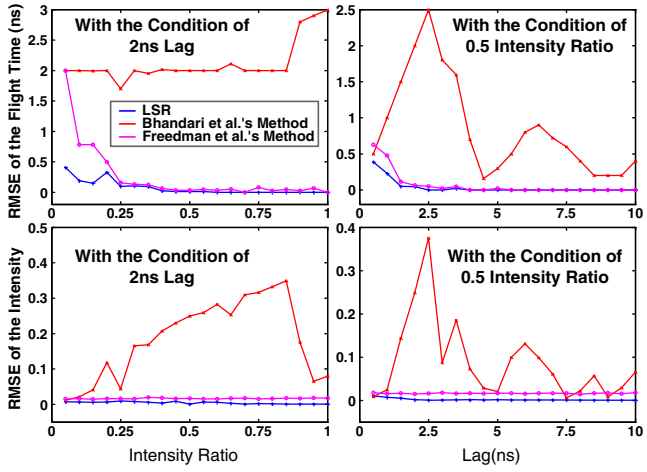

Fig. 4. Comparison of transient time profile reconstruction methods. Run 20 times at each situation.

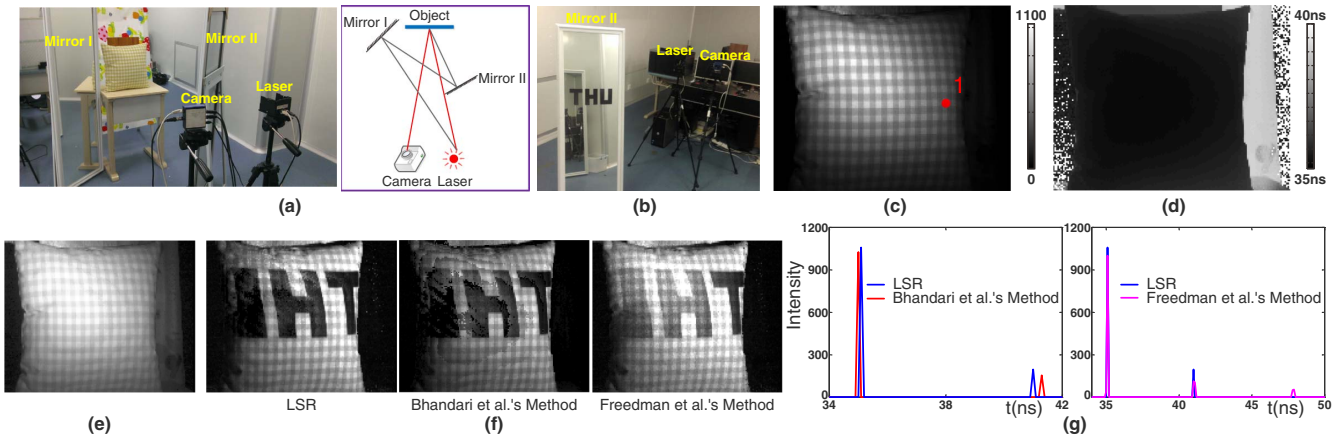

Fig. 5. Reconstructed transient image of the real scene. (a) The experimental setup. (b) The letters "THU." (c) Amplitude map. (d) Depth map. (e) First reflection of the transient image recovered by our LSR method. (f) Second reflections of the recovered transient images. (g) Recovered time profiles at point 1.

time interval, which cannot be sensed by a traditional camera. Thus the letters "THU" sticking on the mirror II, shown in Fig. 5(b), cannot be seen in the amplitude map and the depth map from the position of the camera. When the whole time profile of the optical response is recovered, the information hidden in the second pulse, i.e., the letter on mirror II, can be obtained, as shown in Fig. 5(f). For better comparison, the recovered time profiles of point 1 are shown in Fig. 5(g). Our LSR method detects the number of pulses automatically and decomposes them correctly, whereas Bhandari *et al.*'s method [7] requires the knowledge of the number of pulses, and Freedman *et al.*'s method [6] failed to resolve correct number of pulses. Moreover, in Fig. 5(f), the letters are recovered properly by our LSR method, while artifacts exist in that recovered by Bhandari *et al.*'s method due to intensity error.

In conclusion, we exploit a multi-frequency approach for decomposing specular MPI into its components using log-sum sparse regularization. LSR takes advantage of the strong sparse enhancement introduced by the log-sum term and greatly improves the temporal resolution for resolving MPI. Our LSR method is robust to measurement noise and does not require prior information about the number of pulses. LSR has been experimentally validated on both synthetic and real data, and the results demonstrate its superior performance. Future work includes further improvement of the computation time and the temporal resolution.

This work was supported by the Project of NSFC (Nos. 61327902 and 61120106003).

References

1. S. Fuchs, in *Proceedings of the ICPR* (IEEE, 2010), pp. 3583–3586.
2. S. Fuchs, M. Suppa, and O. Hellwich, in *Proceedings of the ICVS* (Springer, 2013), pp. 31–41.
3. A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, Proc. SPIE **7864**, 786404 (2011).
4. F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, ACM Trans. Graph. **32**, 45 (2013).
5. J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, in *Proceedings of the CVPR* (IEEE, 2014), pp. 3230–3237.
6. D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, in *Proceedings of the ECCV* (Springer, 2014), pp. 234–249.
7. A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, and R. Raskar, Opt. Lett. **39**, 1705 (2014).
8. F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, Opt. Express **22**, 26338 (2014).
9. E. J. Candes and T. Tao, IEEE Trans. Inf. Theory **51**, 4203 (2005).
10. Y. Deng, Q. Dai, R. Liu, Z. Zhang, and S. Hu, IEEE Trans. Neural Netw. Learn. Syst. **24**, 383 (2013).
11. E. J. Candes, M. B. Wakin, and S. P. Boyd, J. Fourier Anal. Appl. **14**, 877 (2008).
12. C. S. Foo, C. B. Do, and A. Y. Ng, in *Proceedings of the ICML* (IEEE, 2009), pp. 321–328.
13. F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, in *Proceedings of the CVPR* (IEEE, 2014), pp. 3222–3229.