Research Article Vol. 31, No. 5 / 27 Feb 2023 / Optics Express 7303
Optics EXPRESS

TDI-based continuous window compressed spatio-temporal imaging capable of flexible voxels post-interpretation

YUN-HUI LI,[1,2,3,*] XIAO-DONG WANG,[1,3] AND WEN-GUANG LIU[1,3]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]Key Laboratory of Space Photoelectric Detection and Perception (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, Nanjing 211106, China
[3]Key Laboratory of On-orbit Manufacturing and Integration for Space Optics System, Chinese Academy of Sciences, Changchun 130033, China
*liyunhui@ciomp.ac.cn

Abstract: To achieve high frame rates and continuous streaming simultaneously, we propose a compressed spatio-temporal imaging framework implemented by combining time-delay-integration sensors and coded exposure. Without additional optical coding elements and subsequent calibration required, this electronic-domain modulation enables a more compact and robust hardware structure, compared to the existing imaging modalities. By exploiting the intra-line charge transfer mechanism, we achieve a super-resolution in both temporal and spatial domains, thus multiplying the frame rate to millions of frames-per-second. In addition, the forward model with post-tunable coefficients, and two reconstruction strategies proposed therefrom, facilitate a flexible voxels post-interpretation. Finally, the effectiveness of the proposed framework is demonstrated by both numerical simulations and proof-of-concept experiments. With the prominent advantages of prolonged time window and flexible voxels post-interpretation, the proposed system will be suitable for imaging random, non-repetitive, or long-term events.

1. Introduction

To visualize various instantaneous phenomena and fast dynamic processes, there is an ever-growing performance demand for high-speed imaging systems. A higher temporal resolution is highly desired for all sorts of scientific research and industrial applications [1–6]. However, the frame rate is generally limited to tens of frames-per-second by image sensors in conventional systems. With the advantages of high integration and parallelized quantization, some CMOS sensors can reach hundreds of frames-per-second, which can be further improved by sacrificing spatial resolution through the region of interest (ROI) operation. For those high-speed systems using image sensors specially designed for high frame rates, on-board digital caching is a solution to deal with the transmission pressure caused by large data flux. Further, in ultra-high-speed imaging beyond the quantization rate, the analog charge will be cached in the in-pixel storage through a special sensor structure, but at the expense of degraded spatial resolution and filling factor [7–10]. Whether by means of on-board memory or in-pixel storage, there is a limited recording time capacity.

In addition to the efforts at the sensor end, under the conventional imaging framework, an alternative path to increase the frame rate will focus on the system modality. The camera array comes to mind first. With a fast rotating mirror prism sweeping an array of 128 cameras, allowing up to 25 million frames-per-second with a full sequence of 128 [11]. Using a stationary beam-splitter instead, along with the image intensifier gated, the frame rate can be increased to 200 million frames-per-second [12]. By staggering each camera's exposure window, a dense array of cameras geometrically aligned is demonstrated for capturing thousands of frames-per-second

#481104 https://doi.org/10.1364/OE.481104
Journal © 2023 Received 17 Nov 2022; revised 20 Jan 2023; accepted 21 Jan 2023; published 14 Feb 2023

Optics EXPRESS

with continuous streaming supported [13]. Commonly, this hardware stacking in exchange for performance improvement has resulted in a bulky architecture, as well as complicated calibration. Another burst capturing modality is sequentially timed all-optical mapping photography (STAMP), which equivalently achieves trillions of frames-per-second by temporal and spatial separation, yet falls short in number of frames [14]. By mapping the spatial image into a serial time-domain data stream, the serial time-encoded amplified microscopy (STEAM) enables a continuous operation at millions of frames-per-second [15]. However, it encodes the image into the spectrum of a broadband pulse, sacrificing the spectral information, and thus depriving its ubiquitous adaption. Despite being accessible to hundreds of billions of frames-per-second, the streak camera can only provide one-dimensional images, as well as a narrow time window [16].

In summary, within the conventional imaging framework, high-speed imaging always leads to either a bulky and complicated system, or a shallow sequence depth, or a degraded spatial resolution. An approach to circumvent these problems is the compressed sensing-based imaging framework, in which dynamic events can be reconstructed from the modulated sampling data that is far less than required by Nyquist [17–20]. This low sampling rate reduces the frame rate and pixel count required for the image sensor, and thus slashing the data flux by a large amount, so that both high frame rate and continuous time window become feasible within a compact system. Moreover, by means of multiplexed sampling and applying specific priori in reconstruction, its immunity to noise is enhanced, which is critical in high-speed imaging due to insufficient exposure. However, the dynamic scene needs to be modulated controllably before being projected to the sensor, which introduces an additional operation in the optical path, usually implemented with mechanical or time-varying optical elements. Typically, mechanical translation of a passive coded aperture by the piezoelectric stage is employed in coded aperture compressed temporal imaging (CACTI) [21,22], and it can be improved by a more flexible spatial light modulator (SLM) [23–26]. These elements modulate the incident scene at a rate faster than the sensor, thereby multiplying the achievable frame rate. However, limited by their refreshing rate, the frame rate is generally clamped at several thousand frames-per-second. In contrast, the galvanometer scanner can achieve millions of frames-per-second through temporal shearing. But confined by the sensor's frame rate, only single-shot operation is available, depriving the continuous window capability [27].

All these physical-based approaches rely on relatively significant modifications or additions to the imaging system, as well as complicated calibrations. By comparison, without employing any mechanical or optical scanning device, the electronic-domain approaches allow a more compact system, with the advantages of high speed and high accuracy. Temporal shearing by a streak camera, compressed ultrafast photography (CUP) enables an acquisition rate of up to 100 billion frames-per-second [28]. Intermittent exposure or pixel-wise exposure realized in specially manufactured CMOS sensors are also effective means of spatio-temporal modulation, capable of hundreds of millions of frames-per-second [29,30]. Limited by the electronic readout speed of the sensor, these approaches do not have continuous acquisition capability.

The rolling-shutter operation, which is commonly used in CMOS sensors, theoretically has a temporal resolution of several microseconds by leveraging the subtle time delay between adjacent lines, and the data can be continuously read out in lines. However, its mechanism determines that there is no spatial multiplexing, so essentially, it is not a quite efficient modulation, although several systems based on it have been proposed [31–33]. With the same temporal resolution and continuous readout capability, the time-delay-integration (TDI) operation multiplexes temporal and spatial information, enabling a more efficient acquisition. By utilizing a pseudorandom binary mask and a TDI camera, the high-speed system developed by J. Park et al. can record the dynamic events at a 200kHz frame rate in the continuous streaming mode [34]. In addition, restricted by their stationary optical frameworks, the mapping relationship between the data cube voxels and the image sensor pixels is fixed in the existing systems. Therefore, for better performance, the

Optics EXPRESS

flexible voxels post-interpretation will be a remedy for non-repetitive or difficult-to-reproduce events, which is appreciated [35,36].

To achieve high frame rates and continuous streaming simultaneously, with voxels post-interpretation considered, in this paper, we propose a compressed spatio-temporal imaging framework implemented by combining TDI sensors and coded exposure, enabling a continuous sampling at millions of frames-per-second. Compared to the existing high-speed imaging modalities, this electronic-domain approach promotes a more efficient and flexible modulation implementation. By leveraging the on-chip TDI operation, the spatio-temporal multiplexing no longer requires additional optical coding elements and subsequent calibration, resulting in a more compact and robust hardware structure. Furthermore, with regular TDI as the baseline, both temporal resolution and spatial resolution are subdivided by exploiting the intra-line charge transfer mechanism, thereby multiplying the achievable frame rate. In addition, the continuous streaming is arbitrarily separable, which ensures that segments of interest can be selectively reconstructed, with significantly reduced computational complexity and memory pressure in reconstruction. Finally, the forward model with post-tunable coefficients, and two reconstruction strategies proposed therefrom, facilitate a flexible voxels post-interpretation.

Our contributions can be summarized as follows:

1) We propose a compressed spatio-temporal imaging framework based on TDI combined with coded exposure, and specify two system architectures. Compared to conventional TDI-based systems, it can achieve super-resolution in both temporal and spatial domains, resulting in millions of frames-per-second with continuous streaming capability.

2) Given the same sampling results, we realize a flexible voxels post-interpretation through the forward model with post-tunable spatio-temporal merging coefficients. Two reconstruction strategies are then derived, which can flexibly specify the spatial and temporal resolution for the target concerned.

With the prominent advantages of prolonged time window and flexible voxels post-interpretation, the proposed system will be suitable for imaging random, non-repetitive, or long-term events.

2. System architecture and modulation principle

TDI operation was originally designed to improve the sensor's sensitivity in dynamic scene applications. For this purpose, the charge line transfer needs to be synchronized with the image motion to achieve staring imaging equivalently, thus extending the integration time without causing blur. For a general scene without image motion matching constraint, the charge packet transfers one line at a time in a stepwise manner, as shown in Fig. 1(a), and keeps continuously exposed throughout the transfer process. After performing all the line transfer steps, namely the integral stages, each charge packet to be read out contains both temporal information and one-dimensional spatial information along the TDI direction. In this case, the spatial resolution is determined by the pixel size, and the temporal resolution depends on the line transfer rate, which can generally reach hundreds of thousands of frames-per-second. This off-the-shelf sensor performs spatio-temporal multiplexing by leveraging the on-chip TDI operation, with high temporal resolution and continuous time window capability, making it a high-quality candidate for compressed spatio-temporal imaging.

Furthermore, by insight into the driving timing of TDI, we find that the charge packet inter-line transfer is realized through a relay of multiple driving signals. Taking the three-phase TDI sensor as an example, the line transfer process consists of six subdivided phases. Therefore, this intra-line transfer can be reorganized into multiple sub-steps, each of which moves a fraction of the pixel, and a continuous transfer mode is adopted to provide equal time intervals for each sub-step. The size of the charge packet within the sensor, as well as its readout and quantization operation,

Optics EXPRESS

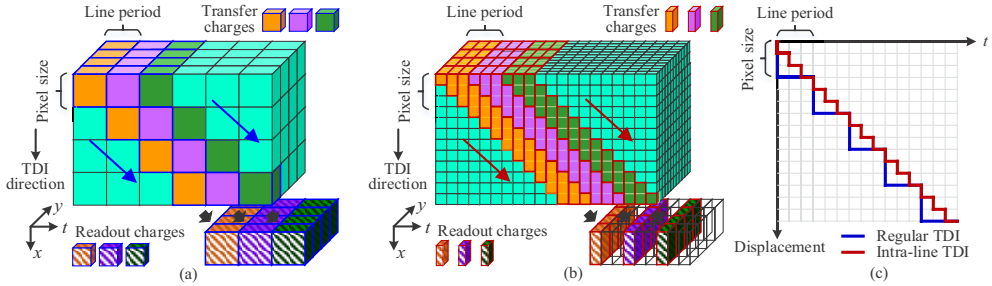

Fig. 1. Illustration of how TDI works. (a) and (b) show the inter-line and intra-line charge transfer and readout processes, respectively. (c) Comparison of charge displacement between these two cases.

determines that the sampling at each sub-step remains implemented in full pixels. As shown in Fig. 1(b), the inter-line charge transfer is refined into three sub-steps, with each displacement of 1/3 pixel, and the transfer trajectories of three adjacent charge packets are represented in three distinguished colors. The charge transfer displacements in these two cases are compared in Fig. 1(c), which shows that the intra-line charge transfer also synchronously subdivides the temporal resolution into fractional line periods. In this way, with the same sensor architecture and data flux, we have further improved the spatial and temporal resolution simultaneously, as shown in Fig. 1(b), compared to Fig. 1(a) for the division of the spatio-temporal data cube.

However, information multiplexing alone is insufficient, and it needs to be supplemented by coding modulation. Based on the intra-line TDI, we propose a compressed spatio-temporal imaging framework, which is implemented by combining coded exposure and TDI. The intra-line charge transfer depicted in Fig. 1(b) multiplies both the one-dimensional spatial resolution and temporal resolution of the spatio-temporal data cube, while the spatial resolution perpendicular to the TDI direction remains at the pixel level. To this end, we present a dual-arm system as shown in Fig 2(a), which subdivides the remaining spatial dimension through two orthogonally installed TDI cameras, and also doubles the sampling rate. The system consists of a coding light source, an objective lens, a beam splitter, two relay lens and TDI cameras, a synchronization trigger, and the data acquisition and reconstruction computer. After being reflected by the target scene, the coded beam is incident to the beam splitter through the objective lens, and then equally divided into two parts, which are finally projected to the TDI cameras through the relay lenses, respectively. The two TDI cameras are installed orthogonally and maintain sub-pixel geometric registration.

Directed by the data acquisition and reconstruction computer, the synchronization trigger drives the light source to flash with a preset coding sequence, which is a pseudo-random binary one composed of 0's and 1's, and synchronously generates the TDI intra-line transfer signal with the same period. The data acquisition and reconstruction computer captures the TDI cameras' data in real time, and then implements video reconstruction in accordance with the selected time segment and spatio-temporal resolution requirements.

Figure 2(b) presents an alternative compact version within the same principle framework. The coded beam reflected by the target scene is incident to two Dove prisms through the objective lens, and then projected to two separate areas of the TDI camera through the relay lenses, respectively. To achieve two orthogonal projections, these two Dove prisms have a relative rotation angle of 45° along the axis. Compared with the former, it only uses one TDI camera to realize sampling in both two orthogonal directions, fully utilizing the horizontal pixel count margin, and thus has the advantages of lower cost and smaller volume. The above two systems have the same modulation effect on the spatio-temporal data cube, which is depicted in Fig. 3. The original spatio-temporal

Optics EXPRESS

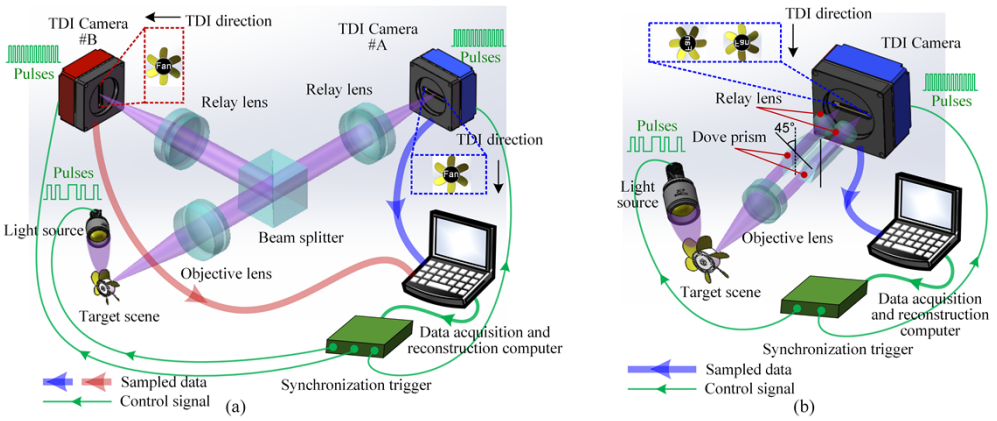

Fig. 2. The hardware composition of the proposed system. (a) The dual-arm version, in which the incident beam is divided into two parts through a beam splitter and projected onto two orthogonally installed TDI cameras, respectively. (b) An alternative compact version, which projects two orthogonal images to separate areas of the TDI camera by leveraging the image rotation function of Dove prism.

data cube is modulated in temporal domain through coded exposure, and then spatially sheared through TDI in the horizontal and vertical directions, respectively. Finally, these projections are merged at a pixel-wise scale to form the sampled data.

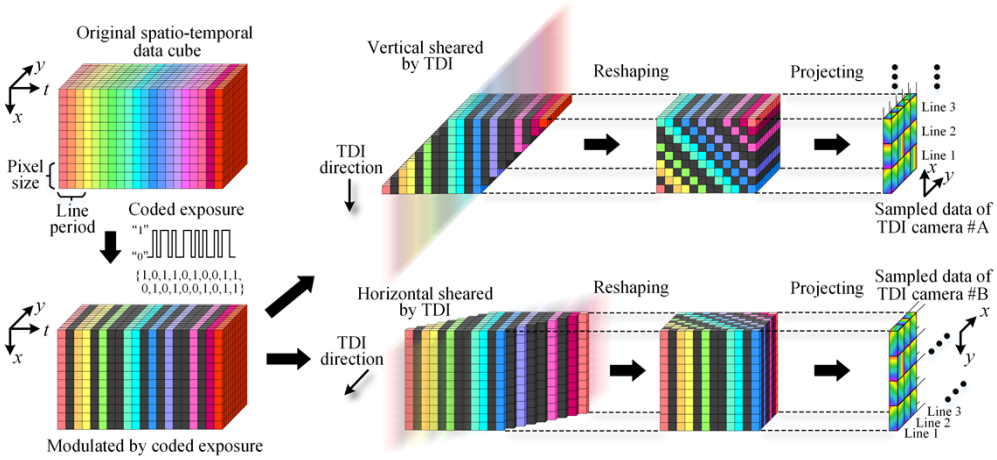

Fig. 3. The modulation effect of the proposed system on the spatio-temporal data cube, which is modulated in the temporal dimension by coded exposure, and then spatially sheared by TDI in the vertical and horizontal directions, respectively. Finally, they are projected to the sensors and spatially merged by pixel sampling to obtain the sampled data.

Temporal coding is selected here instead of spatial coding. Actually, the projections along highly redundant dimensions, such as time or spectrum, are preferred by most high-dimensional compressed imaging systems. In this case, spatial coding is a more general and effective solution, and temporal coding is not feasible. While the projection under the TDI mechanism is along the spatio-temporal direction, as shown in Fig. 1, that is, the sampling includes both temporal and spatial multiplexing. This particularity determines the validity of temporal coding, even

Optics EXPRESS

if the one-dimensional coding it presents is still with a poor capability. Benefiting from this dual-arm architecture, temporal coding can approximate quasi-two-dimensional coding through orthogonal projection. Although this is still inferior in coding capability compared to spatial coding, it has achieved comparable results. In addition to performance considerations, we are committed to removing the spatial coding elements from the system to obtain a more compact and robust hardware structure. Therefore, balancing system performance and implementation, we ended up choosing the current solution.

Compared with the existing compressed imaging systems, the advantages of the proposed one are summarized as follows: First, the electronic-domain modulation based on TDI sensors combined with coded exposure simplifies the system architecture and promotes a more efficient and flexible modulation implementation. In addition, no optical coding element required, eliminating the need for calibration, which is beneficial to improve the robustness. Second, the modulation based on intra-line TDI has both high-speed and continuous streaming capability, thus having an infinitely prolonged time window, which is suitable for recording random, or long-term events. Moreover, compared with the regular TDI, both temporal resolution and spatial resolution are further multiplied. Third, its spatio-temporal voxels post-interpretation enables a more flexible reconstruction for the target scene with the given sampled data. This will be described in detail in the following section.

3. Forward model and reconstruction strategy

3.1. Forward sensing model with variable coefficients

Referring to the visual modulation effect presented in Fig. 3, we establish a mathematical model to describe the modulation and acquisition process of the proposed framework on the spatio-temporal data cube. With the reference coordinates defined in Fig. 3, the sampling result $Y(i, j)$, where (i, j) are the row and column coordinates, can be expressed as follows:

$$Y(i,j) = \sum_{c=(j-1)\cdot S+1}^{j\cdot S} \left(\sum_{u=0}^{S-1} \left(\sum_{r=1}^{M\cdot S} X(r, c, r+u+(i-1)\cdot S) \cdot E(r+u+(i-1)\cdot S) \right) \right) + \delta(i,j), \quad (1)$$

where M is the TDI integral stages, and S is the TDI intra-line steps, namely the theoretical spatial super-resolution times. $X \in \mathrm{R}^{(M\cdot S)\times(M\cdot S)\times T}$ represents the target spatio-temporal data cube, $E \in \mathrm{R}^{T}$ represents the coded exposure sequence, and $\delta \in \mathrm{R}^{L_R \times M}$ signifies the measurement noise . T is the time span of the reconstructed video, there is $T = S \times (M + L_R)$-1, where L_R is defined as the number of lines read out by each TDI camera. The physical meaning of Eq. (1) is that the innermost summation represents the regular TDI operation, the middle and the outermost summation represent the merging operation along the TDI direction and perpendicular to the TDI direction, respectively. Note that the latter two are mandatory because the sensor is always read out in full pixels, with no fractional pixels accessible.

Derived from Eq. (1), the sampling results of both TDI cameras in the proposed system can be reformulated into a matrix representation:

$$\begin{cases} Y_A = R \cdot D_A \cdot X_A \cdot C = A_A \cdot X_A \cdot C \\ Y_B = R \cdot D_B \cdot X_B \cdot C = A_B \cdot X_B \cdot C \end{cases}, \quad (2)$$

where D_A, $D_B \in \mathrm{R}^{(L_R\cdot S)\times(M\cdot S\cdot T)}$ are TDI operation matrices. Both $R \in \mathrm{R}^{L_R \times (L_R\cdot S)}$ and $C \in \mathrm{R}^{(M\cdot S)\times M}$ are merging matrices, which realize spatial merging along the TDI direction and perpendicular to the TDI direction, respectively. To simplify the model representation, define R multiplied by D_A, D_B as the projection matrix A_A, $A_B \in \mathrm{R}^{L_R \times (M\cdot S\cdot T)}$. Y_A, $Y_B \in \mathrm{R}^{L_R \times M}$ are the sampling result matrices, and the target spatio-temporal data cube is flattened into two-dimensional matrices X_A, $X_B \in \mathrm{R}^{(M\cdot S\cdot T)\times(M\cdot S)}$ along the temporal dimension.

Optics EXPRESS

Each matrix in Eq. (2) is visualized in Fig. 4. The projection matrix A is sparse with a fixed structure, which is jointly determined by the TDI stages M and intra-line steps S, and its internal elements are composed of exposure sequences. The spatial merging matrix C has a diagonal-like structure, which is also determined by the TDI stages and intra-line steps.

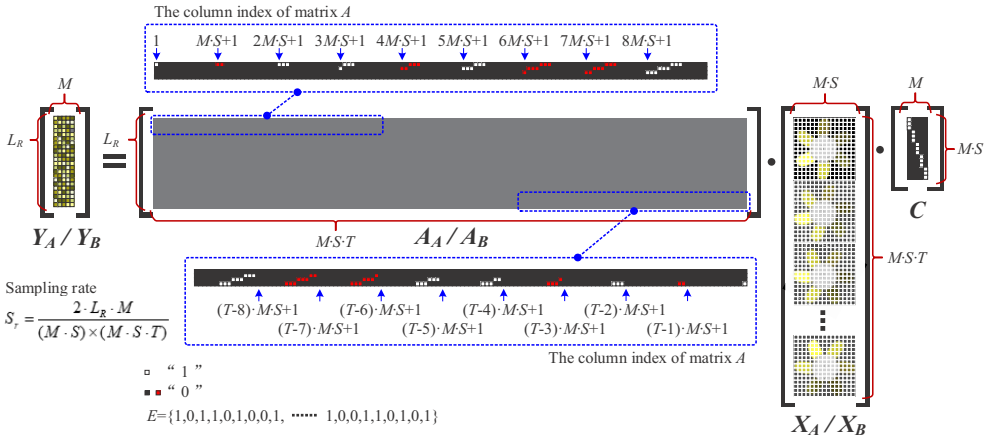

Fig. 4. A graphical representation of the matrix in the forward model. In the two dotted boxes are the enlarged display of local areas in the upper left corner and lower right corner of matrix A, respectively. The target spatio-temporal data cube is flattened into a two-dimensional matrix X along the temporal dimension.

The formula for calculating the sampling rate S_r of the above model is given in Fig. 4. When $L_R \to \infty$, there is $S_r \to 2/(M \cdot S^3)$, which means that its upper limit is extremely low, so it would be quite difficult to reconstruct high-quality video frames. Therefore, we consider downscaling the model in the spatial or temporal dimensions to improve the sampling rate. Then, the forward model described in Eq. (2) can be transformed into a spatio-temporal downscaled version:

$$\begin{cases} Y_A = A_A \cdot M_T \cdot M_N \cdot X'_A \cdot C_N \\ Y_B = A_B \cdot M_T \cdot M_N \cdot X'_B \cdot C_N \end{cases}, \tag{3}$$

where $M_T \in \mathrm{R}^{(M \cdot S \cdot T) \times (M \cdot S \cdot T/Tm)}$ is the temporal merging matrix, and $M_N \in \mathrm{R}^{(M \cdot S \cdot T/Tm) \times (M \cdot S/Nm \cdot T/Tm)}$ is the spatial merging matrix. They are used to merge the time-related and space-related columns of projection matrix A for downscaling operation, respectively. $C_N \in \mathrm{R}^{(M \cdot S/Nm) \times M}$ is a matrix that compensates for the spatial scale difference between the sampling results and the merged target in the direction perpendicular to TDI. X'_A, $X'_B \in \mathrm{R}^{(M \cdot S/Nm \cdot T/Tm) \times (M \cdot S/Nm)}$ are the target matrices with downscaled temporal and spatial resolution. T_m is the temporal merging coefficient, and N_m is the spatial merging coefficient, indicating the descending times in temporal and spatial resolution, respectively.

According to the geometric relationship between the two TDI cameras, if the target matrix X'_B is defined as a frame-by-frame transpose of X'_A, that is, $X'_B = \mathrm{transpose}(X'_A)$ by frame of size $(M \cdot S)$, then the projection matrix A_B is a group-by-group reversed column order of A_A, that is, there is $A_B = \mathrm{fliplr}(A_A)$ by group of size $(M \cdot S)$. In this case, the two discrete parts in Eq. (3) can be integrated into a standard compressed sensing model:

$$\begin{pmatrix} \mathrm{vec}(Y_A) \\ \mathrm{vec}(Y_B) \end{pmatrix} = \begin{pmatrix} ((A_A \cdot M_T \cdot M_N) \otimes C_N) \cdot P \\ (A_B \cdot M_T \cdot M_N) \otimes C_N \end{pmatrix} \cdot \mathrm{vec}(X'_B) \text{ (abbreviated as } Y_V = \Lambda' \cdot X'_V), \tag{4}$$

Optics EXPRESS

where $\otimes$ denotes the Kronecker product operation, and vec$(\cdot)$ denotes the vectorization operation of splicing matrix by row. $P \in \mathrm{R}^{(M^2 \cdot S^2 / Nm^2 \cdot T/Tm) \times (M^2 \cdot S^2 / Nm^2 \cdot T/Tm)}$ is a remapping matrix, which is responsible for the mapping operation from vec(X'_B) to vec(X'_A), that is, vec(X'_A)=P·vec(X'_B). Its abbreviated expression is included in the right bracket, where $Y_V \in \mathrm{R}^{(2L_R \cdot M) \times 1}$ is the result vector, $\Lambda' \in \mathrm{R}^{(2L_R \cdot M) \times (M^2 \cdot S^2 / Nm^2 \cdot T/Tm)}$ is the sensing matrix, and $X'_V \in \mathrm{R}^{(M^2 \cdot S^2 / Nm^2 \cdot T/Tm) \times 1}$ is the target vector in this case.

For the forward sensing model described in Eq. (4), we can reconstruct the video frames by solving the following optimization problem:

$$\hat{X}'_V = \arg\min_{X'_V} \left\{ \frac{1}{2} ||Y_V - \Lambda' X'_V||_2^2 + \lambda \varphi(X'_V) \right\}, \tag{5}$$

where $||\cdot||_2$ denotes the ℓ_2 norm. The first quadratic term enforces the fidelity between the measurements and the estimated results. The second term is a regularization function $\varphi(\cdot)$, and λ is its weight coefficient. To solve Eq. (5), we adopt a total variation (TV) regularizer and a two-step iterative shrinkage thresholding (TwIST) algorithm [37].

3.2. Reconstruction strategy

The proposed framework performs a high-speed continuous sampling process, thus recording the entire event with an infinitely extended time window. In response, we segment the sampling results and reconstruct the scene in a piecewise manner, which has the following advantages: First, segmentation reduces the matrix size and computational complexity, thus accelerating reconstruction. Second, multiple segments can be reconstructed in parallel without interfering with each other. Third, the reconstruction can be targeted to the specific segments of interest to avoid wasting computing resources. In addition, this approach also brings higher flexibility in reconstruction, that is, given the sampling results, the reconstruction parameters can be freely post-tuned based on the variable model described in Section 3.1, so as to achieve flexible voxels post- interpretation. Based on the fact that a certain sampling rate should be guaranteed for a successful reconstruction, it is impractical to capture high-speed and high-resolution scenes simultaneously without compromising either of them, so a trade-off between temporal and spatial resolution is inevitable.

Thus, two reconstruction strategies are derived, which are called selective reconstruction and combinatorial reconstruction, respectively. The former is shown in the left half of Fig. 5. In view of the difference in characteristics of the concerned targets, the merging coefficients are selectively set in the model for the same time segment. For high dynamic targets, a small temporal merging coefficient is adopted to ensure high frame rate. While for the targets with high spatial resolution requirements, a small spatial merging coefficient is used to facilitate the capture of detailed information. Note that although only the parallelepiped shown in Fig. 5 has been effectively sampled, the cuboid span from t_0 to t_3 is still taken as the reconstruction target, thus preserving the integrity of the spatio-temporal data cube. In this case, certain areas at both ends are excluded from the reconstruction results for credibility, as shown in Fig. 5, only the part $t_1 \sim t_2$ is retained after reconstruction with the sampling results of $t_0 \sim t_3$.

The other combinatorial reconstruction strategy is shown in the right half of Fig. 5, which splits a complex scene into several sub-segments. With the guidance of small temporal merging coefficient corresponding to high dynamic targets and small spatial merging coefficient corresponding to the targets with high spatial resolution requirements, each sub-segment has its distinguish setting on the merging coefficients. Then, they are reconstructed separately and integrated into a complete video. Note that a certain overlapping area between adjacent parallelepipeds should be ensured for a continuous video streaming. As shown in Fig. 5, the first sub-segment is reconstructed with the sampling results of $t_4 \sim t_8$, and the second sub-segment is reconstructed with the sampling

Optics EXPRESS

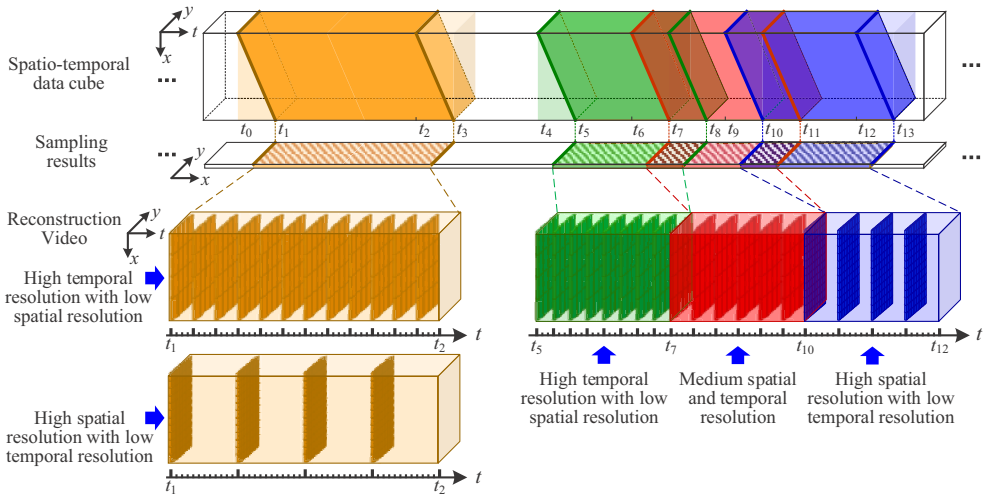

Fig. 5. Schematic diagram of the image reconstruction strategy, which shows two cases. In one case, as shown on the left side, flexible temporal and spatial merging coefficients can be selected for reconstruction for the same sampling process. It is noted that in order to ensure the effectiveness of reconstruction, only the part t_1~t_2 is retained after reconstruction with the sampling results of t_0~t_3. The other case is shown on the right side. A piece of video can be segmented and reconstructed separately with different temporal and spatial merging coefficients. The first segment is reconstructed with the sampling results of t_4~t_8, in which the part t_5~t_7 is retained. The second segment is reconstructed with the sampling results of t_6~t_{11}, in which the part t_7~t_{10} is retained. The third segment is reconstructed with the sampling results of t_9~t_{13}, in which the part t_{10}~t_{12} is retained. In this way, we obtain the continuous reconstructed video streaming of t_5~t_{12}.

results of t_6~t_{11}. In this way, we get the reconstruction results of t_5~t_7 and t_7~t_{10}, respectively, thus realizing a seamless connection.

4. Numerical results

In the TwIST algorithm implementation, to make full use of the temporal redundancy in the wide time window, we first sparse the temporal dimension with a discrete cosine transformation (DCT) base before adopting the TV denoiser. The TV regularization function is of an isotropic type and oriented towards two-dimensional images. Therefore, the spatial image slices in the video will be de-noised separately. Like most other algorithms, as a hyper-parameter, the weight coefficient λ needs to be manually fine-tuned according to the scene characteristics and spatial resolution, and the larger it is, the smoother the image and vice versa. Its value involved in all the following experiments is set in the range of 0.1 to 10. On the premise of the segment selected in the experiment and the computer configuration described in Section 5.1, it takes about ten more minutes to complete a reconstruction, and a larger merging coefficient means a significantly shorter time. Generally, it will converge within 200 iterations.

To verify the effectiveness of the proposed system, we first carried out numerical simulations. In order to quantify the improvement in spatial and temporal resolution brought by intra-line TDI, we designed a simulation experiment to explore the upper limit of the proposed system's spatial and temporal resolution, taking inter-line TDI as the baseline for comparison. In the following two evaluations, the scene is set to have a spatial resolution of 192×192, and a temporal resolution of 980, with a time span of 0.54 ms in the case of 1,800,000 frames-per-second.

Optics EXPRESS

In the evaluation of spatial resolution, the USAF 1951 test target remains static in order to avoid interference caused by the spatial displacement. The results obtained by inter-line TDI and intra-line TDI, both based on the dual-arm architecture, are shown in Fig. 6(a) and Fig. 6(b), respectively. To facilitate quantitative analysis, the contrast ratios, covering all elements in Group2 and Group3, are plotted in Fig. 6(c). Compared to a mosaic effect in the former, the intra-line TDI enables a significantly finer image, and it can recognize up to the fifth element in Group3, with a contrast of 0.1 as the threshold.

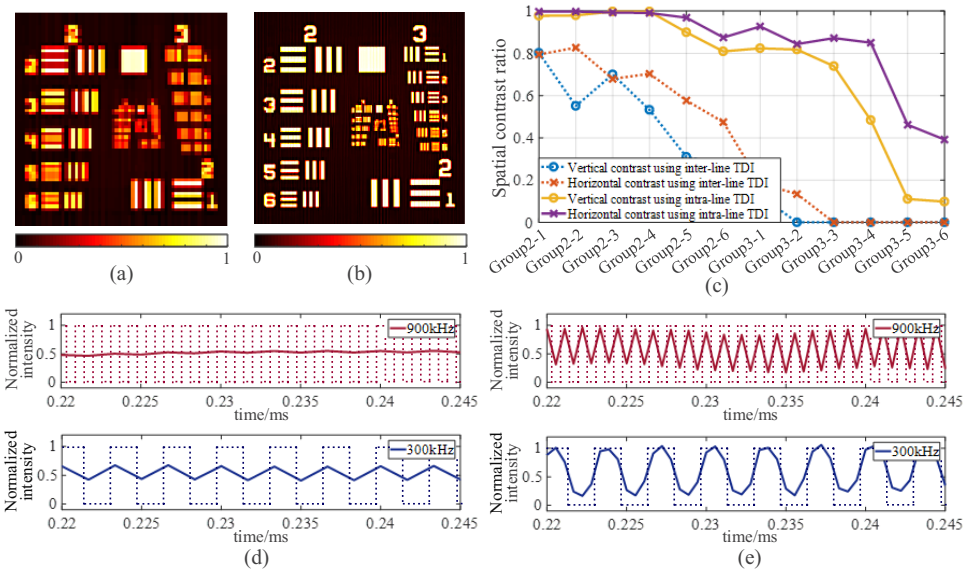

Fig. 6. Evaluation results of spatial and temporal resolution. In the evaluation of spatial resolution, (a) and (b) are the results obtained by inter-line TDI and intra-line TDI, respectively, and their contrast ratios are plotted in (c). In the evaluation of temporal resolution, the intensity response curves of inter-line TDI and intra-line TDI are plotted in (d) and (e), respectively. Note that the dotted line in (d) and (e) represents the theoretical pulse intensity.

In the evaluation of temporal resolution, the scene also remains static but with intensity fluctuations, which are modulated by pulses of 300kHz and 900kHz, respectively, and the former is the Nyquist frequency referring to the inter-line TDI process. Fig. 6(d) and Fig. 6(e) are the response curves of inter-line TDI and intra-line TDI to these two pulses, respectively, which are defined as the average intensity over time. In contrast, the pulse in Nyquist frequency can be captured by intra-line TDI with a higher fidelity. It can even capture intensity fluctuations at 900kHz.

Following the above reconstruction strategy, we specifically designed two different scenes. One is a scene that contains quasi-static target and high dynamic target simultaneously to verify the selective reconstruction strategy. The other is that there are targets with various dynamic attributes distributed in different sub-segments of the time segment concerned, which is used to verify the combinatorial reconstruction strategy.

In order to reduce the computational complexity and memory pressure in the video reconstruction, while retaining sufficient sampling intensity, the first goldfish scene is set to have a spatial resolution of 162×162, and a temporal resolution of 4100, with a time span of 2.28 ms in the case of 1,800,000 frames-per-second. This corresponds to a hardware architecture where the TDI integral stages M is 54 with an intra-line steps S of 3, and the line rate is 600kHz, which is the highest one available to date. The motion trajectories of the two goldfish are shown in

Optics EXPRESS

Fig. 7(a), where the broad-finned fish makes a fast circular swimming counterclockwise, while the zebra fish rotates slowly clockwise.

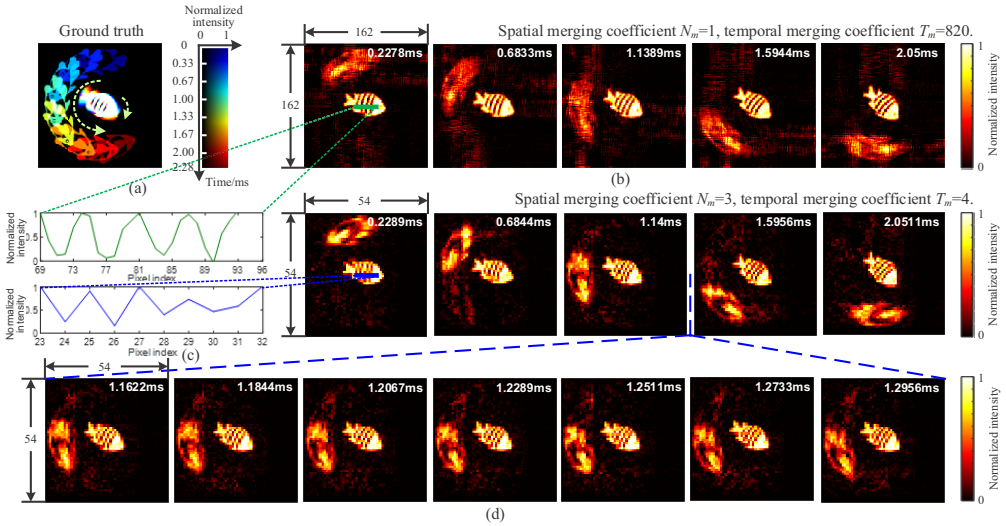

Fig. 7. Simulation results for the goldfish scene. The trajectories of the two goldfish are shown in (a), where the broad-finned fish makes a fast circular swimming counterclockwise, while the zebra fish rotates slowly clockwise. When the merging coefficients are set to $N_m = 1$ and $T_m = 820$, the reconstructed video frames are shown in (b), in which the zebra fish's stripes can be clearly discerned, while the broad-fined fish is severely motion blurred. When the merging coefficients are set to $N_m = 3$ and $T_m = 4$, the reconstructed video frames are shown in (d). Compared to the frames at the same time as the former, the broad-fined fish can be clearly captured, but the contrast of zebra fish's stripes degrades, which can also be confirmed by the normalized intensity curves depicted in (c). Note that the time is marked in the upper right corner of each reconstructed video frame. See Visualization 1 for the full video.

When the merging coefficients are set to $N_m = 1$ and $T_m = 820$, the reconstructed video frames are shown in Fig. 7(b). Since the spatial resolution remains at 162×162, the zebra fish's stripes can be clearly discerned. However, the temporal resolution is degraded significantly due to a large temporal merging coefficient, resulting in the broad-finned fish being severely motion blurred. With the merging coefficients set to $N_m = 3$ and $T_m = 4$, the reconstructed video frames are shown in Fig. 7(d). Compared to the frames at the same time as the former, the high temporal resolution ensures that the motion trajectory of broad-fined fish can be clearly captured. Along with this, however, the spatial resolution drops to 54×54, degrading the contrast of zebra fish's stripes, which can also be confirmed by the normalized intensity curves depicted in Fig. 7(c). To sum up, given the sampling results, the first set of coefficients is preferred to explore the spatial information of zebra fish in detail, while the second one is suitable for capturing the time-varying information of broad-fined fish.

It is noted that compared with the results in Ref. [34], more scene contents are included here at a lower spatial resolution, and the sampling rate is further significantly reduced due to the spatio-temporal subdivision. For these two reasons, T_m in the second set of coefficients is kept at least 4 to maintain a considerable reconstruction quality. The lower T_m values will be verified in the following experimental section.

With the same spatial resolution and frame rate as the above scene, as well as the hardware configuration, the second dynamic scene consists of three time sub-segments. To match the

Optics EXPRESS

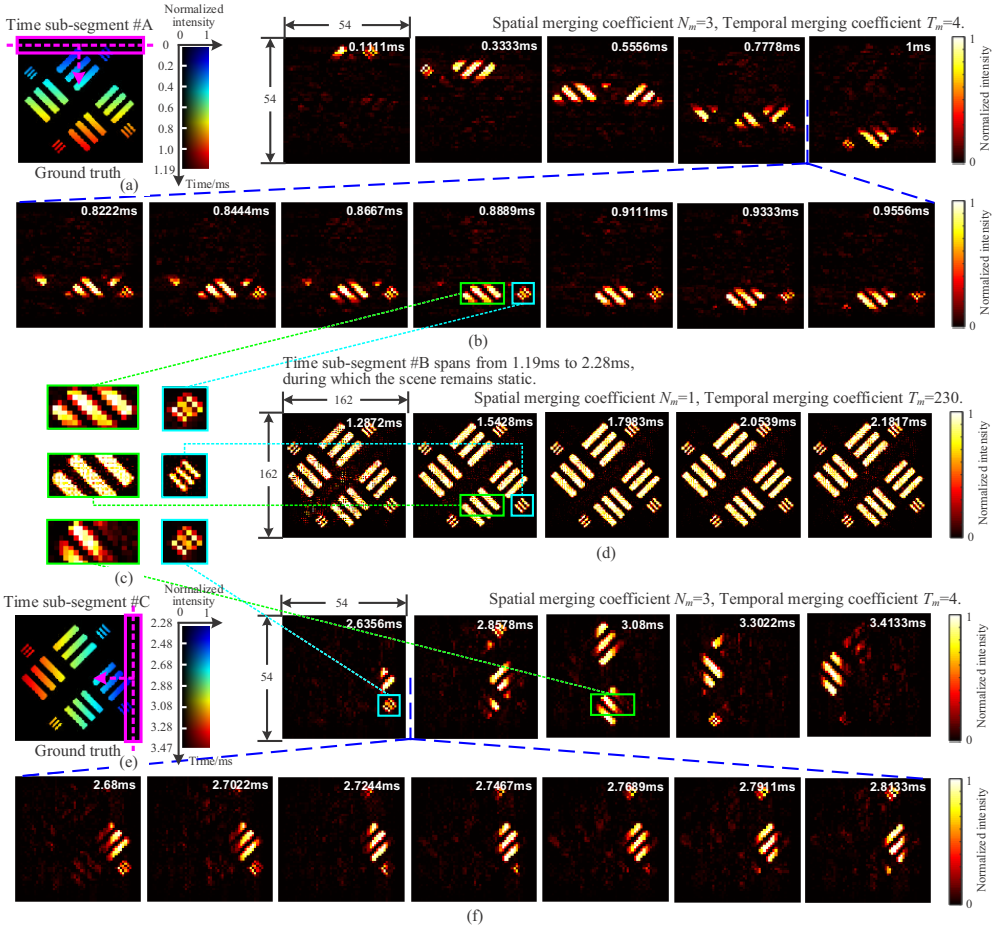

Fig. 8. Simulation results for the dynamic test target. The scene is divided into three time sub-segments. A narrow rectangular window shown in (a) and (e) slides vertically downward and horizontally to the left, respectively, generating the time sub-segments #A and #C. The time sub-segments #B between them is a static process with the fully open window. In the case of $N_m = 3$ and $T_m = 4$, the reconstructed video frames of sub-segments #A and #C are shown in (b) and (f). While in the case of $N_m = 1$ and $T_m = 230$, the reconstruction results of sub-segment #B are given in (d). Close-up views of the target bars in these three sub-segments are collected in (c). See Visualization 2 for the final integrated video.

overlapping relationship of adjacent sub-segments shown in Fig. 5, sub-segment #A and #C are both 2136 frames with a time span of 1.19 ms, and sub-segment #B sandwiched between them has a total of 1972 frames and lasts for 1.09 ms. By using a narrow open window on the static test target to slide vertically downward and horizontally to the left, respectively, we get the ground truth scenes of sub-segments #A and #C, as shown in Fig. 8(a) and Fig. 8(e).

In reconstruction, the merging coefficients $N_m = 3$ and $T_m = 4$ are set for these two sub-segments to ensure a clear capture of the sliding processes, and the reconstructed video frames are shown in Fig. 8(b) and Fig. 8(f). Sub-segment #B is a static process with the fully open window, which is suitable for taking the merging coefficients as $N_m = 1$ and $T_m = 230$ to explore the target details. The reconstructed video frames are shown in Fig. 8(d). See Visualization 2 for their final integrated video. In terms of resolving for the target bars, the reconstructed frames

Optics EXPRESS

for sub-segment #B will be significantly better than the other two due to no spatial resolution degradation, which can also be confirmed by the close-up views collected in Fig. 8(c). Note that although we cannot accurately predict the target characteristics before reconstruction, both the scene prior and the tentative reconstruction by multiple sets of coefficients can still enable an optimized set of coefficients and satisfactory reconstruction effects.

We observed in Fig. 7 and Fig. 8 that the system has slight artifacts when imaging targets with spatial displacement, which presents that the target in other time periods will appear slightly in the current frame. After analysis, it is concluded that the artifacts are caused by low sampling rate and the resulting mediocre reconstruction quality. To obtain a considerable target image with such a small amount of sampled data, we performed a sparse transformation along the temporal dimension to squeeze its redundancy during reconstruction, and then the limited bandwidth effect is also responsible for the artifacts. In reconstruction, a better deep denoiser or neural network is expected to suppress these artifacts and further improve the video quality.

5. Experiments

5.1. Experimental setup

Following the dual-arm system depicted in Fig. 2(a), we built a proof-of-concept setup to verify the effectiveness of the proposed framework. The light beam emitted from the dynamic scene, whose generation will be described separately below, is incident to the beam splitter (BS013, 50:50 Non-Polarizing, Thorlabs, Inc.) through the objective lens (MVL50M23, 50 mm EFL, f/2.8, Thorlabs, Inc.), and then the transmitted and reflected beams are projected to the focal planes of two orthogonally installed TDI cameras (VT-4K5C-H100, Vieworks Co., Ltd.) through the relay lenses (MVL35M23, 35 mm EFL, f/2.0, Thorlabs, Inc.), respectively. Each camera is equipped with a line start trigger interface to facilitate synchronization. We developed a FPGA (XC4VSX55, Advanced Micro Devices, Inc.)-based synchronization trigger to synchronize coded exposure and TDI line transfer. The computer configured with an Intel Core i7-10750 H CPU and 16 G RAM is responsible for real-time data acquisition and reconstruction algorithm execution.

In particular, the controllable intra-line charge transfer described in the theoretical model is not supported in the off-the-shelf TDI cameras, so we use the inter-line charge transfer to simulate it equivalently, which is convincing because of the same physical mechanism. With the TDI intra-line steps S set to 3, the TDI camera with a pixel size of 5μm and an integral stage set to 192 in the experimental setup will be equivalently merged to a model with a pixel size of 15μm and an integral stage of 64. The maximum line rate of 100kHz will also be reduced to 33.33kHz. Besides, the raw sampled data should implement a 3×3 merging in the digital domain to get the equivalent sampling results.

After the above equivalence, the experimental setup has a spatial resolution of 192×192 in case of no downscaling, and a frame rate of up to 100,000 frames-per-second. Note that the latter does not represent the upper limit of the system framework described in this paper. Since the charge transfer rate in TDI sensors is in nanoseconds, the line rate is mainly limited by the electronic readout and quantization operation, as well as data transmission bandwidth, which can reach hundreds of kilohertz at present. On this premise, benefiting from the intra-line TDI operation, the proposed framework will be further improved by several times in terms of temporal resolution, reaching millions of frames-per-second capable of continuous streaming.

5.2. Results

In the above numerical simulations, the effectiveness of the proposed framework, and the capability of flexible voxels post-interpretation, as well as the reconstruction strategy derived from it, have been demonstrated. Therefore, we focus on its ability to capture high-dynamic scenes in the experimental verification. In order to demonstrate its adaptability to scenes with

Optics EXPRESS

different characteristics, we designed three fast changing scenes in total, covering changes in intensity, morphology, and their mixture. In addition, with such a high frame rate requirement, it becomes costly to capture the ground truth directly with a high-speed camera. To circumvent this issue, we use the controllable dynamic scenes instead, so as to verify the experimental results by comparing them with the theoretical truth value. These dynamic scenes are generated with the help of manually manipulated LEDs and an optical chopper, respectively. Note that in the implementation of the former, the LEDs will take on the functions of both dynamic scene generation and temporal coding, so a logical "and" operation is required between the coded exposure signal and the modulated signal used for high dynamic analog, before finally driving the LEDs. Additionally, in both simulations and experiments, we have tried to generate multiple sets of coding sequences with different seeds, and the results show that the reconstruction quality is quite robust to the coding sequence. So there is no specific coding sequence requirement within the pseudo-random coding framework.

In Scene #1, we use four independently driven LEDs to illuminate targets with different shapes, where x-shape, triangle, and circle are implemented with sinusoidal driving signals of 5kHz, 2.5kHz, and 1kHz, respectively, by means of high-frequency pulse width modulation. While the equal-sign-shape keeps a constant intensity by fixing the duty cycle. The two TDI cameras' sampling results are shown in Fig. 9(a). From the continuous streaming, we cut out 263 lines of readout data, spanning from 52 ms to 59.89 ms. Then, with the merging coefficients set to $N_m = 3$ and $T_m = 1$, a video data cube with a size of $64 \times 64 \times 598$ is reconstructed, which spans from 52 ms to 57.98 ms at a frame interval of 10μs. See Visualization 3 for the full video. The representative frames are shown in Fig. 9(b), in which the close-up views of the three shapes visualize the fluctuation of their intensity over time, referring to the frame number and time marked within the image. To facilitate comparison with the theoretical truth value, the average intensity of the local areas covering these three shapes is calculated separately, and then plotted in Fig. 9(c). As expected, they all appear as sinusoids with the same frequency as the theoretical one, except that the circle's intensity curve is slightly distorted compared to the other two. By analyzing this distortion, we found that its LED entered the near-saturated operating area, resulting in a nonlinearity between the luminous intensity and the driving current, which was also effectively identified by our proposed system.

With the intensity change similar to that in Scene #1 preserved, Scene #2 also contains an embodiment of changes in spatial information. Two seven-segment displays (HDSP-7513, Broadcom, Inc.) are used as target generating devices, mounted in left and right, and display Arabic numerals alternately in the manner described in Fig. 10(a). The intensity of each digit fluctuates in triangular waves with a display period of 1.28 ms. With the same merging coefficients setting as above, and also the sampling results spanning from 52 ms to 59.89 ms are cut out for reconstruction, we obtain a reconstructed video of Scene #2, including a total of 598 consecutive frames, as seen in Visualization 4. The representative frames are shown in Fig. 10(b). In the frame with peak intensity, we can clearly capture and identify the Arabic numerals, as shown in the close-up views. To present the intensity change in the scene, we calculated the average intensity of the small rectangular boxes marked in Fig. 10(b), and their changes over time are plotted in Fig. 10(c). The two areas corresponding to the left side of the figure are fully activated during the digital display period, so their intensity presents a continuous triangular wave. Since they belong to the left and right digits respectively, there is a 180-degree phase difference between them, which is consistent with the theoretical value shown in Fig. 10(a). As for the area corresponding to the right side of Fig. 10(c), when 1-3-5-7 alternates, it is only activated at 3-5 with the rest blanked, so its average intensity shows an intermittent fluctuation as in the figure.

Finally, to demonstrate its ability to capture high-speed motion, we designed Scene #3. The setup used to generate this dynamic scene is shown in Fig. 11(a), which uses an optical chopper to sweep across the static USAF target at high speed in the middle image plane, forming the

Optics EXPRESS

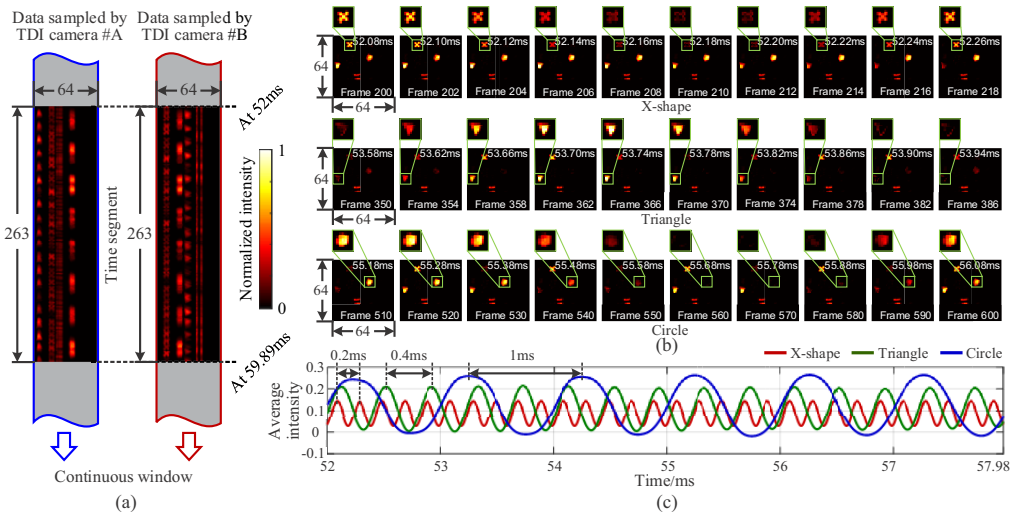

Fig. 9. Experimental results for Scene #1. (a) The segments cut out from the sampling results of two TDI cameras, with a resolution of 263×64, spanning from 52 ms to 59.89 ms. (b) The representative frames, in which the close-up views of the three shapes visualize the fluctuation of their intensity over time. (c) The average intensity of the local areas covering these three shapes varies with time.

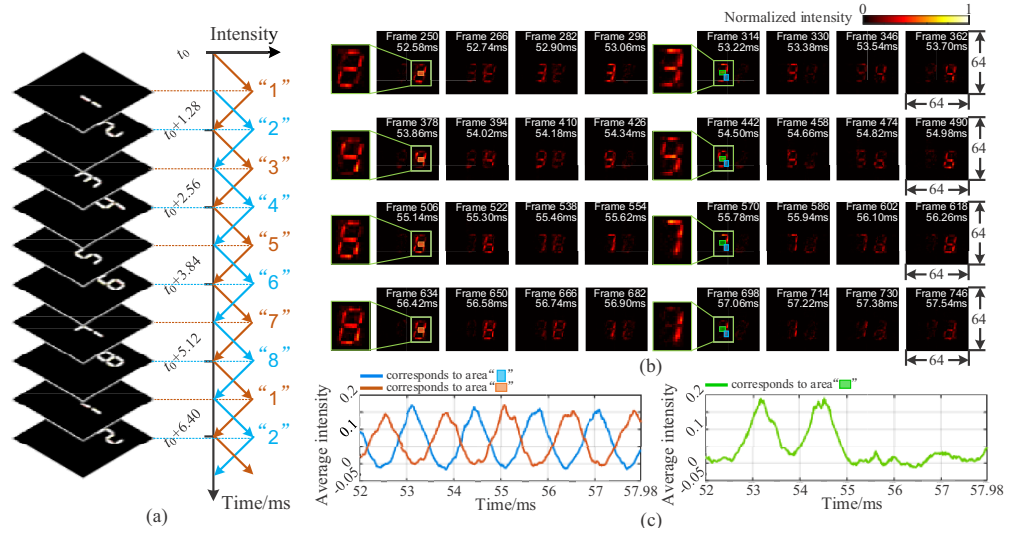

Fig. 10. Experimental results for Scene #2. Theoretically, the Arabic numerals displayed alternately by two seven-segment displays and their intensity changes are depicted in (a). The representative reconstructed frames are shown in (b), in which the close-up views present identifiable Arabic numerals. The average intensity fluctuates over time is plotted in (c), corresponding to the small rectangular boxes marked in (b), respectively.

Optics EXPRESS

trajectory shown in the lower left corner of the figure. From the continuous sampling streaming, we cut out a segment that contains a complete sweeping process, spanning from 30 ms to 70 ms. With the merging coefficients set to $N_m = 3$ and $T_m = 1$, the size of the reconstructed video corresponding to this process will reach $64 \times 64 \times 4000$. To reduce the computational complexity and memory pressure in the reconstruction, it is divided into 7 sub-segments overlapped in the manner depicted on the right side of Fig. 5, which are then reconstructed separately, and finally combined together. The representative video frames are shown in Fig. 11(b), which clearly present a complete sweeping process. See Visualization 5 for the full video.

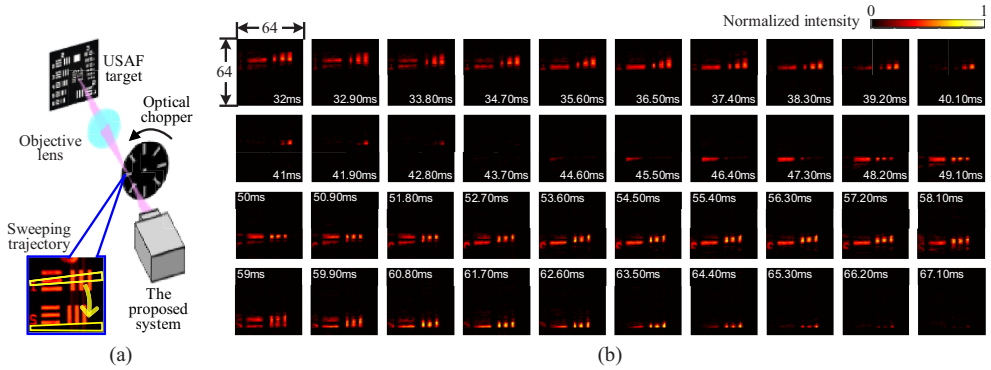

Fig. 11. Experimental results for Scene #3. (a) The setup used to generate dynamic Scene #3, which consists of a USAF target, an objective lens, and an optical chopper. (b) The time-stamped representative reconstructed frames, which record a complete sweeping process.

6. Conclusion and discussion

In this paper, we propose a compressed spatio-temporal imaging framework implemented by combining TDI sensors and coded exposure. Since no additional optical coding element and subsequent calibration are required, it has a more compact and robust hardware structure, compared to other existing high-speed imaging modalities. By exploiting the intra-line charge transfer mechanism, we achieve a super-resolution in both temporal and spatial domains, thus multiplying the frame rate to millions of frames-per-second. In addition, based on the forward model with post-tunable coefficients and the reconstruction strategies derived from it, we realize a flexible voxels post-interpretation, which can flexibly specify the spatial and temporal resolution for the target concerned. Subsequently, the effectiveness of the proposed system, as well as its capability of flexible voxels post-interpretation, was demonstrated in numerical simulations. Furthermore, we built a proof-of-concept setup with a spatial resolution of 192×192, and a frame rate of up to 100,000 frames-per-second. Note that the latter does not represent the upper limit of the imaging framework described in this paper. In experiments, its adaptability to transient scenes with various characteristics was verified.

In the future work, we will focus on the improvement in pixel count and frame rate. To fully utilize the large pixel count of TDI cameras, based on the proof-of-concept setup, its resolution of 4640×256 can be remapped to an incident scene of 1536×768 by the image-transmitting optical fiber, thus producing images with formats that conform with conventional photography. And it can be further multiplied by the intra-line TDI operation. On the other hand, since the charge transfer rate in TDI sensors is in nanoseconds, the line rate is mainly limited by the electronic readout and quantization operation, as well as data transmission bandwidth, which is currently clamped at hundreds of kilohertz. These advances in TDI cameras can be anticipated for a higher

Optics EXPRESS

frame rate in the near future. Additionally, the deep learning framework is also preferred to improve the quality of reconstruction and provide real-time acquisition.

For applications, based on a simplified and robust framework, it provides a cost-effective alternative to high-speed imaging. Especially, with the prominent advantages of prolonged time window and flexible voxels post-interpretation, the proposed system will be suitable for imaging random, non-repetitive, or long-term events.

Funding. National Natural Science Foundation of China (62005266); Youth Innovation Promotion Association of the Chinese Academy of Sciences (2022219); Open Project Funds for the Key Laboratory of Space Photoelectric Detection and Perception (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology (NJ2022025-2); Fundamental Research Funds for the Central Universities (NJ2022025).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. K. Goda, A. Ayazi, D. R. Gossett, J. Sadasivam, C. K. Lonappan, E. Sollier, A. M. Fard, S. C. Hur, J. Adam, C. Murray, C. Wang, N. Brackbill, D. D. Carlo, and B. Jalali, "High-throughput single-microparticle imaging flow analyzer," Proc. Natl. Acad. Sci. U.S.A. **109**(29), 11630–11635 (2012).
2. S. S. Howard, A. Straub, N. G. Horton, D. Kobat, and C. Xu, "Frequency-multiplexed in vivo multiphoton phosphorescence lifetime microscopy," Nat. Photonics **7**(1), 33–37 (2013).
3. X. Liu, A. Skripka, Y. Lai, C. Jiang, J. Liu, F. Vetrone, and J. Liang, "Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging," Nat. Commun. **12**(1), 6401 (2021).
4. I. Radu, K. Vahaplar, C. Stamm, T. Kachel, N. Pontius, H. A. Dürr, T. A. Ostler, J. Barker, R. F. L. Evans, R. W. Chantrell, A. Tsukamoto, A. Itoh, A. Kirilyuk, Th. Rasing, and A. V. Kimel, "Transient ferromagnetic-like state mediating ultrafast reversal of antiferromagnetically coupled spins," Nature **472**(7342), 205–208 (2011).
5. W. Yang and R. Yuste, "In vivo imaging of neural activity," Nat. Methods **14**(4), 349–359 (2017).
6. H. Mikami, L. Gao, and K. Goda, "Ultrafast optical imaging technology: principles and applications of emerging methods," Nanophotonics **5**(4), 497–509 (2016).
7. T. G. Etoh, D. V. T. Son, T. Yamada, and E. Charbon, "Toward one giga frames per second–Evolution of in situ storage image sensors," Sensors **13**(4), 4640–4658 (2013).
8. M. Suzuki, M. Suzuki, R. Kuroda, and S. Sugawa, "A preliminary chip evaluation toward over 50Mfps burst global shutter stacked CMOS image sensor," in *Proc. IS&T Int'l. Symp. on Electronic Imaging: Image Sensors and Imaging Systems* 30, 398 (2018).
9. T. Arai, J. Yonai, T. Hayashida, H. Ohtake, H. van Kuijk, and T. G. Etoh, "A 252-v/lux·s, 16.7-million-frames-per-second 312-kpixel back-side-illuminated ultrahigh-speed charge-coupled device," IEEE Trans. Electron Devices **60**(10), 3450–3458 (2013).
10. Y. Tochigi, K. Hanzawa, Y. Kato, R. Kuroda, H. Mutoh, R. Hirose, H. Tominaga, K. Takubo, Y. Kondo, and S. Sugawa, "A global-shutter CMOS image sensor with readout speed of 1-Tpixel/s burst and 780-Mpixel/s continuous," IEEE J. Solid-State Circuits **48**(1), 329–338 (2013).
11. C. T. Chin, C. Lancée, J. Borsboom, F. Mastik, M. E. Frijlink, and N. de Jong, "Brandaris 128: A digital 25 million frames per second camera with 128 highly sensitive frames," Rev. Sci. Instrum. **74**(12), 5026–5034 (2003).
12. V. Tiwari, M.A. Sutton, and S.R. McNeill, "Assessment of High Speed Imaging Systems for 2D and 3D Deformation Measurements: Methodology Development and Validation," Exp. Mech. **47**(4), 561–579 (2007).
13. B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, 294–301 (2004).
14. K. Nakagawa, A. Iwasaki, Y. Oishi, R. Horisaki, A. Tsukamoto, A. Nakamura, K. Hirosawa, H. Liao, T. Ushida, K. Goda, F. Kannari, and I. Sakuma, "Sequentially timed all-optical mapping photography (STAMP)," Nature Photonics **8**(9), 695–700 (2014).
15. K. Goda, K. K. Tsia, and B. Jalali, "Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena," Nature **458**(7242), 1145–1149 (2009).
16. Y. Tsuchiya, "Advances in streak camera instrumentation for the study of biological and physical processes," IEEE Journal of Quantum Electronics **20**(12), 1516–1528 (1984).
17. D. L. Donoho, "Compressed Sensing," IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006).
18. E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory **52**(2), 489–509 (2006).
19. E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling," IEEE Signal Process. Mag. **25**(2), 21–30 (2008).
20. E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Problems **23**(3), 969–985 (2007).

Optics EXPRESS

21. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," Opt. Express **21**(9), 10526–10545 (2013).
22. R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. K. Katsaggelos, "High spatio-temporal resolution video with compressed sensing," Opt. Express **23**(12), 15992–16007 (2015).
23. Q. Zhou, J. Ke, and E. Y. Lam, "Near-infrared temporal compressive imaging for video," Opt. Lett. **44**(7), 1702–1705 (2019).
24. D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: programmable pixel compressive camera for high speed imaging," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 329–336 (2011).
25. M. Qiao, X. Liu, and X. Yuan, "Snapshot spatial–temporal compressive imaging," Opt. Lett. **45**(7), 1659–1662 (2020).
26. J. Ke, L. Zhang, Q. Zhou, and E. Y. Lam, "Broad dual-band temporal compressive imaging with optical calibration," Opt. Express **29**(4), 5710–5729 (2021).
27. X. Liu, J. Liu, C. Jiang, F. Vetrone, and J. Liang, "Single-shot compressed optical-streaking ultra-high-speed photography," Opt. Lett. **44**(6), 1387–1390 (2019).
28. L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**(7529), 74–77 (2014).
29. F. Mochizuki, K. Kagawa, S. Okihara, M. Seo, B. Zhang, T. Takasawa, K. Yasutomi, and S. Kawahito, "Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor," Opt. Express **24**(4), 4155–4176 (2016).
30. Y. Luo, J. Jiang, M. Cai, and S. Mirabbasi, "CMOS computational camera with a two-tap coded exposure image sensor for single-shot spatial-temporal compressive sensing," Opt. Express **27**(22), 31475–31489 (2019).
31. G. Weinberg and O. Katz, "100,000 frames-per-second compressive imaging with a conventional rolling-shutter camera by random point-spread-function engineering," Opt. Express **28**(21), 30616–30625 (2020).
32. F. Guzmán, P. Meza, and E. Vera, "Compressive temporal imaging using a rolling shutter camera array," Opt. Express **29**(9), 12787–12800 (2021).
33. E. Vera, F. Guzmán, and N. Díaz, "Shuffled rolling shutter for snapshot temporal imaging," Opt. Express **30**(2), 887–901 (2022).
34. J. park and L. Gao, "Continuously streaming compressed high-speed photography using time delay integration," Optica **8**(12), 1620–1623 (2021).
35. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "Flexible Voxels for Motion-Aware Videography," in *Proceeding of European Conference on Computer Vision* 6311, 100–114 (2010).
36. G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, "Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging," Nat. Methods **7**(3), 209–211 (2010).
37. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twIst: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. on Image Process. **16**(12), 2992–3004 (2007).