Single-shot compressed ultrafast photography based on U-net network

ANKE ZHANG,1 JIAMIN WU,1 JINLI SUO,1 LU FANG,1 HUI QIAO,1,2 DAVID DAY-UEI LI,2 SHIAN ZHANG,3 JINTAO FAN,1,3 DALONG QI,3 QIONGHAI DAI,1,4 AND CHENGQUAN PEI1,5

1Department of Automation, Tsinghua University, Beijing 100084, China
2Faculty of Science, University of Strathclyde, Glasgow G4 0RE, UK
3State Key Laboratory of Precision Spectroscopy, School of Physics and Materials Science, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China
4daiqionghai@tsinghua.edu.cn
5zihaopei@mail.tsinghua.edu.cn

Abstract: The compressive ultrafast photography (CUP) has achieved real-time femtosecond imaging based on the compressive-sensing methods. However, the reconstruction performance usually suffers from artifacts brought by strong noise, aberration, and distortion, which prevents its applications. We propose a deep compressive ultrafast photography (DeepCUP) method. Various numerical simulations have been demonstrated on both the MNIST and UCF-101 datasets and compared with other state-of-the-art algorithms. The result shows that our DeepCUP has a superior performance in both PSNR and SSIM compared to previous compressed-sensing methods. We also illustrate the outstanding performance of the proposed method under system errors and noise in comparison to other methods.

© 2020 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

The capture of transient scenes at high imaging speed is essential for various applications [1–5] and it can help us to extend our understanding of transient processes. With rapid developments in CCD, CMOS sensor, and single-photon avalanche diode (SPAD) technologies, the imaging speed has been increased from several frames per second (fps), achieved by an intermittent camera [6], to one billion frames per second [7]. Currently, the predominant approaches for capturing transient events are all-optical mapping photography [8–10] (STAMP), serial time-encoded amplified imaging [11–12] (STEAM), and compressed ultrafast spectral-temporal (CUST) photography [13]. All of them have been used in physical chemistry [14–17], materials science [18], and nonlinear optics [19]. Moreover, imaging scattered dynamics within picoseconds or even femtoseconds are meaningful in biomedicine, such as blood flow velocity [20] and tissue elasticity [21]. Research in light scattering imaging has also been increasingly featured in recent progress in biomedicine [22]. Besides, the analysis of temporal fluctuation in the scattered light signal reveals many optical properties of biological tissues [22,23].

This characteristic has enabled a diverse range of applications, such as assessments of food and pharmaceutical products [24] and studies of protein aggregation diseases [25]. Streak cameras (SCs) are ultrafast imaging tools that can convert the time variations of an ultrafast signal to a spatial profile and achieve picoseconds or even femtosecond measurements with high spatial resolution. However, the images on CCD should be narrow enough to read out the time information due to the shearing operation. Thus, the camera can only capture one-dimensional images. Therefore, a narrow entrance slit (50um) is inserted in front of the camera lens and it can limit the imaging field of view (FOV) to a line. To achieve two-dimensional imaging, this system should equip additional optical scanning mirrors. Although this method is capable of capturing
the transient event, the event itself must be repetitive following the same spatiotemporal pattern while the entrance slit of a streak camera steps across the entire FOV.

In cases where the physical phenomena are not repetitive, such as shock wave, nuclear explosion, or synchrotron radiation, this 2D streak imaging method is inapplicable. To overcome this limitation, a computational photography method for streak cameras was proposed, which can capture 2D dynamic images with a temporal resolution of picoseconds. In this method, the spatial domain was encoded by a pseudo-random binary pattern, followed by a streak camera with a fully opened entrance slit. In this processing, the three-dimensional (3D: x, y, t) scene is then measured by a 2D detector array with a single snapshot, and the reconstruction from 2D to 3D can be translated into a convex optimization problem. It is also a snapshot compressed imaging (SCI) system. Gao, Liang and Shian Zhang have developed this system with the reconstruction method called TwIST [26], and they have achieved exciting results [27–31]. However, TwIST is sensitive to input parameters, and the low peak signal to noise ratio (PSNR) and the structural similarity index (SSIM) are poor. Some algorithms used in compressed sensing can be adopted to reconstruct images for this compressive ultrafast photography (CUP) system. In [32], a new reconstruction framework was proposed, adopting the rank minimization as an intermediate step during reconstruction. Specifically, by integrating the compressive sampling model in SCI with the weighted nuclear norm minimization (WNNM) for video patch groups, a joint model for SCI reconstruction was formulated [32,33]. To solve this problem, the alternating direction method of multipliers (ADMM) [34] method was employed to develop an iterative optimization algorithm for SCI reconstruction. DeSCI and GAP-TV are two typical methods [35,36].

In this paper, a deep learning method was developed to reconstruct CUP images in a single shot. We take Mnist, UCF-101 [36] and Runner [37] datasets to simulate this CUP system with a perfect mask which do not contains any noise aberration, and distortion. The results show that the deep learning method has advantages over DeSCI, GAP-TV, and TwIST in PSNR and SSIM. However, the codes used in real experiments is extracted from an image captured by this CUP system for code mask with static mode. It contains strong noise, aberration, and distortion in real experiments, which degrade the reconstruction performance. The robust performance was also evaluated by taking Mnist data and a mask captured in the real experiment to simulated this CUP system. The results show that the PSNR and SSIM calculated by the proposed method outperforms the other methods. Besides, the computing efficiency of this deep learning method is also better than TwIST and GAP. It is possible to implement a real-time ultrafast imaging system. We set up an experiment to record the transient process of a femtosecond laser passing through the water added a little milk, and the result showed that this system can achieve the temporal resolution of 4ps.

2. Experiment setup

To validate our method, we imaged propagation of femtosecond laser pulses in real-time. A system as shown in Fig. 1 was established. The system consists of two relay lenses (Relay lens1, Nikon 35F2D; Relay lens2, Nikon 50-1.4D) and a mask which is made of 512×512 random codes and the size of every code is 75um×75um. The dynamic scattering scene is imaged by the camera lenses, to an intermediate plane. Then the light is encoded by a mask firstly and finally captured by a streak camera (Hamamatsu C7700) with 1ns scanning time and 5mm slit width. Inside the streak camera, a sweeping voltage with an ultrafast slope is applied along the y-axis, deflecting the encoded image frames towards different y locations according to the times of arrival.

The final temporally dispersed image is captured by a streak camera (1016×1344) with a single exposure. In the experiment, the dynamic scattering scene, namely, a femtosecond laser (800nm) passing through water tinted with milk, was imaged, and the field of view is about 600ps. Mathematically, the compression process is the same as CUP system [27]. The process
is equivalent to successively applying a spatial encoding operator, $C$, and a temporal shearing operator, $S$, to the intensity distribution from the input dynamic scene, $I(x, y, t)$, the processing of streak camera can be derived as,

$$I_c(x', y', t) = SCI(x, y, t)$$  \hspace{1cm} (1)

where, $I(x, y, t)$ represents the original dynamic scene. Then, $I_c(x', y', t)$ is the encoded and sheared scene by the streak camera. Then, CCD compress and accumulate the sheared scene with a resolution $R \times (N + vt)$; $v$ is the shearing speed of the streak camera. The compressed image $E(m, n)$ is:

$$E(m, n) = TI_c(x, y, t)$$  \hspace{1cm} (2)

where the operator $T$ represents compressing process of streak camera CCD. When encoding the image, the mask samples the scene sparsely. Assuming the size of mask is the same as pixel size, the encoded video can be expressed as:

$$I_c(x, y, t) = \sum_{i,j,k} I_{i,j,k} C_i j - v t rect\left(\frac{x}{d} - \left( i + \frac{1}{2} \right), \frac{x}{d} - \left( j + \frac{1}{2} \right), \frac{t}{\Delta t} - \left( k + \frac{1}{2} \right)\right)$$  \hspace{1cm} (3)

where $d$ is the pixel size, and $I_{i,j,k}$ is the discretized pixel imaged on sensor, $C_i$ is the binary mask value in $x$-$y$ direction, $\Delta t = \frac{d}{v}$ is the scanning time interval on each pixel. Then, the streak camera CCD compressed the burst video series. The process can be viewed as a sum in $t$ direction:

$$E(m, n) = \sum_{k=0}^{n-1} I_c(m, n, k)$$  \hspace{1cm} (4)

To estimate the original scene from the CCD measurement, we need to solve the inverse problem of Eq. (4). This process can be formulated in a more general form as:

$$\arg \min \left\{ \frac{1}{2} || E - TSCI(x, y, t) ||^2 + \beta \Phi(I) \right\}$$  \hspace{1cm} (5)

$$\Phi(I) = \sum_{k=0}^{N_t-1} \sum_{i=1}^{N_x N_y} \sqrt{(\Delta^h I_k)^2 + (\Delta^v I_k)^2} + \sum_{k=0}^{N_t} \sum_{i=1}^{N_x N_y} \sqrt{(\Delta^h I_m)^2 + (\Delta^v I_m)^2} + \sum_{k=0}^{N_t} \sum_{i=1}^{N_x N_y} \sqrt{(\Delta^h I_n)^2 + (\Delta^v I_n)^2}$$  \hspace{1cm} (6)

Where, $\Phi(I)$ is the regularization function, $\beta$ is the regularization parameter and $\Delta^h$, $\Delta^v$ are horizontal and vertical first-order local difference operators on a 2D lattice. For example, to
reconstruct a dynamic scene with dimensions $N_x \times N_y \times N_t$ ($N_x$, $N_y$, are respectively the numbers of voxels along $x$, $y$), where the coded mask itself has dimensions $N_x \times N_y$, the actual matrix $E$ used in Eq. (4) will have dimensions $N_x \times (N_y + N_t - 1)$, with zeros padded to the ends. $N_t$ is reconstructed frames finally.

3. Neural network design

3.1. Encoder

We designed a novel Deep Compressive Ultrafast Photography (DeepCUP) to accomplish single-shot-single-mask decompression. By assuming all the training and testing data using the same mask in compression encoding, the neural network can be trained to extract the reverse process from a compressed single image to an uncompressed video sequence. Thus, a $l_0$ optimization problem of compressed sensing is converted to a recognition and extraction problem. The compressed video is can be interpreted as a set of sparsely sampled sequences that are related in position and time, and the compressed image is a feature space that can be remapped to a decoupled space through series of nonlinear re-projections. To decode the time series of spatial information, three essential components are required: a feature extractor that presents the meaning of distribution, a projector that can separate all spatial and time-sequential information, and multiple extractors that discern and collect the information of specific frames. As shown in Fig. 2, The network starts from a 3-layer Convolutional encoder for feature extraction cascaded by 15 Res blocks of high-level feature mapping. The result of the feature mapping is decoded by 8 convolutional decoders separately.

![Fig. 2. The architecture of neural network proposed in this paper.](image)

The 3-layer convolutional encoder takes in a sparsely encoded image $I$ in space $R^{M \times N}$, and the network extracts and refines the feature space of the sparsely encoded image with a channel number 64, 128, 256 consecutively. After each convolutional operation, normalization and nonlinear operations are applied. The output of Layer $i$ and Channel $k$ is formulated by,

$$X^{(i+1)}_k = \text{Relu} \left( \frac{CX^{(i)}_k}{||CX^{(i)}_k||} \right), \quad (7)$$

where

$$CX^{(i)}_k = b_k + \sum_{j=1}^{C_i} w^{(i)}_{j,k} \ast X^{(i-1)}_j \quad (8)$$

And $C_i$ is the input channel number of the current convolutional layer, $w^{(i)}_{j,k}$ is weight of neural network, and there is a $3 \times 3$ kernel corresponding to the input channel $C_j$ and output channel
Ck in Layer i. The input image can be viewed as a single channel input. The 3-layer encoder extracts and converts the input image into a high-level feature space for the cascaded remapping process.

3.2. ResNet re-projection

After the encoding process of DeepCUP, the feature space is consecutively remapped by 15 ResNet blocks formulated by,

\[ x_{k+1} = x_k + \text{Relu}(T(x_k)) \]  

where \( x_k \) is the output of the \( k \)th layer Res-block and \( T \) is the transformation applied in the Res-block. We used 15 res-blocks in series for the whole re-projection process, and the transformation function \( T \) is:

\[ T(x_k^{(i)}) = \frac{b_k + \sum_{j=1}^{C_{in}} w_{jk}^{(i)} \ast x_{j}^{(i-1)}}{\left| b_k + \sum_{j=1}^{C_{in}} w_{jk}^{(i)} \ast x_{j}^{(i-1)} \right|} \]  

The specific parameters of the transformation are shown in Table 1.

| Outputs          | Layer structure            |
|------------------|-----------------------------|
| \( 256 \times 64 \times 64 \) | Conv2d 3 × 3 Stride 1 Batch Normalization Relu |
|                  | Conv2d 3 × 3 Stride 1      |
|                  | Batch Normalization        |

The remapping process is cascaded by a convolution layer, a normalization layer, and a Relu activation layer followed by another convolution layer and a normalization layer. The function of the Resnet block is to convert and separate these high-level features by a series of non-linear transformations. The decoder has the ability of feature extraction and recombination, but the encoded information might not contain decoupled information that can be directly extracted. Thus, we need a series of re-projections such that the high-level feature space is remapped to a state that can be decoded by multiple parallel designed decoders.

3.3. Decoder

The last layer of the res-block is connected to 8 decoders in parallel, and each decoder extracts and recombines the remapped information in one particular frame. Each decoder is consisting of 3 cascaded layers of transpose convolution, which use Eq. (7) with the input being zero-padded with stride 2. Because the 8 decoder is in a parallel design, the final output should be in \( R^{8 \times M \times N} \) when processing the input in \( R^{(M+vt) \times (N)} \).

3.4. Training

In the forward path, the video of dynamic scene is converted into an compressed image, and we want to find function \( \phi \), \( R^{x \times y \times vt} \) to \( R^{x \times y \times t} \) such that \( I = \phi(E) \), where, \( E \) is compressed image, \( I \) is constructed image. To train the network that minimizes the error of the reconstructed burst video
series $I'$, we constrain and optimize the model with the following objective function,

$$L = \mu_a \| I(x, y, t) - I_d(x, y, t) \| + \mu_b \| TSC \tilde{I}(x, y, t) - TSCI_d(x, y, t) \| + \mu_c \Phi(\tilde{I})$$

(11)

where, $I$ is the output, $I_d$ is the ground truth, TSC is the forward process of compressing, shearing, and encoding and $\mu_a, \mu_b, \mu_c$ are hyperparameters. $\Phi$ is the total variation formulated as,

$$\Phi(\tilde{I}) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{(\Delta_x I'_m)^2 + (\Delta_y I'_n)^2} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{(\Delta_x I'_m)^2 + (\Delta_y I'_n)^2} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \sqrt{(\Delta_x I'_m)^2 + (\Delta_y I'_n)^2}$$

(12)

Here we assume that the discretized form of $I'$ has dimensions $N_x \times N_y \times N_t$, and $m, n, k$ are indices. $I'_m, I'_n, I'_k$ denote the 2D lattices along the dimensions $m, n, k$, respectively. The training process enforces the decoded video close to the original video, but this constraint does not provide strong restrictions on the total intensity and forward model. Thus, we added two extra constraints on the forward model: one by the encoded streak camera image, and the other by the black/white CCD. We also added the total variation constraint to enforce the smoothness of the final output. We set the $\mu_a$ to be 1, $\mu_b, \mu_c$ to be 0.1 when training both Flying MNIST Dataset and UCF-101 Dataset.

4. Experiment and results

Various methods can be applied for SCI reconstruction, but most of them are based on spatial-varying mask [36] reconstruction instead of static mask reconstruction. In this paper, we compared with two spatial-varying mask reconstruction methods, DeSCI and GAP-TV, and one static mask reconstruction method, TwIST.

4.1. PSNR and SSIM analysis based on simulation

We trained the model with 2 datasets: UCF-101 and Flying Mnist datasets. The UCF-101 dataset consists of 13220 videos from 101 human action categories. We generate 65000 clips with 8 frames per clip across all sorts of human actions. The UCF-101 dataset is used for training a model for daily scenes, which is smoother and texture-rich. The Flying Mnist dataset is a self-generated dataset with several Mnist numbers flying in random directions. Each clip also has a time length of 8 frames. The training set contains 65,000 clips, and the validation set contains 1000 clips. Such a scene resembles the streak camera detection, which is usually a simple scene but in a high dynamic range. For the real dataset, we take pictures based on the CUP system, which images femtosecond events in burst mode. Because the incoming light will encounter optical-electrical-optical conversion, massive noise is brought to the final image. The UCF-101 dataset was performed as a series of linear operations based on the CUP system and reconstructed it with TwIST, GAP-TV, DeSCI, and the DeepCUP proposed in this paper. In our simulations, the UCF-101 dataset was encoded by a perfect mask and the reconstruction result is shown in Fig. 3.

To prove the generalization ability of the model, we further trained the model on UCF-101 dataset to solve general compress-sensing problem. The random 65,000 video clips were generated to train the model. To benchmark the training result, we choose the widely-used drop dataset [38] and Runner dataset [37], simulated with a series of operations based on the CUP system with perfect encoding, and it was reconstructed with TwIST, DeSCI, GAP-TV and DeepCUP, respectively. The results are shown in Fig. 4 and Fig. 5. Moreover, PSNR and SSIM were also evaluated and the results are summarized in Table 2.

We can conclude from Fig. 3, Fig. 4, Fig. 5 and Table 2 that for the simulated data, DeepCUP outperforms other state-of-the-art methods in PSNR and SSIM. Because GAP-TV and DeSCI are optimized for spatial-varying mask problems, they perform generally worse than TwIST when converted to static mask problems, which mainly focused on optimizing the total variation. In a static mask problem, the equivalent movements of the camera sensor introduce a large
Fig. 3. Reconstruction results of Flying mnist dataset encoded by perfect mask and the comparison among (a) DeepCUP, (b) DeSCI, (c) GAP-TV, (d) TwIST, (e) Ground truth.

Fig. 4. Result of drop dataset comparison among (a) DeepCUP, (b) DeSCI, (c) GAP-TV, and (d) TwIST, (e) Ground truth.
Fig. 5. Result of Runner dataset comparison among (a) DeepCUP, (b)TwIST, (c)GAP-TV, and (d)DeSCI, (e) Ground truth.

**Table 2. Average PSNR and SSIM results**

| Methods  | Flying Mnist | Drop | Runner |
|----------|--------------|------|--------|
| DeepCUP  | PSNR: 30.63 SSIM: 0.933 | PSNR: 27.73 SSIM: 0.8458 | PSNR: 26.54 SSIM: 0.836 |
| TwIST    | PSNR: 26.45 SSIM: 0.7393 | PSNR: 27.72 SSIM: **0.8803** | PSNR: 24.84 SSIM: 0.764 |
| DeSCI    | PSNR: 18.32 SSIM: 0.6591 | PSNR: 22.51 SSIM: 0.7267 | PSNR: 21.83 SSIM: 0.63 |
| GAP-TV   | PSNR: 24.78 SSIM: 0.7393 | PSNR: 21.17 SSIM: 0.6875 | PSNR: 21.04 SSIM: 0.612 |

variation to the system [36]. The performance is also affected by an unchanged encoding mask. TwIST can solve the static mask SCI reconstruction decently, but a lot of details are missing during the reconstruction process. The proposed DeepCUP method can learn the static mask SCI reconstruction method as a global prior successfully, and produce the most reliable reconstruction.

4.2. Noise analysis

Most SCI reconstruction assumed the encoding process is perfectly binary shown in Fig. 6(a), from which an encoded result is formed by a series of sparse images.

In the real experiment, the imaged mask usually contains noise, aberration, and distortion, and the real mask is shown in Fig. 6(b). To analyze the robust performance of the DeepCUP method, the Mnist dataset was encoded by a mask that was captured from the experiment. The proposed deep learning method and other state-of-the-art methods were performed to reconstruct it and the results are shown in Fig. 7. PSNR and SSIM were also calculated as shown in Table 3.

From Fig. 7 and Table 3, we can see it can be seen that the proposed method outperforms other state-of-the-art methods in PSNR and SSIM. In the streak camera, the imaging light is going through optical-electrical-optical conversion, making the encoded results highly noisy. In such a scenario, the result of the forward model is no longer from a sparse image sequence, and the reconstruction result is highly jeopardized by using conventional compressed sensing methods. However, the proposed deep learning method can learn the reconstruction function
Fig. 6. Designed perfect mask and experimentally-captured mask.

Fig. 7. Comparison of the results of Flying Mnist experimentally-captured mask encoded datasets using (a) DeepCUP, (b) DeSCI, (c) GAP-TV, (d) TwIST, and (e) Ground truth.

Table 3. Average PSNR and SSIM results

| Methods | Flying Mnist (Real Encoding) |
|---------|------------------------------|
| DeepCUP | PSNR: 24.58, SSIM: 0.8651    |
| TwIST   | PSNR: 19.04, SSIM: 0.5848    |
| DeSCI   | PSNR: 18.42, SSIM: 0.6583    |
| GAP-TV  | PSNR: 18.42, SSIM: 0.6539    |
under a fixed deflection and a noise model by inputting data encoded by a specific flawed mask. The model finds the best fit of the reconstruction results. In this paper, the model assumes that it remains unchanged when applied to different datasets. As long as the mask with the same noise and aberration is applied to the real experiment, the model has the potential to handle the noise robustly. Thus, the proposed deep-learning based method can tolerate such system imperfections, by learning from a large augmented dataset including various distributions. Besides, computing efficiency is also much better. We took drop, Mnist datasets and the video in Ref. [27] as dynamic scenes to simulate the CUP system and then reconstructed it with TwIST, GAP, and Deep CUP, respectively. The reconstruction time required on a computer with RTX 2080Ti GPU and 11 GB RAM is summarized in Table 4.

| Dataset          | TwIST (s) | GAP (s) | DeepCUP (s) |
|------------------|-----------|---------|-------------|
| Drop             | 139.6     | 5.9     | 0.05763     |
| Mnist            | 45.84     | 4.6     | 0.05489     |
| Wang Li Hong Result [27] | 46.19 | 5.0     | 0.05618     |

Table 4 shows that DeepCUP is much more efficient than the other methods. Therefore, it is possible to implement an ultrafast imaging system for real-time applications.

4.3. Experiment

To test the generalization and robustness of the proposed method, we adopted the experimental data used in Ref. [27]. Although the decoding model was trained on Flying MNIST dataset, it can still readily recover transient videos with totally different contents. The video was encoded with an experimentally imaged mask containing both noises and aberrations. The recovered burst image still resembles the original video series as shown in Fig. 8.

In the real experiment, we imaged the process of femtosecond laser pulse passing through water tinted with milk as a scattering medium. In these experiments, to scatter light from the media to the CUP system, we evaporated dry ice into the light path in the air. We can observe the transmitting and scattering effects when the light interacts with milk solution as shown in Fig. 9. Comparing all the reconstruction results, the proposed method stands out in terms of noise rejection.

In our experiment, the shearing velocity of the streak camera was set to $v = 13.6 \text{ mm/ns}$ and the temporal resolution is 4ps. The spatially encoded, temporally sheared images were acquired by an internal CCD camera (ORCA-R2, Hamamatsu) with a sensor size of 1344×1024 binned pixels (2×2 binning; binned pixel size $d = 12.9 \mu m$). The reconstructed frame rate, $r$, was determined by $r = v/d$ to be nearly 100 billion frames per second.
5. Conclusions

In summary, we designed a novel reconstruction method based on deep learning to reconstruct the single compressed imaging. The UCF-101 dataset and Mnist dataset were adopted to train the networks, moreover, those datasets were also performed by a series of operations based on DeepCUP system and reconstructed it with the method of deep learning proposed in this paper and compressed sensing methods such as TwIST, GAP-TV, and DeSCI, and the results show that DeepCUP has better performances than the other methods in PSNR and SSIM. It is also applicable to the drop dataset, where the network has been trained by UCF-101 and Mnist datasets. Besides, the imaged mask which is obtained by streak camera usually contains noise, aberration, and distortion in the experiments, DeepCUP outperforms compressed sensing methods in terms of PSNR, SSIM and anti-noise performance. An experiment was conducted to image the transient process of the femtosecond laser passing through water tinted with milk, and the results show that DeepCUP can achieve nearly 100 billion frames per second.

Funding

Postdoctoral Research Foundation of China.

Disclosures

The authors declare no conflicts of interest.

References

1. P. B. Corkum and F. Krausz, “Attosecond science,” Nat. Phys. 3(6), 381–387 (2007).
2. Z. Li, R. Zgadzaj, X. Wang, Y.-Y. Chang, and M. C. Downer, “Single-shot tomographic movies of evolving light-velocity objects,” Nat. Commun. 5(1), 3085 (2014).
3. C. Pei, S. Wu, D. Luo, W. Wen, J. Xu, J. Tian, M. Zhang, P. Chen, J. Chen, and R. Liu, “Traveling wave deflector design for femtosecond streak camera,” Nucl. Instrum. Methods Phys. Res., Sect. A 855, 148–153 (2017).
4. G. Gao, K. He, J. Tian, C. Zhang, J. Zhang, T. Wang, S. Chen, H. Jia, F. Yuan, L. Liang, X. Yan, S. Li, C. Wang, and F. Yin, “Ultrafast all-optical solid-state framing camera with picosecond temporal resolution,” Opt. Express 25(8), 8721–8729 (2017).
5. B. Heshmat, G. Satat, C. Barsi, and R. Raskar, “Single-shot ultrafast imaging using parallax-free alignment with a tilted lenslet array,” in CLEO: 2014, OSA Technical Digest (online) (Optical Society of America, 2014), paper Tu3E.7.
6. P. W. W. Fuller, “An introduction to high speed photography and photonics,” Imaging Sci. J. 57(6), 293–302 (2009).
7. Invisible Vision Ltd., UBSi-True 1 billion fps ultra-high-speed framing camera, invisiblevision.com, http://www.invisiblevision.com/pdf/UBSi_(1Bn_fps_Camera).pdf (2016).
8. Z. Li, R. Zgadzaj, X. Wang, Y.-Y. Chang, and M. C. Downer, “Single-shot tomographic movies of evolving light-velocity objects,” Nat. Commun. 5(1), 3085 (2014).
9. 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),” Nat. Photonics 8(9), 695–700 (2014).
10. T. Suzuki, F. Isa, L. Fujii, K. Hiroswa, K. Nakagawa, K. Goda, Ichiro Sakuma, and Fumihiko Kannari, “Sequentially timed all-optical mapping photography (STAMP) utilizing spectral filtering,” Opt. Express 23(23), 30512–30522 (2015).
11. 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).
12. 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. Di Carlo, and B. Jalali, “High-throughput single-microparticle imaging flow analyzer,” Proc. Natl. Acad. Sci. U. S. A. 109(29), 11630–11635 (2012).
13. Y. Lu, T. W. Wong, F. Chen, and L. Wang, “Compressed Ultrafast Spectral-Temporal Photography,” Phys. Rev. Lett. 122(19), 193904 (2019).
14. V. Andreas, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, “Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging,” Nat. Commun. 3(1), 745 (2012).
15. C. B. Schaffer, N. Nishimura, E. N. Glezer, A. M. T. Kim, and E. Mazur, “Dynamics of femtosecond laser induced breakdown in water from femtoseconds to microseconds,” Opt. Express 10(3), 196–203 (2002).
16. G. Garijo, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, “Single-photon sensitive light-in-fight imaging,” Nat. Commun. 6(1), 6021 (2015).
17. A. Stolow, A. E. Bragg, and D. M. Neumark, “Femtosecond time-resolved photoelectron spectroscopy,” Chem. Rev. 104(4), 1719–1758 (2004).
18. P. Hockett, C. Z. Bisgaard, O. J. Clarkin, and A. Stolow, “Time-resolved imaging of purely valence-electron dynamics during a chemical reaction,” Nat. Phys. 7(8), 612–615 (2011).
19. R. Trebino, K. W. DeLong, D. N. Fittinghoff, J. N. Sweetser, M. A. Krumbügel, B. A. Richman, and D. J. Kane, “Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating,” Rev. Sci. Instrum. 68(9), 3277–3295 (1997).
20. J. L. Prince and J. M. Links, “Medical Imaging Signals and Systems,” (Pearson Prentice Hall, 2006), chap. 5, 328–332.
21. J. Bercoff, M. Tanter, and M. Fink, “Supersonic shear imaging: A new technique for soft tissue elasticity mapping,” IEEE Trans. Ultrason. Ferroelec. Freq. Control 51(4), 396–409 (2004).
22. A. Wax and V. Backman, “Biomedical Applications of Light Scattering (McGraw-Hill Professional),” 2009.
23. C. Zhu and Q. Liu, “Review of Monte Carlo modeling of light transport in tissues,” J. Biomed. Opt. 18(5), 050902 (2013).
24. I. Bargigia, A. Tosi, A. B. Shehata, A. D. Frera, A. Farina, A. Bassi, P. Taroni, A. D. Mora, F. Zappa, R. Cubeddu, and A. Pifferi, “Time-resolved diffuse optical spectroscopy up to 1700nm by means of a time-gated InGaAs/InP single-photon avalanche diode,” Appl. Spectrosc. 66(8), 944–950 (2012).
25. J. D. Gunton, A. Shiryayev, and D. L. Pagan, “Protein Condensation: Kinetic Pathways to Crystallization and Disease” (Cambridge Univ. Press, 2007).
26. J. Bioucas-Dias and M. Figueiredo, “A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration,” IEEE Trans Image Process 16(12), 2992–3004 (2007).
27. 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).
28. J. Liang, L. Zhu, and L. V. Wang, “Single-shot real-time femtosecond imaging of temporal focusing,” Light: Sci. Appl. 7(1), 42 (2018).
29. J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, “Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse,” Sci. Adv. 3(1), e1601814 (2017).
30. Q. Qi, S. Zhang, C. Yang, Y. He, F. Cao, J. Yao, P. Ding, L. Gao, T. Jia, J. Liang, Z. Sun, and L. V. Wang, “Single-shot compressed ultrafast photography: a review,” Adv. Photonics 2(1), 014003 (2020).
31. C. Yang, D. Qi, F. Cao, Y. He, X. Wang, W. Wen, J. Tian, T. Jia, Z. Sun, and S. Zhang, “Improving the image reconstruction quality of compressed ultrafast photography via an augmented Lagrangian algorithm,” J. Opt. 21(3), 035703 (2019).
32. S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2862–2869 (2014).
33. L. Zhu, Y. Chen, J. Liang, L. Gao, C. Ma, and L. V. Wang, “Improving image quality in compressed ultrafast photography with a space- and intensity constrained reconstruction algorithm,” Proc. SPIE 9720, High-Speed Biomedical Imaging and Spectroscopy: Toward Big Data Instrumentation and Management, 972008 (2016).
34. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Foundations and Trends in Machine Learning 3(1), 1–122 (2010).
35. X. Yuan, “Generalized alternating projection based total variation minimization for compressive sensing,” IEEE International Conference on Image Processing (ICIP), 2539–2543 (2016).
36. Y. Liu, X. Yuan, J. Suo, D. Brady, and Q. Dai, “Rank minimization for snapshot compressive imaging,” IEEE Trans. Pattern Anal. Mach. Intell. 41(12), 2990–3006 (2019).
37. “Runner data,” https://www.videvo.net/video/elite-runner-slow-motion/4541/.
38. “Drop data,” http://www.phantomhighspeed.com/Gallery.