Institut national de la recherche scientifique

Université du Québec

Centre Énergie Matériaux Télécommunications

Development and Applications of Single-shot Compressed Optical-Streaking Ultrahigh-speed Imaging

Par

Xianglei Liu

Thèse présentée pour l'obtention
du grade de Philosophiae Doctor (Ph.D.)
en sciences de l'énergie et des matériaux

Jury d'évaluation

Président du jury et examinateur interne	Professeur Fiorenzo Vetrone INRS-ÉMT
Examinateur externe	Professeur Qiyin Fang Université McMaster
Examinateur externe	Docteur Joel N. Bixler Laboratoire de recherche de l'armée de l'air
Directeur de recherche	Professeur Jinyang Liang INRS-ÉMT

DEDICATION

Dedicated to my all teachers,
to my all friends,
to my beloved family members.

ACKNOWLEDGEMENTS

I would like to express my gratitude to many people without whom I cannot finish this dissertation. First and foremost, my greatest appreciation goes to my supervisor, Prof. Jinyang Liang, for his guidance, unfailing availability, encouragement, and all his many contributions to my research. As his first Ph.D. student, I feel a lot proud. In addition, I am grateful to Prof. Fiorenzo Vetrone, Prof. Tiago H. Falk, Prof. Aycan Yurtsever, Prof. Tsuneyuki Ozaki, and Prof. Shian Zhang for their insights and solid support for my research. I appreciate Dr. Artiom Skripka, Dr. Chengshuai Yang, João Monteiro, Isabela Albuquerque, and Yujia Xue for many valuable discussions and enjoyable cooperation.

I would like to thank my laboratory colleagues Patrick Kilcullen, Cheng Jiang, Jingdan Liu, Yingming Lai, Miao Liu, Siqi Wang, Haojia Sun, Qianwan Yang, Xiaowen Shi, Miguel Marquez for their assistance and helpful suggestions. I would like to thank my college colleagues, these include but are not limited to Xin Liu, Shi Li, Lifa Ying, Tingzhou Yang, Na Xu, Dalin Cui, Jipeng Hui, Jiabin Liu, Chao Wang, Hui Zhang, Jiyun Chen, Yuting Lei, Mengyang Fan, Minghui Hao, Haixia Wang, Naser Mohammadi, Xiaoying Zheng, Xin Chai, Xavier Ropagnol, Qinzhe Zhang, Fan Yang, Shengyun Huang, Ting Yu, Yong Wang, Chen Wang, Ruiqi Yang, Guolong Song, Wanting He, Qiliang Wei, Yanqin Fu, Zonghua Pu, Mingjie Wu, Fuyong Yue, Jin Xin, Junliang Dong, Pei You for their help.

I acknowledge Prof. Fiorenzo Vetrone, Prof. Qiyin Fang, and Dr. Joel Bixler for their services as members of my dissertation committee. I would also like to acknowledge the funding support from Natural Sciences and Engineering Research Council (NSERC), Fonds de Recherche du Québec–Nature et Technologies (FRQNT), Ministère des Relations Internationales et de la Francophonie du Québec, Computer Canada, and Calcul Québec, and John R. Evans Leader’s Fund from Canada Foundation of Innovation to name a few. Finally, I would like to give my special thanks to my family members for their guidance and emotional support.

ABSTRACT

Higher temporal resolution has been extensively pursued in photography for decades. Although currently charge-coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) technologies have revolutionized high-speed and ultrahigh-speed photography, further increasing their temporal resolution is fundamentally limited by the bandwidth of the electronics and on-chip storage. Existing ultrahigh-speed CCD/CMOS cameras have achieved the temporal resolution of the sub-microsecond level at the cost of pixel count, size, and price.

Compressed ultrafast photography (CUP) is an emerging two-dimensional (2D) computational imaging modality that synergistically combines compressed sensing (CS) with streak imaging. CUP enables capturing non-repetitive time-evolving events at picosecond-level temporal resolution and has led to a variety of exciting discoveries and applications in physics such as the observation of optical chaos, dissipative soliton dynamics, and photonics Mach cone. Despite the salient advantage in temporal resolution, CUP cannot record long-lasting dynamics (e.g., upconversion luminescence processes on the order of microseconds and milliseconds) in a single shot, due to optoelectronic sweeping time of less than nanoseconds. Furthermore, a limited quantum efficiency (QE) and the space-charge effect in the optoelectronic streak cameras restrain the signal-to-noise ratios (SNRs) and spatial resolution of the CUP's measurement, respectively. To overcome these limitations, this dissertation focuses on efforts in developing cost-efficient and compact ultrahigh-speed imaging (i.e., sub-microsecond temporal resolution) hardware, high-fidelity image reconstruction software, and pertinent applications.

First, single-shot ultrahigh-speed imaging is of great significance to capture transient phenomena in physics, biology, and chemistry in real time. Existing techniques, however, have a restricted application scope, a low sequence depth, or a limited pixel count. To overcome these limitations, we developed single-shot compressed optical-streaking ultrahigh-speed photography (COSUP) with an imaging speed of 1.5 million frames per second (fps), a sequence depth of 500 frames, and an (x, y) pixel count of 0.5 megapixels per frame. COSUP's single-shot ultrahigh-speed imaging ability was demonstrated by recording single laser pulses illuminating through transmissive targets and by tracing a fast-moving object. As a universal imaging platform, COSUP is capable of increasing imaging speeds of a wide range of CCD and CMOS cameras by four orders of magnitude. We envision COSUP to be applied in widespread applications in biomedicine and materials science.

Second, COSUP is applied to biophotonics. The photoluminescence lifetime imaging of upconverting nanoparticles is increasingly featured in recent progress in optical thermometry. Despite

remarkable advances in photoluminescent temperature indicators, existing optical instruments lack the ability of wide-field photoluminescence lifetime imaging in real time, thus falling short in dynamic temperature mapping. Here, we further develop COSUP to single-shot photoluminescence lifetime imaging thermometry (SPLIT) for video-rate upconversion temperature sensing in a wide field. SPLIT first records wide-field luminescence intensity decay compressively in two views in a single shot. Then, an algorithm, built upon the plug-and-play alternating direction method of multipliers, is used to reconstruct the video, from which the extracted lifetime distribution is converted to a temperature map. Using the core/shell $NaGdF_4:Er^{3+}$, $Yb^{3+}/NaGdF_4$ upconverting nanoparticles as the lifetime-based temperature indicators, we apply SPLIT in longitudinal wide-field temperature monitoring beneath a thin scattering medium. SPLIT also enables video-rate temperature mapping of a moving biological sample at single-cell resolution.

Third, we developed a snapshot-to-video autoencoder (S2V-AE) for COSUP. S2V-AE is a new deep neural network that maps a compressively recorded 2D image to a movie. The S2V-AE preserves spatiotemporal coherence in reconstructed videos and presents a flexible structure to tolerate changes in input data. Implemented in compressed ultrahigh-speed imaging, the S2V-AE enables the development of single-shot machine-learning assisted real-time (SMART) COSUP, which features a reconstruction time of 60 ms and a large sequence depth of 100 frames. SMART-COSUP is applied to wide-field multiple-particle tracking at 20 thousand frames-per-second (kfps). As a universal computational framework, the S2V-AE is readily adaptable to other modalities in high-dimensional compressed sensing. SMART-COSUP is also expected to find wide applications in applied and fundamental sciences

Fourth, the concept of COSUP has been extended to electron microscopy. Bringing ultrafast (nanosecond and below) temporal resolution to transmission electron microscopy (TEM) has historically been challenging. Despite significant recent progress in this direction, it remains difficult to achieve sub-nanosecond temporal resolution with a single electron pulse, in real-time imaging. To address this limitation, here, we propose a methodology that combines laser-assisted TEM with computational imaging methodologies based on CS. In this technique, a 2D transient event [i.e. (x, y) frames that vary in time] is recorded through a CS paradigm, which consists of spatial encoding, temporal shearing via streaking, and spatiotemporal integration of an electron pulse. The 2D image generated on a camera is used to reconstruct the datacube of the ultrafast event, with 2D in space and 1D in time, via a CS-based image reconstruction algorithm. Using numerical simulation, we find that the reconstructed results are in good agreement with the ground truth, which demonstrates the applicability of CS-based computational imaging methodologies to laser-assisted TEM. Our proposed method, complementing the existing ultrafast stroboscopic and nanosecond single-shot techniques, opens up the possibility for single-shot, real-time, spatiotemporal imaging of irreversible structural phenomena with sub-nanosecond temporal resolution.

Keywords: computational imaging; compressed sensing; deep learning; temperature sensing; high-dimensional data acquisition; ultrahigh-speed photography

RÉSUMÉ

Une résolution temporelle plus élevée a été largement recherchée en photographie depuis des décennies. Bien qu'actuellement, les technologies de dispositifs à couplage de charge (CCD) et de métal-oxyde-semi-conducteur (CMOS) aient révolutionné la photographie à haute et ultra-haute vitesse, l'augmentation supplémentaire de leur résolution temporelle est fondamentalement limitée par leur bande passante d'électronique et de stockage sur puce. Les caméras CCD/CMOS ultra-rapides existantes ont atteint une résolution temporelle inférieure à la microseconde au détriment du nombre de pixels, de la taille et du prix.

La photographie ultrarapide compressée (CUP) est une nouvelle modalité d'imagerie computationnelle bidimensionnelle (2D) qui combine de manière synergique la détection compressée (CS) avec l'imagerie par stries. CUP permet de capturer des événements évoluant dans le temps non répétitif à une résolution temporelle de niveau picoseconde et a conduit à une variété de découvertes et d'applications passionnantes en physique telles que l'observation du chaos optique, la dynamique dissipative des solitons et le cône de Mach photonique. Malgré l'avantage saillant de la résolution temporelle, CUP ne peut pas enregistrer une dynamique de longue durée (par exemple, des processus de luminescence de conversion ascendante de l'ordre de la microseconde et de la milliseconde) en un seul coup, en raison du temps de balayage optoélectronique inférieur à la nanoseconde. De plus, une efficacité quantique (QE) limitée et l'effet de charge d'espace dans les caméras optoélectroniques à balayage limitent respectivement les rapports signal sur bruit (SNR) et la résolution spatiale de la mesure du CUP. Pour surmonter ces limitations, cette thèse se concentre sur les efforts de développement de matériel d'imagerie à ultra-haute vitesse (c'est-à-dire une résolution temporelle inférieure à la microseconde), d'un logiciel de reconstruction d'image haute-fidélité et d'applications pertinentes.

Tout d'abord, l'imagerie ultra-rapide à un seul coup est d'une grande importance pour capturer les phénomènes transitoires en physique, biologie et chimie en temps réel. Les techniques existantes, cependant, ont une portée d'application restreinte, une faible profondeur de séquence ou un nombre de pixels limité. Pour surmonter ces limitations, nous avons développé la photographie à ultra-haute vitesse à stries optiques compressées (COSUP) avec une vitesse d'imagerie de 1.5 million d'images par seconde (fps), une profondeur de séquence de 500 images et un pixel (x, y) compte de 0.5 mégapixels par image. La capacité d'imagerie ultra-rapide à un seul coup de COSUP a été démontrée en enregistrant des impulsions laser uniques éclairant à travers des cibles transmissives et en traçant un objet en mouvement rapide. En tant qu'une plate-forme d'imagerie universelle, COSUP est capable d'augmenter les vitesses d'imagerie d'une large gamme de caméras CCD et CMOS de quatre ordres de grandeur. Nous

envisageons que COSUP soit appliqué dans des applications généralisées en biomédecine et en science des matériaux.

Deuxièmement, COSUP est appliqué à la biophotonique. L'imagerie à vie par photoluminescence des nanoparticules à conversion ascendante est de plus en plus présente dans les progrès récents de la thermométrie optique. Malgré des progrès remarquables dans les indicateurs de température photoluminescents, les instruments optiques existants n'ont pas la capacité d'imagerie à vie de photoluminescence à champ large en temps réel, ce qui ne permet pas de cartographier dynamiquement la température. Ici, nous développons davantage COSUP en thermométrie d'imagerie à durée de vie par photoluminescence (SPLIT) à un seul coup pour la détection de température de conversion ascendante à débit vidéo dans un champ large. SPLIT enregistre d'abord la décroissance de l'intensité de la luminescence à champ large de manière compressive dans deux vues en une seule exposition. Ensuite, un algorithme, basé sur la méthode de direction alternative plug-and-play des multiplicateurs, est utilisé pour reconstruire la vidéo, à partir de laquelle la distribution de durée de vie extraite est convertie en une carte de température. En utilisant les nanoparticules de conversion ascendante $NaGdF_4:Er^{3+}$, $Yb^{3+}/NaGdF_4$ noyau/coquille comme indicateurs de température basés sur la durée de vie, nous appliquons SPLIT dans la surveillance longitudinale de la température à champ large sous un milieu de diffusion mince. SPLIT permet également la cartographie vidéo de la température d'un échantillon biologique en mouvement à une résolution unicellulaire.

Troisièmement, nous avons développé un auto-encodeur instantané vers vidéo (S2V-AE) pour COSUP. S2V-AE est un nouveau réseau neuronal profond qui mappe une image 2D enregistrée par compression à un film. Le S2V-AE préserve la cohérence spatio-temporelle des vidéos reconstruites et présente une structure flexible pour tolérer les changements dans les données d'entrée. Mis en œuvre dans l'imagerie ultra-rapide compressée, le S2V-AE permet le développement d'un COSUP en temps réel assisté par apprentissage automatique (SMART) à un seul coup, qui présente un temps de reconstruction de 60 ms et une grande profondeur de séquence de 100 images. SMART-COSUP est appliqué au suivi de particules multiples à champ large à 20, 000 images par seconde. En tant que cadre de calcul universel, le S2V-AE est facilement adaptable à d'autres modalités de détection compressée à haute dimension. SMART-COSUP devrait également trouver un large champ d’applications dans les sciences appliquées et fondamentales

Quatrièmement, le concept de COSUP a été étendu à la microscopie électronique. Apporter une résolution temporelle ultrarapide (nanoseconde et inférieure) à la microscopie électronique à transmission (TEM) a toujours été un défi. Malgré des progrès récents significatifs dans cette direction, il reste difficile d'atteindre une résolution temporelle inférieure à la nanoseconde avec une seule impulsion d'électron, en

imagerie en temps réel. Pour remédier à cette limitation, nous proposons ici une méthodologie qui combine la TEM assistée par laser avec des méthodologies d'imagerie computationnelle basées sur CS. Dans cette technique, un événement transitoire 2D [c.-à-d. (x, y) trames qui varient dans le temps] est enregistré à travers un paradigme CS, qui consiste en un codage spatial, un cisaillement temporel via des stries et une intégration spatio-temporelle d'une impulsion électronique. L'image 2D générée sur une caméra est utilisée pour reconstruire le cube de données de l'événement ultrarapide, avec deux dimensions spatiales et une dimension temporelle, via un algorithme de reconstruction d'image basé sur CS. À l'aide de la simulation numérique, nous constatons que les résultats reconstruits sont en bon accord avec la vérité terrain, ce qui démontre l'applicabilité des méthodologies d'imagerie computationnelle basées sur CS à la TEM assistée par laser. Notre méthode proposée, complétant les techniques stroboscopiques ultrarapides et nanosecondes existantes, ouvre la possibilité d'une imagerie spatio-temporelle en temps réel de phénomènes structurels irréversibles avec une résolution temporelle inférieure à la nanoseconde.

Mots-clés : imagerie computationnelle; détection compressée; apprentissage en profondeur; détection de température; acquisition de données haute dimension; ultra-rapide photographie

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	One-dimensional
2D	Two dimensional
3D	Three-dimensional
mm	Millimeter
μm	Micrometer
nm	Nanometer
ms	Millisecond
μs	Microsecond
ns	Nanosecond
Hz	Hertz
fps	Frames per second
kfps	Thousand fps
Mfps	Million frames per second
CS	Compressed sensing
CCD	Charge-coupled device
CMOS	Complementary metal-oxide-semiconductor
SNRs	Signal to noise ratios
QE	Quantum efficiency
CUP	Compressed ultrafast photography
COSUP	Compressed optical-streaking ultra-high-speed photography
SMART-COSUP	Single-shot machine-learning assisted real-time COSUP
AE	Autoencoder
S2V-AE	Snapshot-to-video AE
TEM	Transmission electron microscopy
CUTEM	Compressed ultrafast TEM
DS-CUTEM	Dual-shearing CUTEM
CACTI	Coded aperture compressive temperature imaging
P2C2	Programmable pixel compressive camera
MA	Multiple-aperture
DMD	Digital micromirror device
LCOS	Liquid crystal on the silicon

GS	Galvanometer scanner
SPI	Single-pixel imaging
CI	Computational imaging
CSI	Compressed spatial imaging
CT	Computational tomography
CTI	Compressed temporal imaging
TwIST	Two steps iteration shrink/thresholding
PnP	Plug-and-play
ADMM	Alternative direction multiplier methods
BM3D	Block-matching and 3D filtering
TV	Total variation
RMSEs	Root-mean-square errors
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity index measure
PLI	Photoluminescence lifetime imaging
UCNPs	Upconverting nanoparticles
TCSPC	Time-correlated single-photon counting
FOV	Field of view
EM	Electron-multiplying
TTL	Transistor-transistor logic
XRD	X-ray powder diffraction
CUSP	Compressed ultrafast spectral photography
FLIM	Fluorescence lifetime imaging
LLE	Lossless-encoding
PpLIM	Phosphorescence lifetime imaging microscopy
DNNs	Deep neural networks
MLPs	Multi-layer perceptrons
GANs	Generative adversarial networks
ReLU	Rectified linear unit
Bi-LSTM	Bi-directional long short-term memory
BN	Batch normalization

LIST OF PUBLICATIONS

Peer-reviewed Journals:

[1] Liu, Xianglei, et al. "Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging." Nature Communications 12 (2021): 6401.

[2] Liu, Xianglei, et al. "Single-shot real-time compressed ultrahigh-speed imaging enabled by a snapshot-to-video autoencoder." Photonics Research 9.12 (2021): 2464-2474.

[3] Liu, Xianglei, et al. "Single-shot compressed optical-streaking ultra-high-speed photography." Optics Letters 44 (2019): 1387-1390. [Featured on the front page of the Optical Society of America (now Optica)]

[4] Liu, Xianglei, et al. "Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: Analytical modeling and simulation." Micron 117 (2019): 47-54.

[5] Liu, Jingdan, Charlotte Zaouter, Xianglei Liu, et al. "Coded-aperture broadband light field imaging using digital micromirror devices." Optica 8.2 (2021): 139-142.

[6] Lai, Yingming, Yujia Xue, Christian-Yves Côté, Xianglei Liu, et al. "Single-Shot Ultraviolet Compressed Ultrafast Photography." Laser & Photonics Reviews 14.10 (2020): 2000122.

[7] Lai, Yingming, Ruibo Shang, Christian-Yves Côté, Xianglei Liu, et al. "Compressed ultrafast tomographic imaging by passive spatiotemporal projections." Optics Letters 46.7 (2021): 1788-1791.

[8] Jiang, Cheng, Patrick Kilcullen, Xianglei Liu, et al. "Real-time High-speed Three-dimensional Surface Imaging using CoaXPress-Interfaced Band-Limited Illumination Profilometry" Optics Letters 45.4 (2020): 964-967.

[9] Liu, Jingdan, Yingming Lai, Heide Ibrahim, Katherine Légaré, Xianglei Liu, et al., “Swept coded aperture real-time femtophotography (SCARF)”, under review in Nature Photonics.

[10] Marquez, Miguel, Yingming Lai, Xianglei Liu, et al., “Deep-learning supervised snapshot compressive imaging enabled by an end-to-end adaptive neural network”, under review of IEEE Journal of Selected Topics in Signal Processing.

Conference proceedings:

[1] Liu, Xianglei, et al. " Snapshot-to-video autoencoder for compressed ultrahigh-speed imaging." submitted to OPTO (PW22O), Photonics West, 2022.

[2] Liu, Xianglei, et al. "Compressed-sensing two-dimensional rotating-mirror streak camera for single-shot ultrahigh-speed optical imaging." Photonics North (PN). IEEE, 2021.

[3] Liu, Xianglei, et al. "Single-shot real-time optical imaging using compressed optical-streaking ultra-high-speed photography." High-Speed Biomedical Imaging and Spectroscopy VI. Vol. 11654. International Society for Optics and Photonics, 2021.

[4] Liu, Xianglei, et al. "Compressed ultrafast transmission electron microscopy: a simulation study." Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXVI. Vol. 10883. International Society for Optics and Photonics, 2019.

1 INTRODUCTION

1.1 Motivation

Temporal resolution is an important parameter of photography [1]. To quantitatively assess the temporal resolution, imaging modalities are characterized by the concept of image speed, which is quantified by the frame rate with the unit of frames per second (fps). With early examples being well-known recordings in 1878 of a galloping horse [2] and the 1887 photograph of a supersonic bullet [3], photographers and scientists have continuously sought methods to capture transient scenes at a higher imaging speed [4].

Multiscale imaging speeds cater to different scientific research and application needs. For instance, high-speed imaging [e.g., up to thousands of fps] plays a key role in particle image velocimetry [5], video surveillance [6], and instant replay [7], to name a few. Ultrahigh-speed imaging (e.g., up to millions of fps) is necessary for observing neural activities [8], phosphorescence light emission [9], conformational changes in proteins [10], etc. Ultrafast imaging (e.g., billions of fps) opens scientific research such as monitoring molecular behavior (e.g., molecular cinema) [11], characterizing the fine structure of matter [12], and studying chemical and biological processes [13] that have been inaccessible before.

Single-shot ultrahigh-speed imaging is indispensable for visualizing various microsecond-level phenomena occurring in two-dimensional (2D) space. Existing single-shot ultrahigh-speed imaging techniques can be generally categorized into active-detection and passive-detection domains. The active-detection approaches exploit specially designed pulse trains to probe 2D transient events. However, these methods are not suitable for imaging self-luminescent and color-selective events. By contrast, the passive-detection approaches leverage receive-only ultrahigh-speed detectors to record photons scattered and emitted from the transient scenes. Examples include charge-coupled device (CCD) [14] and complementary metal-oxide-semiconductor (CMOS) [15] technologies that have revolutionized high-speed and ultrahigh-speed photography. Despite these sensors' widespread impact, further increasing frame rates using CCD or CMOS technology is fundamentally limited by the bandwidth of the electronics and on-chip storage [16]. Commercial ultrahigh-speed CCD/CMOS cameras either use customized sensors [17] with expensive prices as shown in Table 1.1 or rely on bulky camera arrays [18].

Compressed ultrafast photography (CUP) [19-21] is an emerging computational imaging modality that synergistically combines compressed sensing (CS) [22] with streak imaging [23].

Table 1.1 Typical commercial high-speed cameras

Part number	*Price (USD)*[Note 1]	*Imaging speed (kfps)*	*Pixel count*	*Manufacturer*
FASTCAM SA-Z	~150,000	20	1,000,000	Photron
i-SPEED 727	~134,000	50	500,000	iX-Cameras
HPV-X2	~250,000	10,000	100,000	Shimadzu

[Note 1] the prices are based on the market of date December 2021.

CUP is renowned for its two-dimensional (2D) dynamic imaging with an image speed of up to 10^{13} fps. Compared to pump-probe-based ultrafast imaging methods (including but not limited to strobe photography) [24-26], CUP enables recording non-repetitive time-evolving events in a single snapshot. So far, CUP has led to a variety of exciting applications in physics such as the observation of optical chaos [27], dissipative soliton dynamics [28], and photonics Mach cone [20]. Despite the salient advantages in imaging speed, CUP cannot record long-lasting dynamics (e.g., upconversion luminescence processes on the order of microseconds and milliseconds) in a single shot, due to optoelectronic sweeping time of less than nanoseconds. Furthermore, due to the photon-to-photoelectron conversion by the photocathode, the quantum efficiency (QE) of the optoelectronic streak cameras is typically <15% for visible light, which limits the quality of acquired data.

A mechanical streak camera, using a one-dimensional (1D) slit placed at its entrance, usually uses a rotating mirror [e.g., a galvanometer scanner (GS) or a polygon mirror] to deflect the light across a 2D imaging plane of an off-the-shelf CCD/CMOS sensor. Since the mechanical sweeping is much slower than the optoelectronic counterpart in CUP, the mechanical streak camera has a longer exposure time to record long-lasting dynamics (i.e., up to milliseconds) and has tunable temporal resolution typically from hundreds of nanoseconds to microseconds to satisfy ultrahigh-speed imaging. Moreover, its all-optical data acquisition allows the flexibly implementing of many high-sensitivity cameras [e.g., electron-multiplying (EM) CCD and scientific CMOS cameras, whose QEs can be >90% for visible light] to obtain superior SNRs in measurements. Thus, the mechanical streak camera becomes a good candidate to overcome the limitations of the size and price in existing ultrahigh-speed CCD/CMOS cameras and to address the limitations of QE and sweeping time in CUP. However, the mechanical streak is a 1D imaging technique due to the entrance slit.

1.2 Objectives

This dissertation focuses on efforts in developing cost-efficient and compact ultrahigh-speed imaging (i.e., sub-microsecond temporal resolution) hardware, high-fidelity image reconstruction software, and pertinent applications. First, we plan to develop a CS-based ultrahigh-speed imaging modality adapted from a mechanical streak camera but with 2D imaging ability in a single shot. To overcome the limitations in existing ultrahigh-speed cameras, the proposed imaging modality targets the image speed over one million fps (i.e., sub-microsecond level temporal resolution), a sequence depth over 100 frames (i.e., millisecond exposure time), and a pixel count in each frame over 100,000 (e.g., Table 1.1). Using these capabilities, we will be able to characterize the behavior of laser pulses through specific targets and track a fast-moving object.

Second, we aim to apply the newly developed ultrahigh-speed imaging system to novel applications with a specific focus on advanced lifetime characterizations in photoluminescence nanoparticles. To enable this application, the deployed system needs to satisfy the requirements of microsecond-level temporal resolution and high-sensitivity low light imaging, high-fidelity video reconstruction algorithm, and temperature-related upconversion nanoparticle as the indicator.

Third, the CS-based video reconstruction using the analytical-modeling-based method is limited by the long processing time and unstable image quality [29]. To resolve these problems, deep learning provides an end-to-end video reconstruction avenue as well as good video reconstruction quality. Thus, we intend to use deep learning methods for video reconstruction in the proposed ultrahigh-speed imaging system. Compared with existing learning-based methods, the new deep neural network aims to have the largest sequence depth (i.e., 100 frames) in a reconstructed video. Furthermore, the new neural network contributes to the development of next-generation ultrahigh-speed cameras, which will be applied to on-time feedback applications like multiple-particle tracking [30].

Fourth, bringing ultrafast imaging speed to transmission electron microscopy (TEM) has historically been challenging [31]. Despite significant recent progress in this direction, it remains difficult to achieve sub-nanosecond temporal resolution and nanometer-level spatial resolution with a single electron pulse in real-time imaging [32, 33]. To achieve this goal, we plan to establish analytical modeling of CS-based TEM, which contributes to the construction of dynamic TEM at INRS. Compared to the pump-probe methods, CS-based TEM has a distinguished advantage in the observation of non-repeatable transient dynamics.

1.3 Organization

This dissertation is organized as follows. Chapter 2 introduces several concepts including computational imaging (CI), single-pixel imaging (SPI), compressive spatial imaging (CSI), compressive temporal imaging (CTI), sparsity, incoherence, etc. Several typical examples of CTI are analyzed and compared. At the end of this chapter, typical CS algorithms in CTI are presented and discussed. Chapter 3 presents the prototype and proof-the-concept experiments of the proposed CS-based ultrahigh-speed imaging modality termed single-shot compressed optical-streaking ultrahigh-speed photography (COSUP). Chapter 4 presents COSUP's application to fast wide-field upconverting luminescence lifetime/temperature sensing. The proposed optical thermometry is called single-shot photoluminescence lifetime imaging thermometry (SPLIT). With the demonstrations in biological tissues with absorption as well as scattering, the SPLIT system records not only 2D lifetime images but wide field temperature sensing for a moving sample for the first time. Chapter 5 presents a new deep neural network, termed as Snapshot-to-Video Auto-Encoder (S2V-AE). With the S2V-AE, single-shot machine-learning assisted real-time (SMART)-COSUP is demonstrated by tracking multiple objects in real time. Chapter 6 presents the concept and configurations of compressed ultrafast (CU)-TEM and dual-shearing (DS) CUTEM. Conclusions and prospectives conclude this dissertation in Chapter 7. At the end of this dissertation is the summary in French.

2 METHODS

In this section, we first introduce the concept of CI to distinguish it from conventional imaging. We then briefly present the concept of single-pixel imaging (SPI) as a typical compressed spatial imaging (CSI) modality, because SPI has well-established CS-based theory analysis and successful applications. For instance, two premises of CS, i.e., sparsity and incoherence are covered here. Upon the understanding of SPI, we naturally extend the topic to compressed temporal imaging (CTI), which is the core of this dissertation. We elaborate on several typical CTI modalities in the following order: system schematic, working principle, and forward model. Finally, we elaborate on three types of CTI reconstruction algorithms and compare their reconstruction speed, image quality, and flexibility.

2.1 Computational imaging (CI)

CI is the process of indirectly forming images from measurements using algorithms [34]. CI involves a tight integration of the “hardware encoder” and the “software decoder” to form the images of interest as shown in Fig. 2.1. In conventional imaging as shown in Fig. 2.1(a), one point on the object plane is imaged as one point on the image plane, following the geometry optics rule. Two adjacent points on the object plane cannot be sampled by one pixel on the image plane, following the Nyquist sampling theorem. In contrast, CI allows one pixel on the image plane to record different points on the object plane, introducing a hardware encoder as shown in Fig.2.1(b). A software decoder is used to retrieve the object and achieves better performances on the field of view (FOV), optical aberrations [35], and spatial bandwidth product [36], to name a few. CI covers a broad range of techniques including light field imaging [37], quantitative phase imaging [38], CSI, and CTI [39], to name a few.

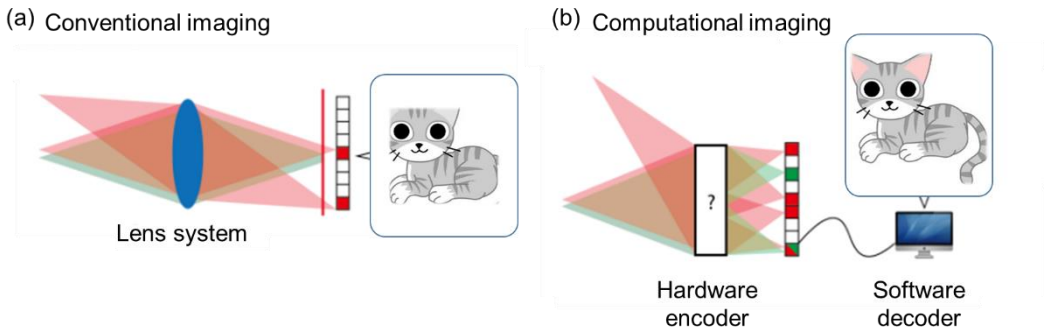


Figure 2.1 Schematics of conventional imaging (a) and computational imaging (b).

2.2 Compressed spatial imaging (CSI)

The SPI system, as shown in Fig. 2.2, employs a single-pixel detector (e.g., photodiode) at the receiving end, offering great potential for optical imaging at wavelengths where pixel-array detectors, such as CCD and CMOS cameras, are either not available or prohibitively expensive. SPI has been demonstrated with great success when operating with infrared light [40], terahertz waves [41], and even photoacoustic signals [42]. Instead of acquiring spatial information through parallel detection, SPI relies on using a digital micromirror device (DMD) to display a series of encoding patterns and then computationally reconstructs spatial information from a series of measurements [43]. The working principle of SPI is shown in Fig. 2.3. The object denoted by an image of the logo of "LACI" is expressed as a vector. Each encoding pattern is expressed as a vector placed at one row of the sensing matrix. Each measurement is the inner product between each encoding pattern and the object. Mathematically, the forward model of SPI can be modeled as a matrix-vector product form,

$$E = \Phi I. \tag{2.1}$$

Here $I = [x_1, x_2, x_3, \dots x_{\mathrm{N}}]^{\mathrm{T}}$ denotes the object, and $E = [y_1, y_2, y_3, \dots y_M]^{\mathrm{T}}$ is the measurement. Note that the number M of available measurements is less than the dimension N of the object. The sensing matrix $\Phi \in \mathbb{R}^{M \times N}$ consists of a series of vectors, each of which $\Phi_j \in \mathbb{R}^N$, represents the j^{th} encoding pattern and is employed to modulate the object, $j = 1,2, \dots M$. Equation 2.1 is an underdetermined linear system and the process of recovering I from the measurement E is ill-posed in general, which means there are many candidate solutions. To this end, we need to bring in CS [44]. Notably, CS relies on two principles: sparsity, which pertains to the object of interest, details in Section 2.2.1, and incoherence, which pertains to the sensing modality, details in Section 2.2.2. Despite the unique advantage across almost the entire spectrum range, the imaging speed of the-state-of-art SPI is relatively slow, e.g., 1,000 fps given pixel count 32×32 [45].

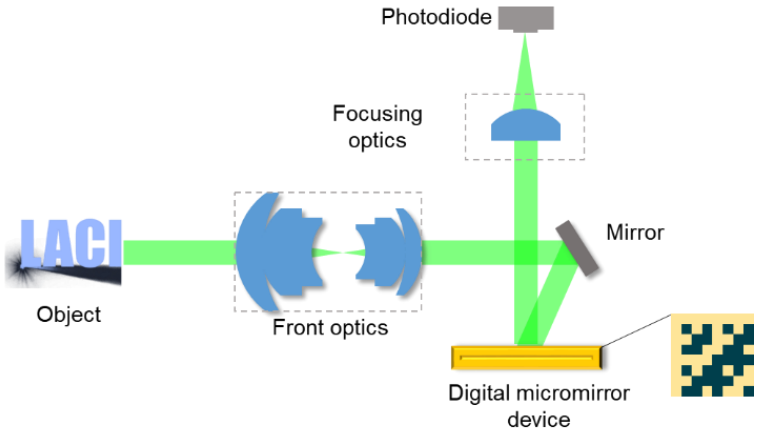


Figure 2.2 Schematic of single-pixel imaging. DMD sequentially displays a group of encoding patterns.

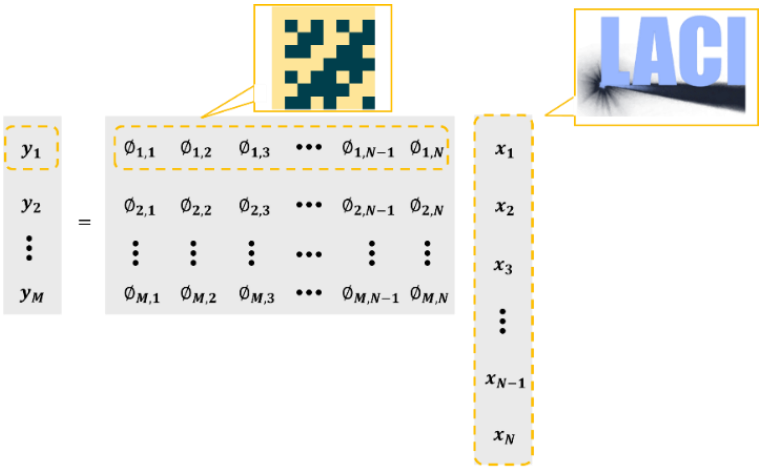


Figure 2.3 Working principle of single-pixel imaging with a dense sensing matrix.

2.2.1 Sparsity

Sparsity expresses the idea that the “information rate” of a continuous-time signal may be significantly smaller than suggested by its bandwidth, or that the degree of freedom of a discrete-time signal is much smaller than its (finite) length [44]. In other words, the sparsity of an image is the percentage of pixels with zero values. Specifically, CS utilizes the fact that many natural signals are sparse or compressible in the sense that they have concise representations when expressed in the proper basis ψ. Fortunately, most, if not all, natural signals have concise representations when expressed on a convenient basis, for instance, the image in Figure 2.4 (a) and its wavelet transform in (b). Despite nearly all the image pixels being nonzero values, the wavelet coefficients provide a concise summary: most coefficients are small, whereas relatively

few large coefficients contain most of the information. As a result of sparsity, one can discard small coefficients without much perceptual loss when a signal has a sparse expansion. Figure 2.4 (c) shows a case where the perceptual loss is barely noticeable between a megapixel image to its approximation obtained by throwing away 97.5% of the coefficients.

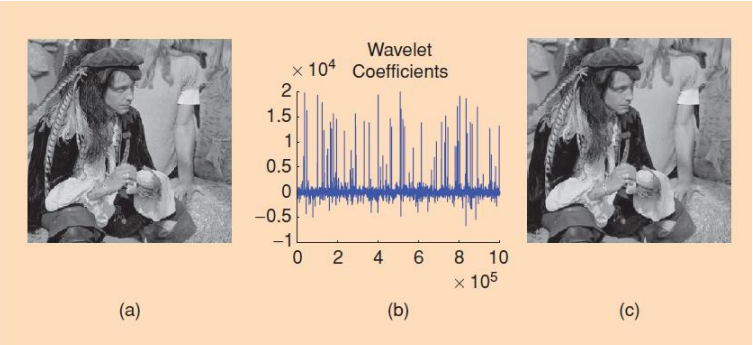


Figure 2.4 Illustration of sparsity. (a) Original megapixel image and (b) its wavelet transform coefficients. (c) The reconstruction was obtained by keeping 25,000 large coefficients [51].

2.2.2 Incoherence

Incoherence extends the duality between time and frequency [44] by expressing that signals having a sparse representation in the basis ψ must be spread out in the domain where they are acquired. For example, a Dirac or a spike is sparse in the time domain, at the same time, it is spread out in the frequency domain. In other words, incoherence indicates that unlike the signal of interest, the sampling/sensing waveforms are extremely densely represented in the basis ψ. It can be concluded that random matrices are largely incoherent with any fixed basis ψ. Using a random matrix as the sampling/sensing waveforms, e.g., white noise, is the standard way to conduct spatial encoding in the CS paradigm.

2.3 Compressed temporal imaging (CTI)

In contrast to CSI, CTI utilizes a pixel-array detector to record a transient scene in a compressive manner, ensuing in a CS-based algorithm to reconstruct the input transient scene. The typical CTI modalities include CUP and [4, 19-21, 46-59] and CACTI [60-64]. One distinguished difference between CACTI and CUP is their encoding scheme and thus their imaging speed. CACTI tries to actively change the encoding mask itself for generating a series of encoding patterns, while CUP tries to passively change encoding patterns with a static encoding mask. As

a result, the maximum imaging speed of the former is thousands of fps while that of the latter is over billions of fps.

2.3.1 Compressed ultrafast photography (CUP)

The system schematic of CUP [19] is shown in Fig. 2.5(a). Through a camera lens, beam splitter, tube lens, and objective, an object is imaged onto a DMD [42]. The modulated object is relayed into the streak camera by the objective, tube lens, and beam splitter. After the fully opened entrance of the streak camera [Fig. 2.5(b)], the modulated object mainly experiences a photon-to-photoelectron conversion by a photocathode, photoelectron sweeping by a pair of electrodes, and spatiotemporal integration of a CCD camera. With the prior information of the object's sparsity, CS-based image reconstruction is used to retrieve the object (e.g., a video). Figure 2.5(b) shows the field of view (FOV) of a streak camera is one-dimensional (1D), limited by a narrow entrance slit (10–50 μm wide). CUP enables the streak cameras to work with a fully opened slit. Figure 2.6 shows the working principle of CUP. Dynamics are encoded by a static encoding pattern. Each encoded frame experiences a temporal shearing operator, and then all spatially encoded and temporally sheared frames are integrated by a spatiotemporal integration operator. Mathematically, the forward model of CUP could be expressed as

$$E = \mathbf{TSC}\, I(x, y, t). \tag{2.2}$$

Here $\mathbf{C}$ denotes spatial encoding, $\mathbf{S}$ denotes temporal shearing, and $\mathbf{T}$ denotes spatiotemporal integration. $I(x, y, t)$ denotes dynamics, and E denotes the optical energy of measurement. Upon the prior knowledge of the operators and the sparsity of the transient scene, $I(x, y, t)$ can be recovered from the measurement E by solving the inverse problem of

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \| E - \mathbf{TSC} I \|_2^2 + \lambda \Phi(I) \right\} \tag{2.3}$$

Here $\|\cdot\|_2^2$ represents the l_2 norm, λ is a weighting coefficient, and $\Phi(\cdot)$ is a regularizer. In practice, $I(x, y, t)$ was recovered by using a CS-based algorithm.

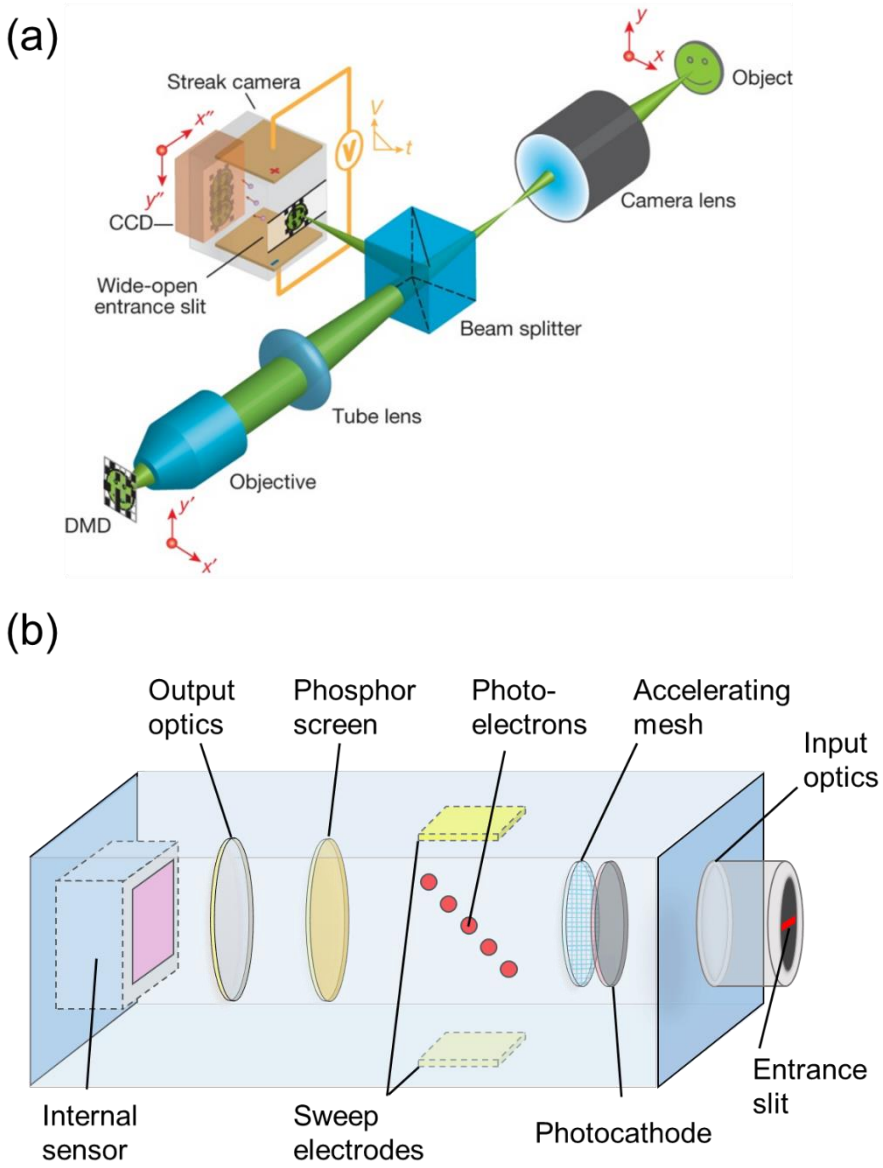


Figure 2.5 Schematics of compressed ultrafast photography (a) and a streak camera (b). DMD, digital micro-mirror device. CCD, charge-coupled device [17].

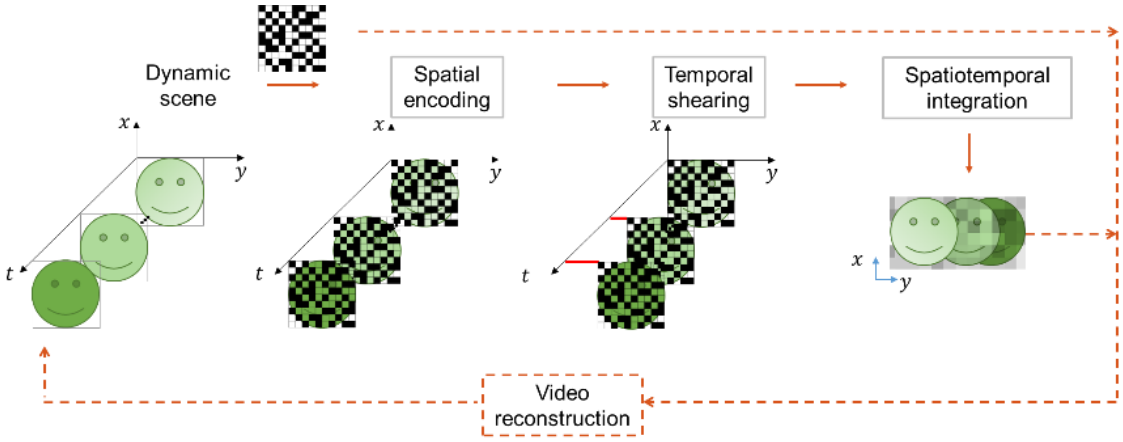


Figure 2.6 Working principle of CUP. The red line denotes the shifting distance along the shearing direction.

2.3.2 Coded aperture compressive temporal imaging (CACTI)

The system schematics of CACTI [60, 61, 63, 64] are shown in Fig. 2.7. A transient scene (e.g., an intensity-decay letter 'C') is imaged by front optics onto an encoding plane, in which a sequence of encoding patterns are generated by a moving printed mask loaded on a piezoelectric stage [Fig. 2.7(a)], a DMD [Fig. 2.7(b)], and an LCOS [Fig. 2.7(c)], respectively. The modulated scene is relayed to a CCD/CMOS camera by relay optics for conducting spatiotemporal integration. Figure 2.8 shows the working principle of CACTI. Each frame of video is modulated by a different encoding pattern. All modulated frames are integrated into a single measurement. Mathematically, the forward model could be expressed as

$$E = \mathbf{T}(\mathbf{SC})I(x, y, t). \tag{2.4}$$

Here $I(x, y, t)$ denotes dynamics and $(\mathbf{SC})$ denotes the spatial encoding and temporal shearing happening simultaneously. $\mathbf{T}$ denotes spatiotemporal integration. E denotes the optical energy of measurement. Compared to Eq. (2.2), CACTI could be viewed as a special case of CUP, where the spatial encoding and temporal shearing occur simultaneously, and CACTI shares the same reconstruction algorithms with CUP.

To compare with CSI, the forward model of CTI could also use matrix-vector product form to express

$$E = \hat{\Phi} I. \tag{2.5}$$

Here, $\hat{\Phi} = \mathbf{TSC}$ or $\mathbf{T}(\mathbf{SC})$ denotes sensing matrix. Figure 2.9 illustrates the i^{th} frame in the video and the i^{th} encoding pattern, $i = 1,2,3 \ldots N_t$. It is worth noting that the sensing matrix is the

concatenation of N_t diagonal matrices. Each diagonal matrix is reshaped from the corresponding encoding pattern. Compared to the dense sensing matrix in CSI in Fig. 2.3, this type of sensing matrix is sparse. Although the problem can be cast as a CS problem, existing CS theory cannot be employed directly to study the CTI, due to the special structure of the sensing matrix. To this end, Ref. [65] proposed a compression-based framework for the theoretical analysis of CTI systems. CACTI leveraged an active way to change encoding patterns with a changing rate higher than an off-the-shelf CCD or CMOS camera's image speed. Since the changing rate of encoding patterns is inherently limited by either the moving speed of a piezoelectric stage or the refreshing rate of DMD (or LCOS), the imaging speeds of these imaging modalities are clamped at several thousand fps.

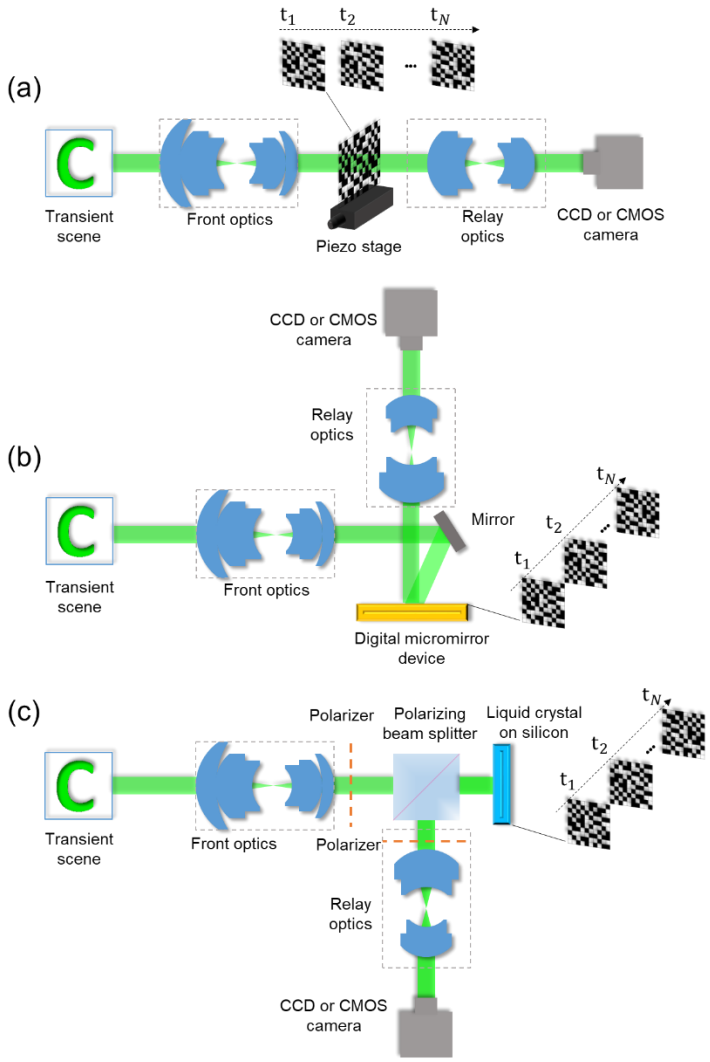


Figure 2.7 Schematics of coded aperture compressive temporal imaging. The encoding mask is loaded on a piezoelectric stage (a), a digital micro-mirror device (b), and a liquid crystal on silicon (c).

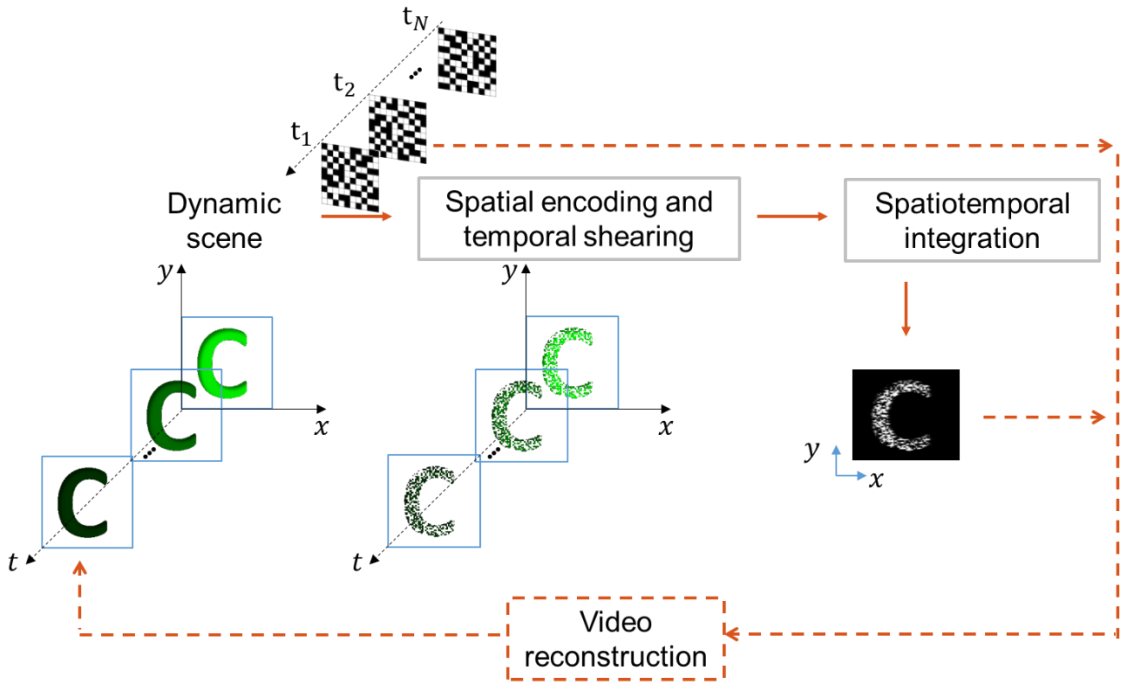


Figure 2.8 Working principle of coded aperture compressive temporal imaging.

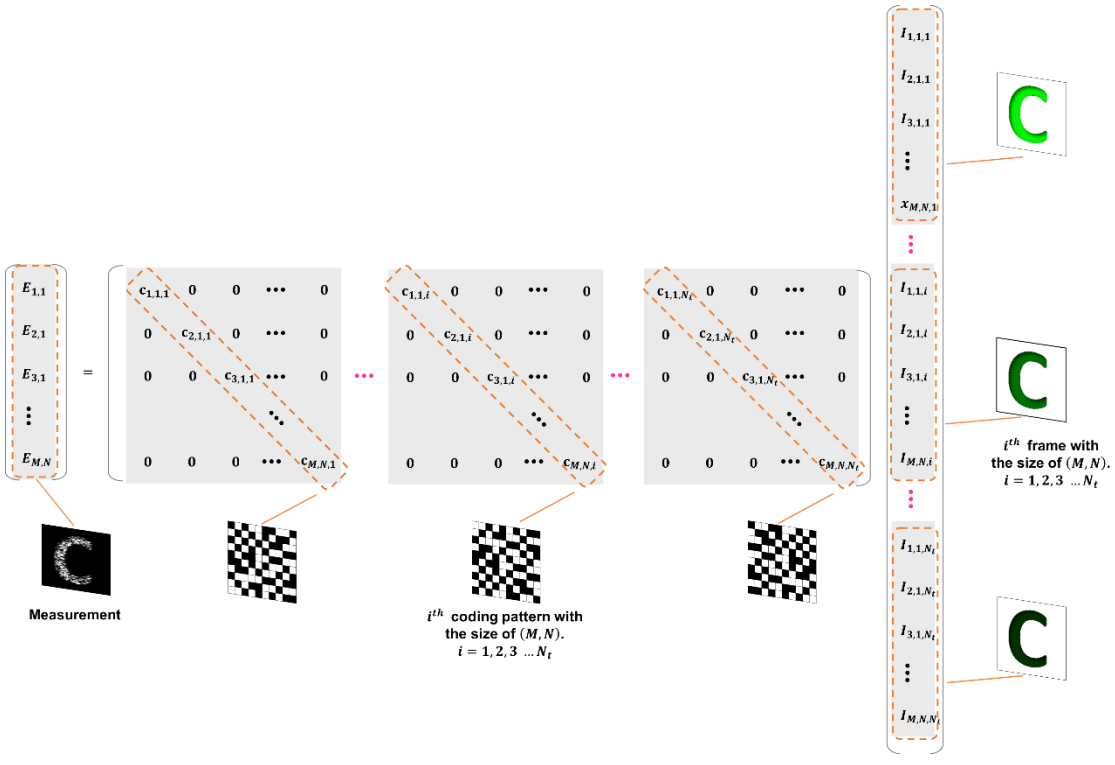


Figure 2.9 Schematic of the matrix-vector product form of compressed temporal imaging.

2.3.3 Investigation of temporal shearing

The concept of temporal shearing is of significance in performing CTI. For instance, in CUP, a static encoding pattern was temporally sheared by a streak tube to generate multiple encoding patterns, which is the key to achieving an ultrafast encoding scheme. Inspired by the tomography concept, the temporal shearing could be viewed as a kind of projection operator. It links the time domain to another domain, such as space, spectrum, and angle, to name a few. For example, the common temporal shearing is the time-to-space projection, which could be mainly grouped into three categories, including single-shearing, dual-shearing, and diagonal-shearing methods. Among them, CUP is the typical single-shearing model as shown in Fig. 2.6. Figure 2.10 shows the working principle of dual-shearing time-to-space conversion. The dynamic scene is temporal shearing first and then is encoded by a static encoding pattern, in which each frame of the dynamic scene goes through a different area on the encoding pattern. Then, inverse temporal shearing is applied to cancel the first temporal shearing effect. After the spatiotemporal integration, all modulated frames are integrated into one compressed image. Compared to the single-shearing time-to-space conversion, the dual-shearing method makes full use of the pixel count of the detector in the shearing direction. Furthermore, each pixel in the measurement is the linear combination of the same-position pixel value in all frames, which maximally preserves the spatial information in the dynamic scene without the effect of the shearing operator and cross-talk. Figure 2.11 shows a diagonal shearing time-to-space conversion. Not only were the encoded frames sheared along the horizontal direction but were sheared along the vertical direction simultaneously. The way of diagonal shearing makes the temporal information mixed with the horizontal and vertical information, which could supply more prior for improving the reconstruction accuracy in time.

In the meantime, temporal shearing provides a way to improve image reconstruction quality. For instance, multiple-view CUP [20] as shown in Fig. 2.12 includes three views, including two time-shearing views (generated by the states of ‘OFF’ and ‘ON’ of a DMD simultaneously) and one time-unshearing view (captured by an external camera). The time-unshearing view was used to preserve the spatial information in the dynamic scene. Meanwhile, both time-shearing views retain temporal information by electronic streaking via time-to-space conversion. Altogether, three views maximally keep rich spatiotemporal information.

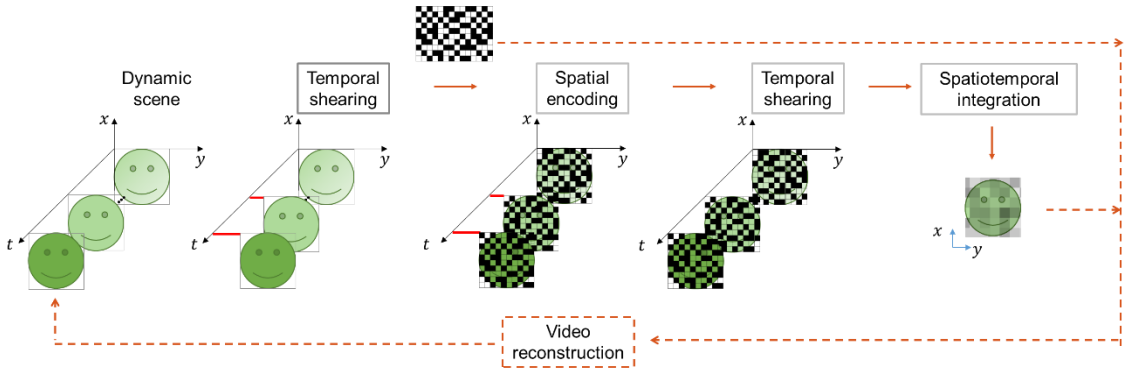


Figure 2.10 Working principle of dual-shearing time-to-space conversion. The red line denotes the shifting distance along the horizontal direction.

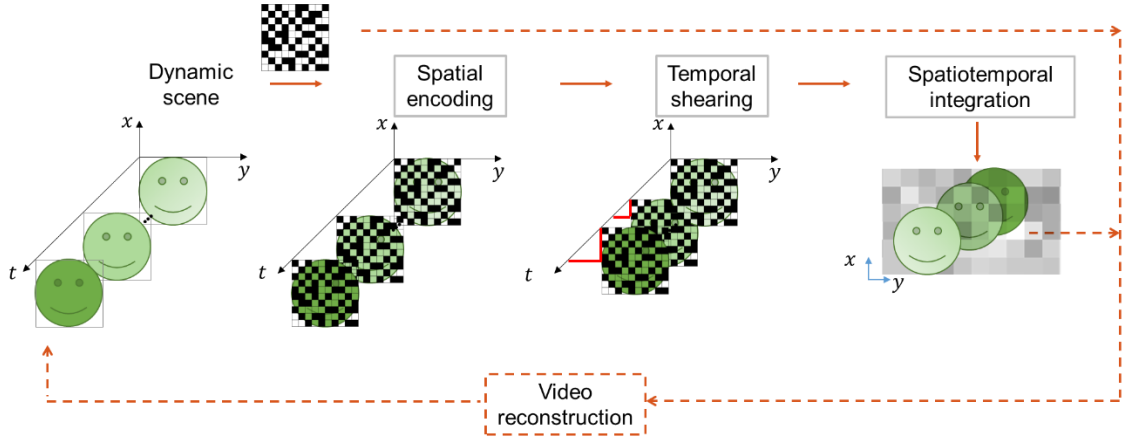


Figure 2.11 Working principle of diagonal shearing time-to-space conversion. The red line denotes the shifting distance along with the horizontal and vertical directions.

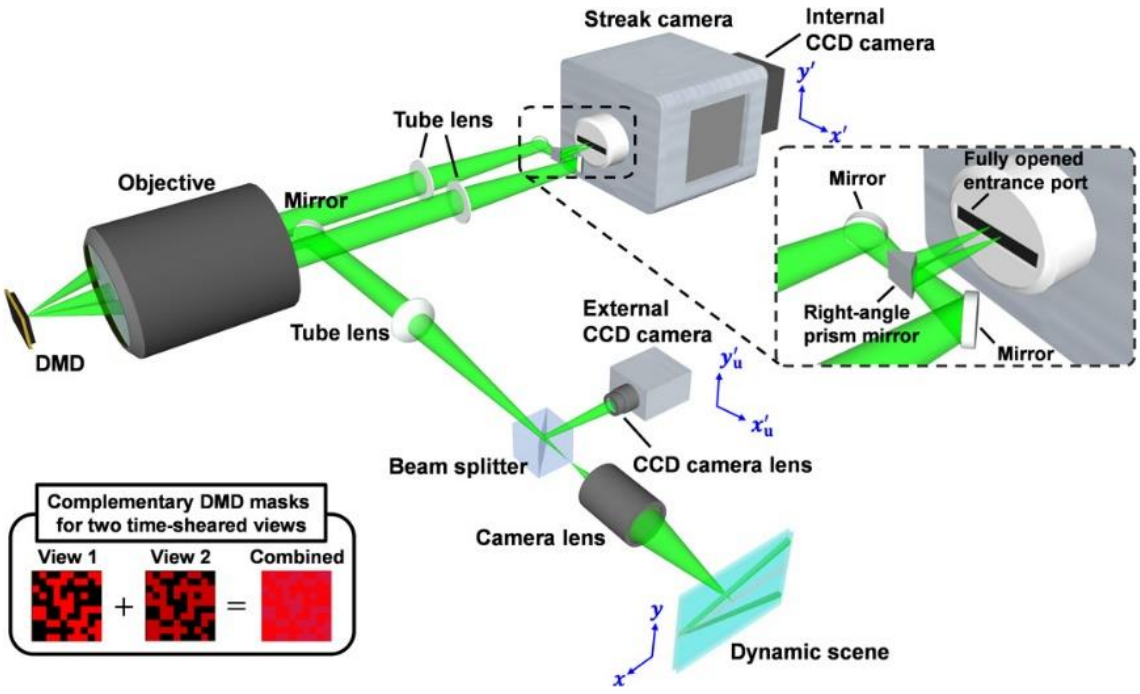


Figure 2.12 Schematic of multiple view CUP with the lossless-encoding scheme [20].

2.4 Reconstruction algorithms

The reconstruction algorithms of CTI could be grouped into two categories. The first one is iterative algorithms with preset priors such as total variation (TV) [66] or deep denoising priors [67]. The typical iterative algorithms include two steps iteration shrink/thresholding (TwIST) [68] and plug-and-play (PnP) alternative direction multiplier methods (ADMM) [69]. The second one is end-to-end learning-based algorithms, in which a measurement (input) directly generates a video (output) through a pre-trained network [70].

2.4.1 TwIST

TwIST is applied to solve an inverse problem (i.e., given E and Φ to find a solution to $E = \Phi I$) by a minimizer of a convex objective function:

$$f(I) = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\Phi I - E\|^2 + \lambda \Phi(I) \right\}. \tag{2.6}$$

Here, I denotes the video to be reconstructed. Φ represents the sensing matrix. $\Phi(\cdot)$ is the regularization function. λ is the known regularization constant. For the j^{th} iteration, the estimation of I becomes

$$I_1 = \mathcal{T}_\lambda(I_0), \tag{2.7}$$

$$I_{j+1} = (1-\alpha)I_{j-1} + (\alpha-\beta)I_j + \beta\mathcal{T}_\lambda(I_j). \tag{2.8}$$

Here, I_0 is the vectorized signal. α and β are the pre-set constants for different reconstruction tasks, which affect the convergence rate of the minimization problem of Eq. (2.6). The designation "two-step" stems from the fact that it depends on both I_{j-1} and I_j, rather than only on I_j. For $j \geq 1$, a function mapping operation $\mathcal{T}_\lambda$: $\mathrm{R}^m \longrightarrow \mathrm{R}^m$ is defined as

$$\mathcal{T}_\lambda(I) = \Psi_\lambda[I + \Phi^T(E - \Phi I)]. \tag{2.9}$$

Here, Φ^T denotes the transpose of Φ. Ψ_λ denotes a denoising operator, whose choice is related to the regularization function $\Phi(\cdot)$. With TV as the regularization function, the denoising operator leverages the Chambolle algorithm [71].

The relative change of the estimated output from the objective function [i.e., Eq. (2.6)] is used as the merit function. The iteration process stops when this change is less than the pre-set tolerance value ϑ (e.g., 0.01), i.e.,

$$\frac{\left|f(I_j) - f(I_{j-1})\right|}{f(I_j)} \leq \vartheta. \tag{2.10}$$

2.4.2 PnP-ADMM

ADMM is an advanced tool for minimizing the sum of multiple separable functions. For simplicity, we use the two-function model as an example. The algorithm works by converting the unconstrained optimization [i.e., Eq. (2.6)] into a constrained problem by introducing a variable v:

$$(\hat{I}, \hat{v}) = \underset{I,v}{\operatorname{argmin}} \left\{\frac{1}{2}\|\Phi I - E\|^2 + \lambda\Phi(I)\right\}, \text{subject to } I = v. \tag{2.11}$$

It considers the augmented Lagrangian function by introducing a Lagrange multiplier u and penalty parameter ρ, so that Eq. (2.11) becomes

$$(\hat{I}, \hat{v}, \hat{u}) = \underset{I,v,u}{\operatorname{argmin}}\{\mathcal{L}(I, v, u)\}. \tag{2.12}$$

where $\mathcal{L}(I, v, u) = \frac{1}{2}\|\Phi I - E\|^2 + \lambda\Phi(v) + u^T(I - v) + \frac{\rho}{2}\|I - v\|^2$. Then, the algorithm finds the solution by seeking a saddle point of $\mathcal{L}$, which involves solving a sequence of sub-problems in the form

$$I^{(j+1)} = \underset{I}{\operatorname{argmin}}\left\{\frac{1}{2}\|\Phi I - E\|^2 + \frac{\rho}{2}\left\|I - \tilde{I}^{(j)}\right\|^2\right\}, \tag{2.13}$$

$$v^{(j+1)} = \underset{v}{\operatorname{argmin}}\left\{\lambda\Phi(v) + \frac{\rho}{2}\left\|v - \tilde{v}^{(j)}\right\|^2\right\}, \text{and} \tag{2.14}$$

$$\bar{u}^{(j+1)} = \bar{u}^{(j)} + \left(I^{(j+1)} - v^{(j+1)}\right). \tag{2.15}$$

Here, $\bar{u}^{(j)} \stackrel{\text{def}}{=} {u^{(j)}}/{\rho}$ is the scaled Lagrange multiplier. $\tilde{I}^{(j)} \stackrel{\text{def}}{=} v^{(j)} - \bar{u}^{(j)}$ and $\tilde{v}^{(j)} \stackrel{\text{def}}{=} I^{(j+1)} + \bar{u}^{(j)}$ are the intermediate variables [72]. Under the mild conditions, one can show that the iterates returned by Eqs. (2.13)–(2.15) converge to the solution of Eq. (2.11).

The idea of plug-and-play(PnP)-ADMM is to modify Eq. (2.14) by observing that it is a denoising step if we treat $\tilde{v}^{(j)}$ as a "noisy" version of v and $\Phi(v)$ as a regularization for v. Based on this observation, we can replace Eq. (2.14) with a denoiser $\mathcal{D}_\sigma$: $\mathrm{R}^m \longrightarrow \mathrm{R}^m$ such that

$$v^{(j+1)} = \mathcal{D}_\sigma\left(\tilde{v}^{(j)}\right), \tag{2.16}$$

where $\sigma = \sqrt{\lambda/\rho}$ is the denoising strength. The choice of $\mathcal{D}_\sigma$ is broad, including TV denoising, deep convolution neural network, and block-matching and 3D filtering (BM3D) [73] to name a few. PnP-ADMM uses the relative change of the estimate in adjacent iterations as the merit function:

$$\text{if } \frac{\left\|I^{(j+1)} - I^{(j)}\right\|_2}{\left\|I^{(j+1)}\right\|_2} < \varpi \text{ and } \rho^{j+1} = \rho^j \ . \tag{2.17}$$

Here, ϖ ($0 < \varpi < 10^{-3}$) is the pre-set tolerance value.

2.4.3 Learning-based CTI

The long-running time of the iteration algorithm precludes wide applications of CTI, especially in some cases, a real-time visualization is desired. Benefitting from recent advances in deep learning, real-time end-to-end reconstruction has been demonstrated in CTI [74]. The deep learning approach first learns, Φ^{-1}, an approximate inverse function of the system forward model in training and then provides instantaneous reconstruction by directly estimating outputs from the input measurements without an iteration process. Mathematically, it could be expressed as:

$$I = \Phi^{-1}E. \tag{2.18}$$

For obtaining the accurate Φ^{-1}, a large number of training data is required for the supervised learning methods, in which labeled datasets are used to train algorithms for classifying data or predicting outcomes accurately. Furthermore, another testing data that have never been seen by

the network is used to test the robustness and inference ability of the trained network. The typical network structures include deep fully connected networks [70], U-net [75] with convolution neural networks, and autoencoder with generative adversarial networks [76]. To overcome the overfitting issue during the training procedure, different merit functions, like pixel-wise mean squared error and the multiscale structural similarity index, are employed.

2.4.4 Comparison between TwIST, PnP-ADMM, and learning-based methods

The PnP-ADMM framework [i.e., Eqs. (2.11)–(2.17)] has a better decomposability than that of the TwIST algorithm [i.e., Eqs. (2.6)–(2.10)] in handling complex and multiple-featured global optimization problems. Using a decomposition-coordination strategy, PnP-ADMM divides the large global optimization problem into small and easier-to-handle sub-problems, whose solutions are coordinated to help pinpoint the global minimization [72]. For instance, the inverse problem Eq. (2.6) is separated into three sub-problems as Eqs. (2.13)–(2.15) in PnP-ADMM. Among them, Eq. (2.14) is cast as a denoising step to leverage advanced denoising functions, which leads to a better image reconstruction quality.

Despite the learning-based methods having the reconstruction speed advantage, it usually requires a deep model, long training time, and a large amount of training data. Furthermore, it is less flexible than the iteration-based algorithms because the model is trained and then works on the system with determined hyper-parameters such as image size, compression ratio, and coding patterns.

In terms of reconstruction speed and image quality, learning-based methods are the best, but they are also the least flexible than the iterative methods, meanwhile, the PnP-ADMM outperforms TwIST as shown in Fig.2.13. Considering overall performance on quality, flexibility, ease of use, cost, and speed, the PnP-ADMM is a good baseline for CTI reconstruction.

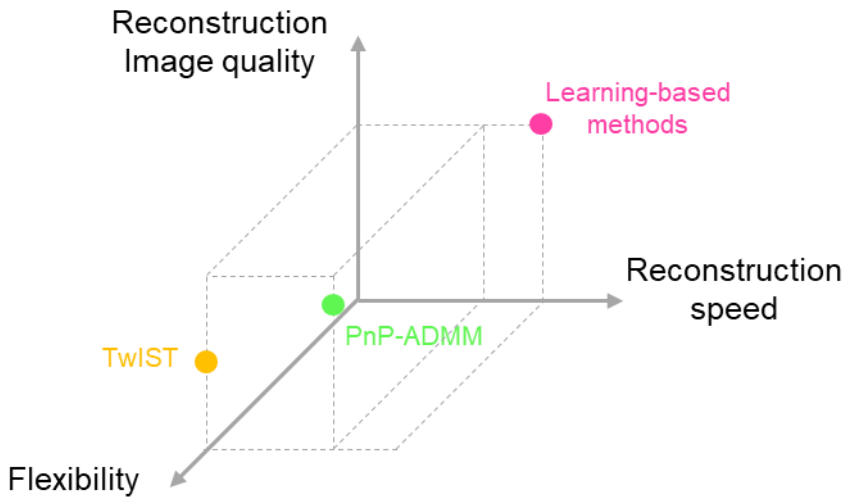


Figure 2.13 Comparison of three types of reconstruction methods in terms of reconstruction image quality, reconstruction speed, and flexibility.

3 CS-BASED ULTRAHIGH-SPEED OPTICAL PHOTOGRAPHY

Single-shot Compressed Optical-Streaking Ultra-high-speed Photography

Photographie à ultra-haute vitesse avec stries optiques compressées en une seule photo

Authors:

Xianglei Liu, Jingdan Liu, Cheng Jiang, Fiorenzo Vetrone, and Jinyang Liang*

Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650 boulevard Lionel-Boulet, Varennes, Québec J3X1S2, CANADA

Publication:

Optics Letters

Vol. 44, Issue 6, pp. 1387-1390 (2019)

https://doi.org/10.1364/OL.44.001387

Contribution of authors: Xianglei Liu built the system, conducted the experiments, and analyzed the data. Jingdan Liu and Cheng Jiang prepared the sample and performed some of the experiments. Prof. Fiorenzo Vetrone initiated the project. Prof. Jinyang Liang initiated the project, proposed the concept, contributed to experimental design, and supervised the project. All authors wrote and revised the manuscript.

Abstract

Single-shot ultra-high-speed imaging is of great significance to capture transient phenomena in physics, biology, and chemistry in real time. Existing techniques, however, have a restricted application scope, a low sequence depth, or a limited pixel count. To overcome these limitations, we developed single-shot compressed optical-streaking ultra-high-speed photography (COSUP) with an imaging speed of 1.5 million frames per second, a sequence depth of 500 frames, and an (x, y) pixel count of 0.5 megapixels per frame. COSUP's single-shot ultra-high-speed imaging ability was demonstrated by recording single laser pulses illuminating through transmissive targets and by tracing a fast-moving object. As a universal imaging platform, COSUP is capable of increasing imaging speeds of a wide range of CCD and CMOS cameras by four orders of magnitude. We envision COSUP to be applied in widespread applications in biomedicine and materials science.

OCIS codes: (110.1758) Computational imaging; (170.6920) Time-resolved imaging; (100.3010) Image reconstruction techniques

3.1 Introduction

Single-shot ultra-high-speed [i.e., 0.1 to 10 million frames per second (Mfps)] imaging [77] is indispensable for visualizing various instantaneous phenomena occurring in two-dimensional (2D) space, such as phosphorescence light emission [78], neural activities [79], and conformational changes in proteins [80]. Existing single ultra-high-speed imaging techniques [14, 18, 81-87] can be generally categorized into active-detection and passive-detection domains. The active-detection approaches exploit specially designed pulse trains to probe 2D transient events [i.e. (x, y) frames that vary in time]. The representative modalities include frequency-dividing imaging [81] and time-stretching imaging [82-85]. However, these methods are not suitable for imaging self-luminescent and color-selective events. By contrast, the passive-detection approaches leverage receive-only ultra-high-speed detectors to record photons scattered and emitted from the transient scenes. Examples include rotatory-mirror-based cameras [18], beam-splitting-based framing cameras [86], the in-situ storage image sensor CCD camera [14], and the global shutter stacked CMOS camera [87]. Nevertheless, these cameras either have a bulky and complicated structure or have a limited sequence depth (i.e., the number of frames in one acquisition) and pixel count.

To circumvent these drawbacks, computational imaging techniques [88], combining physical data acquisition and numerical image reconstruction, were increasingly featured in recent years. In particular, the implementation of compressed sensing (CS) [89] for spatial and/or temporal multiplexing has allowed overcoming the speed limit with a substantial improvement in the sequence depth and pixel count [90, 91]. The representative techniques in this category include programmable pixel compressive camera (P2C2) [92, 93], coded aperture compressive temporal imaging (CACTI) [63, 94], and the multiple-aperture (MA)-CS CMOS camera [95]. However, despite reaching over one megapixel per frame, the imaging speeds of P2C2 and CACTI, inherently limited by the refreshing rate of a spatial light modulation and the moving speed of a piezoelectric stage, are clamped at several thousand fps (kfps). For MA-CS CMOS, despite its ultra-high-speed imaging speeds, it has a limited pixel count of 64×108 with a sequence depth of 32. Thus, existing computational imaging modalities fall short to simultaneously possess satisfying specifications in the frame rates, sequence depth, and pixel count for ultra-high-speed imaging.

To overcome these limitations, in this Letter, we propose single-shot compressed optical-streaking ultra-high-speed photography (COSUP), which is a passive-detection computational

imaging modality with a 2D imaging speed of 1.5 Mfps, a sequence depth of 500, and an (x, y) pixel count of 1000 × 500 per frame.

3.2 Results

3.2.1 Operating principle of COSUP

The schematic of the COSUP system is shown in Fig. 3.1. A transient scene is first imaged onto a digital micromirror device (DMD, AJD-4500, Ajile Light Industries), on which a binary pseudo-random pattern (with an encoding pixel size of 32.4 × 32.4 μm^2) is loaded to conduct spatial encoding. Subsequently, the spatially encoded frames are relayed onto a CMOS camera (GS3-U3-23S6M-C, FLIR) by a 4*f* system. A galvanometer scanner (GS, 6220H, Cambridge Technology), placed at the Fourier plane of this 4*f* system, temporally shears the spatially encoded frames linearly to different spatial locations along the x axis of the CMOS camera according to their time of arrival. The synchronization between the GS' rotation and the camera's exposure is controlled by the sinusoidal signal and the rectangular signal from a function generator (DG1022, Rigol Technologies), as depicted in the inset of Fig. 3.1. Finally, via spatiotemporal integration, the CMOS camera compressively records the spatially encoded and temporally sheared scene as a 2D streak image with a single exposure. It is noted that our work is inspired by recent advances in compressed ultrafast photography [96-98]. However, instead of using a streak camera, we implement the GS [99, 100] for temporal shearing and use an off-the-shelf CMOS camera for detection. This design thus avoids drawbacks—such as the space-charge effect and the low quantum efficiency of the photocathode—that are presented in the streak camera. In the following, we shall demonstrate that this all-optical approach can increase the imaging speed of the CMOS camera by four orders of magnitude to the Mfps level for recording single laser pulses illuminating through transmissive targets and for tracking a fast-moving object.

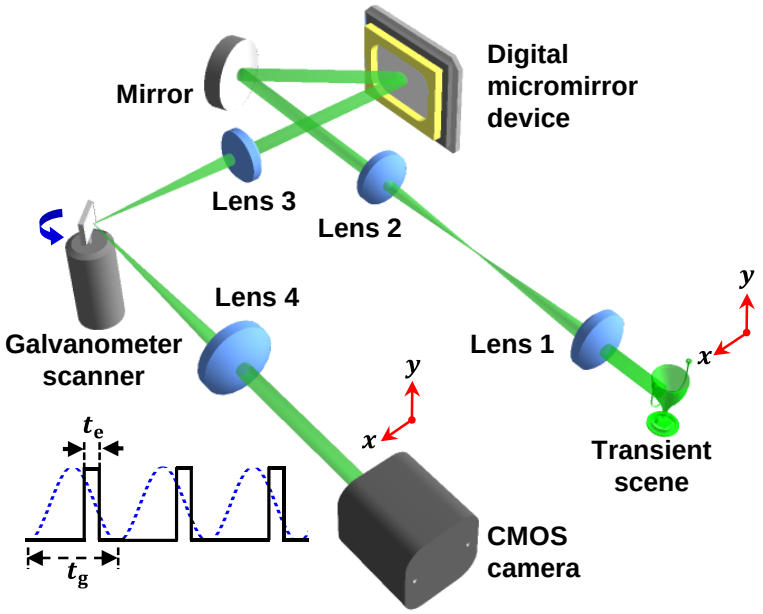


Figure 3.1 Schematic of the COSUP system. Inset: Synchronization between the CMOS camera's exposure (black solid line) with an exposure time of $\boldsymbol{t_e}$ and the galvonometer scanner's sinusoidal control signal (blue dashed line) with a period of $\boldsymbol{t_g}$. Lenses 1 and 4.

The operation of the COSUP system can be described by the following model

$$E = \mathbf{T}\mathbf{S_o}\mathbf{C}\, I(x, y, t), \tag{3.1}$$

where $I(x, y, t)$ is the light intensity of the transient event, **C** represents spatial encoding by the DMD, $\mathbf{S_o}$ represents linearly temporal shearing by the GS (the subscript "o" stands for "optical"), and **T** represents spatiotemporal integration by the CMOS camera. Because of the prior knowledge of the operators and the spatiotemporal sparsity of the transient scene, $I(x, y, t)$ can be recovered from the measurement *E* by solving the inverse problem of

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \| E - \mathbf{T}\mathbf{S_o}\mathbf{C} I \|_2^2 + \lambda \Phi_{\mathrm{TV}}(I) \right\} \tag{3.2}$$

Here $\|\cdot\|_2^2$ represents the l_2 norm, λ is a weighting coefficient, and $\Phi_{\mathrm{TV}}(\cdot)$ is the total variation (TV) regularizer. In practice, $I(x, y, t)$ was recovered by using a CS-based algorithm that was developed upon the two-step iterative shrinkage/thresholding algorithm [101].

To obtain a linear temporal shearing operation, we synchronized the camera's exposure window with the GS' linear rotation region. In particular, a static target was placed at the object plane. By tuning the initial phase of the sinusoidal function, the camera's exposure window was slid to search for the peak or valley of the sinusoidal signal. The search was completed when local features of the static target were precisely matched in the streak image due to the symmetric

back and forth scanning. Finally, 90° was added to the initial phase to find the linear slope region of the sinusoidal function.

The reconstructed movie has a frame rate of

$$r = \frac{\alpha U f_4}{t_g d}. \tag{3.3}$$

Here α=0.07 rad/V is a constant that links the voltage added onto the GS, denoted as *U*, with the deflection angle in its linear rotation range. f_4=75 mm is the focal length of Lens 4, t_g is the period of the sinusoidal voltage waveform added to the GS, and d = 5.86 μm is the CMOS sensor's pixel size. In addition, the preset exposure time of the CMOS camera, t_e, determines the total length of the recording time window. If the entire streak is located within the sensor, the sequence depth can be calculated by $N_t = rt_\mathrm{e}$. The number of pixels in the x axis of each frame, N_x, can be calculated by $N_x \leq N_\mathrm{c} + 1 - N_t$, where $N_\mathrm{c}$denotes the number of pixels in each column of the CMOS sensor. The number of pixels in the y axis of each frame, N_y, is less than or equal to that in each row of the CMOS sensor, N_r [i.e., $N_y \leq N_\mathrm{r}$].

3.2.2 Quantification of system's spatial resolution

To characterize COSUP's spatial frequency responses, we imaged single laser pulses illuminating through a resolution target [Fig. 3.2(a)]. In particular, a 532-nm continuous wave laser was controlled by an external trigger to generate laser pulses with different temporal widths. Five different pulse widths (100, 300, 500, 700, and 900 μs) were used to provide decreased sparsity from 90% to 10% with a step of 20% in the temporal axis for a recording time window of 1 ms. COSUP captured these dynamic scenes at 60 kfps. The illuminated bars (i.e., Elements 4 to 6 in Group 2 and Elements 1 to 6 in Group 3) are shown as the first panel in Fig. 3.2(b). We reconstructed movies for each pulse width and projected these datacubes onto the x-y plane, as shown as the rest of the panels in Fig. 3.2(b). These results reveal that the spatial resolution of COSUP depends on the sparsity of the transient scene. The contrast in the reconstructed image quality degrades with the increased laser pulse widths. Moreover, longer pulse widths produce lower reconstructed intensity. To quantify the system's performance by considering both effects, we used the normalized product of the contrast and the reconstructed intensity as the merit function [Fig. 3.2(c)]. For the 900-μs pulse illumination, Element 3 in Group 3 in the reconstruction has a normalized product below 0.25, which was used as the threshold to determine the resolvable feature. Thus, the COSUP's spatial resolution was quantified to be 50 μm.

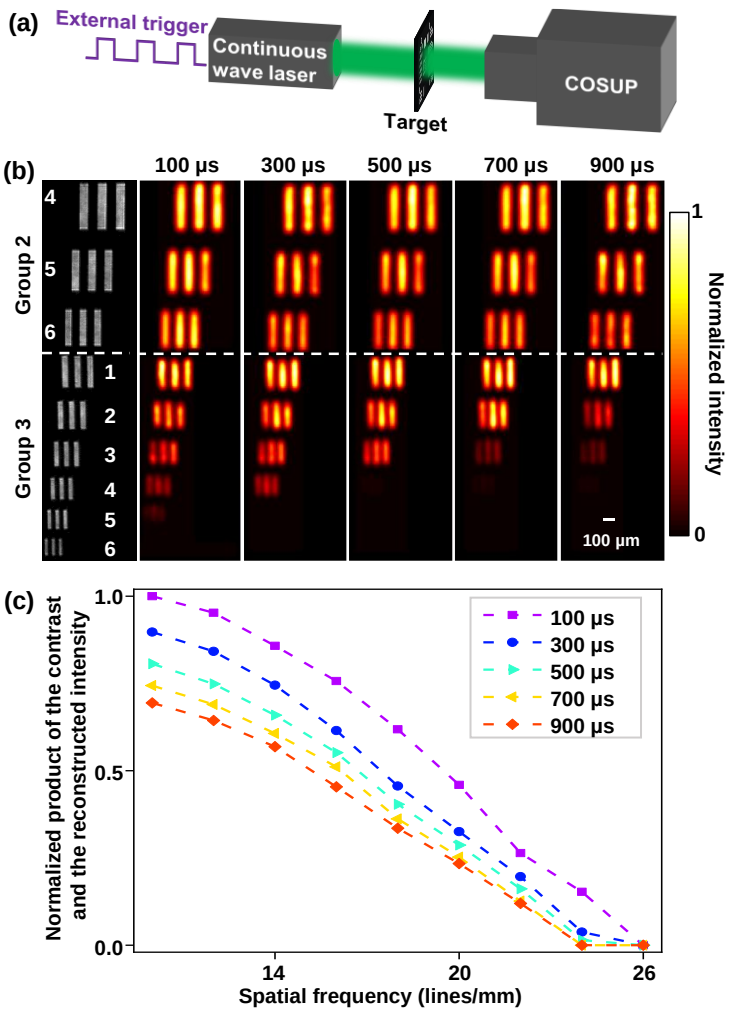


Figure 3.2 Quantifying COSUP's spatial frequency responses. (a) Experimental setup. (b) Illuminated bars on the resolution target. In the first panel, the numbers represent Elements 4 to 6 in Group 2 and Elements 1 to 6 in Group 3. The rest of the panels show the projected images of illuminated bars for different laser pulse widths, calculated by summing over voxels the reconstructed (x, y, t) datacubes along the t axis. (c) Comparison of COSUP's spatial frequency responses with different laser pulse widths.

3.2.3 Quantification of system's temporal resolution

To demonstrate COSUP's multi-scale ultra-high-speed imaging capability, we captured the transmission of single laser pulses through a mask. An incident laser pulse was divided by a beam splitter into two components. The reflected component was recorded by a photodiode, and the transmitted component illuminated a transmissive USAF mask that modulated the laser pulses' spatial profiles. For the first experiment, we generated a pulse train that contained four 300-µs pulses. COSUP's imaging speed was set to 60 kfps. While the CMOS camera, at its intrinsic

imaging speed of 20 fps, could only provide a single image [red solid box in Fig. 3.3(a)] without any temporal information, COSUP recorded the mask's spatial profile and laser pulse's intensity time course in a movie with 240 frames. A representative frame (t = 433 μs) is shown in Fig. 3.3(a). Figure 3.3(b) depicts the normalized intensity of a selected cross section [cyan dashed line in Figs. 3.3(a) and (d)], which demonstrates the well reconstructed spatial features to the ground truth. Moreover, we calculated the average intensity in each frame. The time course shows a good agreement with the photodiode-recorded result [Fig. 3.3(c)]. We then increased the imaging speed to 1.5 Mfps to record a single 10-μs laser pulse. A representative frame (t = 33 μs) is shown in Fig. 3.3(d). The comparison of the time courses of averaged intensity [Fig. 3.3(e)] confirmed consistency between the COSUP and photodiode results under this imaging speed.

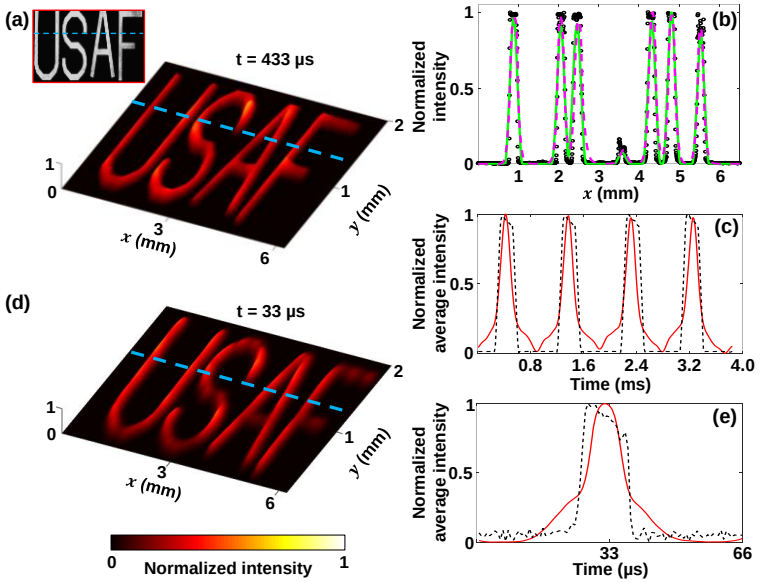


Figure 3.3 Recording transmission of single laser pulses through a mask using COSUP. (a) A represented reconstructed frame showing a 300-μs laser pulse passing through a transmissive USAF pattern. The imaging speed was 60 kfps. Inset: the time-integrated image captured by the CMOS camera with its intrinsic imaging speed (20 fps). (b) Normalized intensity of a selected cross section [cyan dash lines in (a) and (d)] in the ground truth (black circle) and in the representative reconstructed frames using 300-μs (green solid line) and 10-μs (magenta dashed line) laser pulses. (c) Comparison of the measured normalized average intensity of the laser pulse as a function of time using the COSUP system (red solid line) and a photodiode (black dashed line) for the 300-μs laser pulse. (d) As (a), but using a 10-μs laser pulse with a 1.5-Mfps imaging speed. (e) As (c), but for a 10-μs laser pulse.

3.2.4 Application of COSUP to fast-moving objects tracking

To demonstrate COSUP's ability to track fast-moving objects, we imaged an animation of a fast-moving ball [Fig. 3.4(a)]. This animation comprised 40 patterns, which were loaded and played by another DMD (D4100, Digital Light Innovations) at 20 kHz. We shone a collimated laser beam onto the DMD at ~24° from its surface normal. The COSUP system perpendicularly faced the

DMD's surface and collected the light diffracted by the patterns at 140 kfps. Figure 3.4(b) shows a time-integrated image of this dynamic event acquired by the CMOS camera at its intrinsic frame rate of 20 fps. Figure 3.4(c) shows a color-encoded image generated by superimposing 10 representative time-lapse frames (with an interval of 215 μs) of the moving ball from the movie reconstructed by the COSUP system. While the time-integrated image merely presents the overall trace, the time-lapse frames unambiguously show the evolution of the spatial position and the shape (especially the deformation from the round to the elliptical shape at the turning points of its trajectory) at each time point. To evaluate the reconstruction's accuracy, we traced the centroids of the bouncing ball in each reconstructed frame [Fig. 3.4(d)]. The measurement errors were calculated by subtracting the measured position of centroids from the preset ones. Further, the root-mean-square errors (RMSEs) of reconstructed centroids along the x and y axes were calculated to be 22 μm and 9 μm, respectively. These analyses confirm that the COSUP system has good measurement precision with respect to the ground truth. The anisotropy of RMSEs was attributed to the spatiotemporal mixing along the shearing direction.

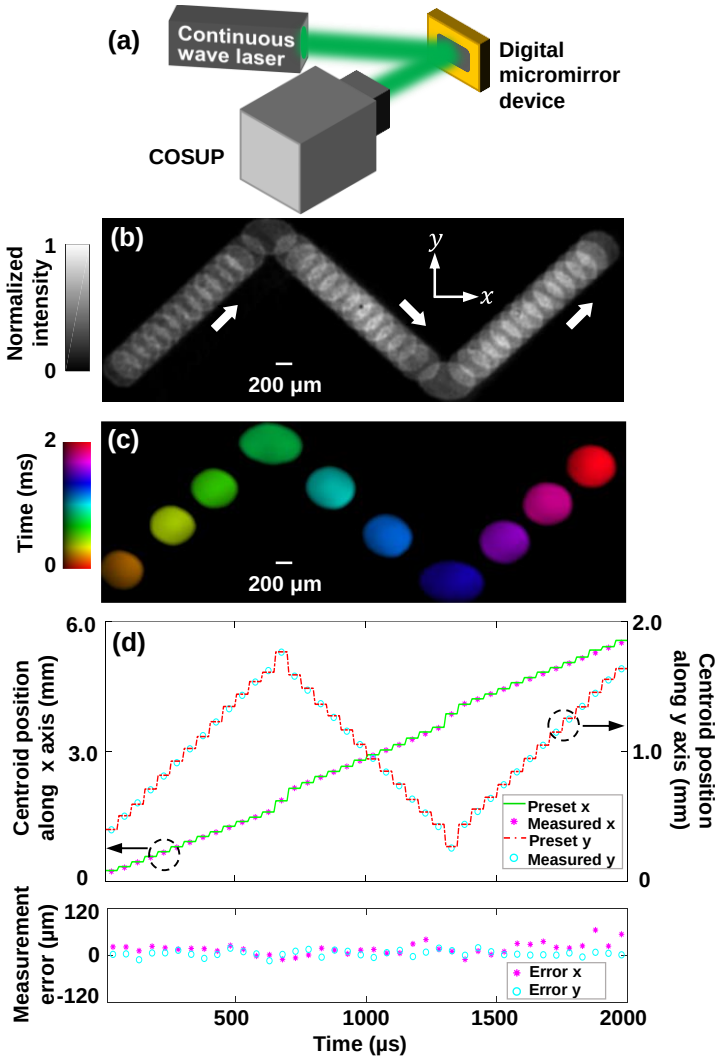


Figure 3.4 Tracing a fast-moving object using COSUP. (a) Experimental setup. (b) Time-integrated image of the fast-moving ball patterns, imaged at the intrinsic frame rate of the CMOS camera (20 fps). (c) Superimposed image of 10 representative time-lapse frames (with an interval of 215 μs) of the same dynamic scene in (b), imaged by using the COSUP system. (d) Comparison of the centroid positions along the x and y axes between the measurement results and the ground truths. To avoid cluttering, only one data point is shown for every seven measured data points.

3.3 Conclusions

In conclusion, we have demonstrated single-shot 2D ultra-high-speed passive imaging using COSUP. Featuring optical streaking using a GS in the 4f imaging system, COSUP endows an off-the-shelf CMOS camera with tunable imaging speeds of up to 1.5 Mfps, which is approximately three orders of magnitude higher than the state-of-art in imaging speed of CS-based temporal imaging using silicon sensors [19-21]. In addition, the system is capable of reaching a sequence depth of up to 500 frames and a pixel count of 0.5 megapixels in each frame. COSUP's ultra-

high-speed imaging capability was demonstrated by capturing the transmission of single laser pulses through a mask and by tracing the shape and position of a fast-moving object in real time.

As a universal imaging platform, COSUP can achieve a scalable spatial resolution by coupling with different front optics in microscopes and telescopes. Moreover, although not demonstrated in this work, COSUP can be easily applied to other CCD or CMOS cameras according to specific studies. For instance, integration of an electron-multiplying CCD camera in the COSUP system will enable high-sensitivity optical neuroimaging of action potential propagating at tens of meters per second [102] under microscopic settings [103]. As another example, an infrared-camera-based COSUP system could enable wide-field temperature sensing in deep tissue using nanoparticles [104]. In summary, by leveraging the advantages of off-the-shelf cameras and sensors, COSUP is expected to find widespread applications in both fundamental and applied sciences.

Funding. Discovery (RGPIN-2017-05959) and Discovery Accelerator Supplement (RGPAS-507845-2017) programs from Natural Sciences and Engineering Research Council (NSERC) of Canada; John R. Evans Leader's Fund from Canada Foundation of Innovation (37146); Établissement de nouveaux chercheurs universitaires (2019-NC-252960) from Fonds de recherche du Québec–Nature et technologies (FRQNT).

4 CS-BASED OPTICAL THERMOMETRY

Fast Wide-field Upconversion Luminescence Lifetime Thermometry Enabled by Single-shot Compressed Ultrahigh-speed Imaging

Thermométrie de durée de vie de luminescence à conversion ascendante rapide à grand champ rendue possible par une imagerie ultra-rapide compressée en une seule prise

Authors:

Xianglei Liu[1,3], Artiom Skripka[1,2,3], Yingming Lai[1], Cheng Jiang[1], Jingdan Liu[1], Fiorenzo Vetrone[1,*], Jinyang Liang[1,*]

[1]Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650 boulevard Lionel-Boulet, Varennes, Québec J3X1S2, CANADA

[2]Present address: Nanomaterials for Bioimaging Group, Departamento de Física de Materiales, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, 28049, SPAIN and The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[3]These authors contributed equally to this work

*Corresponding authors: Fiorenzo.Vetrone@inrs.ca (F.V.) and Jinyang.Liang@inrs.ca (J.Liang)

Publication:

Nature Communications

Vol. 12, pp. 6401 (2021)

https://doi.org/10.1038/s41467-021-26701-1

Contribution of authors: X.L. designed and built the system, conducted the experiments, developed the reconstruction algorithm, and analyzed the data. A.S. prepared the UCNPs, conducted some experiments, and analyzed the data. Y.L. contributed to the algorithm development. C.J. and J.Liu conducted some experiments. F.V. and J.Liang initiated the project. J.Liang proposed the concept, contributed to experimental design, and supervised the project. All authors wrote and revised the manuscript.

Abstract

Photoluminescence lifetime imaging of upconverting nanoparticles is increasingly featured in recent progress in optical thermometry. Despite remarkable advances in photoluminescent temperature indicators, existing optical instruments lack the ability of wide-field photoluminescence lifetime imaging in real time, thus falling short in dynamic temperature mapping. Here, we report video-rate upconversion temperature sensing in wide field using single-shot photoluminescence lifetime imaging thermometry (SPLIT). Developed from a compressed-sensing ultrahigh-speed imaging paradigm, SPLIT first records wide-field luminescence intensity decay compressively in two views in a single exposure. Then, an algorithm, built upon the plug-and-play alternating direction method of multipliers, is used to reconstruct the video, from which the extracted lifetime distribution is converted to a temperature map. Using the core/shell $NaGdF_4:Er^{3+},Yb^{3+}/NaGdF_4$ upconverting nanoparticles as the lifetime-based temperature indicators, we apply SPLIT in longitudinal wide-field temperature monitoring beneath a thin scattering medium. SPLIT also enables video-rate temperature mapping of a moving biological sample at single-cell resolution.

4.1 Introduction

Temperature is an important parameter associated with many physical, chemical, and biological processes [105]. Accurate and real-time (i.e., the actual time during which the event occurs) temperature sensing at microscopic scales is essential to both industrial applications and scientific research, including the examination of internal strains in turbine blades [106], control of the synthesis of ionic liquids [107], and theranostics of cancer [108]. In the past decade, photoluminescence lifetime imaging (PLI) has emerged as a promising approach to temperature sensing [109]. Because photoluminescence can be both excited and detected optically, the resulting non-contact PLI possesses a high spatial resolution [104, 110, 111]. This advantage not only overcomes the intrinsic limitation in spatial resolution of imaging thermography due to the long wavelengths of thermal radiation but also avoids heat-transfer-induced inaccuracy in conventional contact methods [112]. Moreover, independent of prior knowledge of samples' physical properties (e.g., emissivity and Grüneisen coefficient [113, 114]), PLI brings in higher flexibility in sample selection. Furthermore, PLI is less susceptible than the intensity-based measurements to inhomogeneous signal attenuation, stray light, photobleaching, light's path length, and excitation intensity variations [115-119]. Finally, PLI does not rely on the concentration of labeling agents [110], which eliminates the need for special ratiometric probes [120]. Overcoming many challenges in previous methods, PLI is becoming a popular choice for optical thermometry [121-125].

The success of PLI in temperature mapping depends on two essential constituents: temperature indicators and optical imaging instruments. Recent advances in biochemistry, materials science, and molecular biology have discovered numerous labeling agents [126-129] for PLI-based temperature sensing. Among them, lanthanide-doped upconverting nanoparticles (UCNPs) are ideal candidates. Leveraging the long-lived excited states provided by the lanthanide ions, UCNPs can sequentially absorb two (or more) low-energy near-infrared photons and convert them to one higher-energy photon. This upconversion process allows using excitation power densities several orders of magnitude lower than those needed for simultaneous multi-photon absorption [130, 131]. The near-infrared excitation, with smaller tissue extinction coefficients, also gains deeper penetration [132]. Besides, the upconverted luminescence, particularly the Boltzmann-coupled emission bands in co-doped erbium/ytterbium (Er^{3+}/Yb^{3+}) systems, is highly sensitive to temperature changes [133, 134]. Moreover, long-lived (i.e., microseconds to milliseconds) photoluminescence of UCNPs circumvents interferences from autofluorescence and scattering during image acquisition, which translates into improved imaging contrast and

detection sensitivity. Finally, because of advances in their synthesis and surface functionalization coupled with the innovation of core/shell engineering, over the years, UCNPs have become much brighter, photostable, biocompatible, and non-toxic [135]. As a result of these salient merits, UCNPs are one of the frontrunners in temperature indicators for PLI.

Advanced optical imaging is the other indispensable constituent in PLI-based temperature mapping [136]. To detect photoluminescence on the time scale of microseconds to milliseconds, like that produced by UCNPs, most PLI techniques use point-scanning time-correlated single-photon counting (TCSPC) [137]. Although they possess high signal-to-noise ratios, the scanning operation leads to an excessively long imaging time to form a two-dimensional (2D) lifetime map because extended pixel dwell time is required to record the long-lived emission [138]. To accelerate data acquisition, wide-field PLI modalities based on parallel collection in time-domain and frequency-domain have been developed [139]. In the time domain, these techniques extend the TCSPC technique to wide-field imaging (e.g., TimepixCam [140] and Tpx3Cam [141]). Photoluminescence decay over a 2D field of view (FOV) is synthesized from >100,000 frames, which requires the emission to be precisely repeatable. Alternatively, the frequency-domain wide-field PLI techniques [142, 143] use phase difference between the intensity-modulated excitation and the received photoluminescence signal to determine the 2D lifetime distribution. Nevertheless, limited by the range of frequency synthesizers, the measurable lifetimes are mostly restricted to ≤100 μs, which is shorter than the lifetimes of most UCNPs. Akin to the time-domain techniques, these systems rely on the integration over many periods of modulation intensity, during which the sample must remain stationary. Thus far, existing PLI techniques fall short in 2D temperature sensing of moving samples with a micrometer-level spatial resolution.

To surmount these limitations, we report an optical temperature mapping modality, termed single-shot photoluminescence lifetime imaging thermometry (SPLIT). Synergistically combining dual-view optical streak imaging with compressed sensing [144], SPLIT records wide-field luminescence decay of Er^{3+}, Yb^{3+} co-doped $NaGdF_4$ UCNPs in real time, from which a lifetime-based 2D temperature map is obtained in a single exposure. Largely advancing existing optical thermometry techniques in detection capabilities, SPLIT enables longitudinal 2D temperature monitoring beneath a thin scattering medium and dynamic temperature tracking of a moving biological sample at single-cell resolution.

4.2 Results

4.2.1 Operating principle of SPLIT

The schematic of the SPLIT system is shown in Fig. 4.1. A 980-nm continuous-wave laser (BWT, DS3-11312-113-LD) is used as the light source. The laser beam passes through a 4*f* system consisting of two 50-mm focal length lenses (L1 and L2, Thorlabs, LA1255). An optical chopper (Scitec Instruments, 300CD) is placed at the back focal plane of lens L1 to generate 50-µs optical pulses. Then, the pulse passes through a 100-mm focal length lens (L3, Thorlabs, AC254-100-B) and is reflected by a short-pass dichroic mirror (Edmund Optics, 69-219) to generate a focus on the back focal plane of an objective lens (Nikon, CF Achro 4×, 0.1 numerical aperture, 11-mm field number). This illumination scheme produces wide-field illumination (1.5×1.5 mm^2 FOV) to UCNPs at the object plane.

The near-infrared excited UCNPs emit light in the visible spectral range. The decay of light intensity over the 2D FOV is a dynamic scene, denoted by $I(x, y, t)$. The emitted light is collected by the same objective lens, transmits through the dichroic mirror, and is filtered by a band-pass filter (Thorlabs, MF542-20 or Semrock, FF01-660/30-25). Then, a beam splitter (Thorlabs, BS013) equally divides the light into two components. The reflected component is imaged by a CMOS camera (FLIR, GS3-U3-23S6M-C) with a camera lens (Fujinon, HF75SA1) via spatiotemporal integration (denoted as the operator $\mathbf{T}$) as View 1, whose optical energy distribution is denoted by $E_1(x_1, y_1)$.

The transmitted component forms an image of the dynamic scene on a transmissive encoding mask with a pseudo-random binary pattern (Fineline Imaging, 50% transmission ratio; 60-µm encoding pixel size). This process of spatial encoding is denoted by the operator $\mathbf{C}$. Then, the spatially encoded scene is imaged by a mechanical streak camera. In particular, the scene is relayed to the sensor plane of an electron-multiplying (EM) CCD camera (Nüvü Camēras, HNü 1024) by a 4*f* imaging system consisting of two 100-mm focal length lenses (L4 and L5, Thorlabs, AC254-100-A). A galvanometer scanner (Cambridge Technology, 6220H), placed at the Fourier plane of the 4*f* imaging system, temporally shears the spatially encoded frames linearly to different spatial locations along the x_2 axis of the EMCCD camera according to their time of arrival. This process of temporal shearing is denoted by the operator $\mathbf{S}$. Finally, the spatially encoded and temporally sheared dynamic scene is recorded by the EMCCD via spatiotemporal integration to form View 2, whose optical energy distribution is denoted by $E_2(x_2, y_2)$.

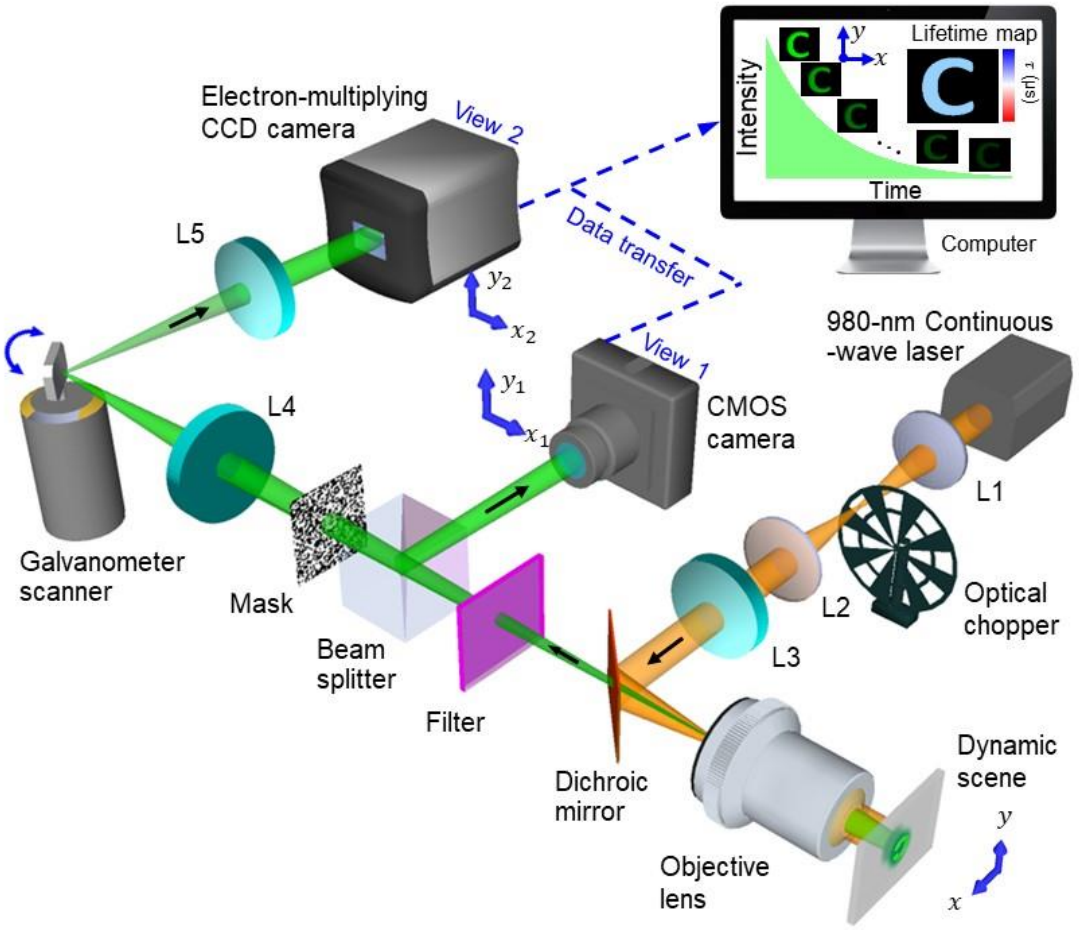


Figure 4.1 Schematic of the SPLIT system. The illustration shows data acquisition and image reconstruction of luminescence intensity decay in a letter “C”. L1–L5, Lens.

By combining the image formation of $E_1(x_1, y_1)$ and $E_2(x_2, y_2)$, the data acquisition of SPLIT is expressed by

$$E = \mathbf{TM}\, I, \tag{4.1}$$

where E denotes the concatenation of measurements $[E_1, \alpha E_2]^T$ (the superscript T denotes the transpose), $\mathbf{M}$ denotes the linear operator $[\mathbf{1}, \alpha\mathbf{SC}]^T$, and α is a scalar factor introduced to balance the energy ratio between the two views during measurement [145]. The hardware of the SPLIT system is synchronized for capturing both views (detailed in Methods) that are calibrated before data acquisition (detailed in Supplementary Note 1 and Supplementary Fig. 4.1).

After data acquisition, E is processed by an algorithm that retrieves the datacube of the dynamic scene by leveraging the spatiotemporal sparsity of the dynamic scene and the prior knowledge of each operator [146, 147]. Developed from the plug-and-play alternating direction method of multipliers (PnP-ADMM) framework [148, 149], the reconstruction algorithm of SPLIT solves the minimization problem of

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{TM}I - E\|_2^2 + R(I) + \mathbf{I}_+(I) \right\}. \tag{4.2}$$

Here, $\|\cdot\|_2$ represents the l_2 norm. The fidelity term, $\frac{1}{2}\|\mathbf{TM}I - E\|_2^2$, represents the similarity between the measurement and the estimated result. $R(\cdot)$ is the implicit regularizer that promotes sparsity in the dynamic scene. $\mathbf{I}_+(\cdot)$ represents a non-negative intensity constraint. Compared to existing reconstruction schemes [150-152], PnP-ADMM implements a variable splitting strategy with a state-of-the-art denoiser to obtain fast and closed-form solutions to each sub-optimization problem, which produces a high image quality in reconstruction (see Supplementary Notes 2 and 3 and Supplementary Fig. 4.2). The retrieved datacube of the dynamic scene has a sequence depth (i.e., the number of frames in a reconstructed movie) of 12–100 frames, each containing 460 × 460 (x, y) pixels. The imaging speed is tunable from 4 to 33 thousand frames per second (kfps) (detailed in Methods).

The reconstructed datacube is then converted to a photoluminescence lifetime map. In particular, for each (x, y) point, the area under the normalized intensity decay curve is integrated to report the value of the photoluminescence lifetime [153]. Finally, using the approximately linear relationship between the UCNPs' lifetime and the physiologically relevant temperature range (20–46 °C in this work) [121, 154], the 2D temperature distribution, $T(x, y)$, is calculated by

$$T(x, y) = c_{\mathrm{t}} + \frac{1}{S_{\mathrm{a}}} \int \frac{\hat{I}(x, y, t)}{\hat{I}(x, y, 0)} dt, \tag{4.3}$$

where c_{t} is a constant, and S_{a} is the absolute temperature sensitivity [136]. The derivation of Equation (4.3) is detailed in Supplementary Note 4. Leveraging the intrinsic frame rate of the EMCCD camera, the SPLIT system can generate lifetime-determined temperature maps at a video rate of 20 Hz.

4.2.2 Quantification of the system's performance of SPLIT

We prepared a series of core/shell UCNP samples to showcase SPLIT's capabilities. These UCNPs shared the same $NaGdF_4$: 2 mol% Er^{3+}, 20 mol% Yb^{3+} active core of 14.6 nm in size, while differed by the thickness of their undoped $NaGdF_4$ passive shell of 1.9, 3.5, and 5.6 nm (Fig. 4.2a and detailed in Supplementary Note 5). All of the UCNP samples were of pure hexagonal crystal phase (Supplementary Fig. 4.3). Under the 980-nm excitation, upconversion emission bands of all samples were measured at around 525/545 nm and 660 nm, which correspond to the $^2H_{11/2}/^4S_{3/2} \rightarrow {}^4I_{15/2}$ and $^4F_{9/2} \rightarrow {}^4I_{15/2}$ radiative transitions, respectively (Figs. 4.2b–c).

To characterize SPLIT's spatial resolution, we covered the 5.6 nm-thick-shell UCNP sample with a negative USAF resolution target (Edmund Optics, 55-622). Operating at 33 kfps, SPLIT recorded the photoluminescence decay. The temporally projected datacube reveals that the intensity and contrast in the reconstructed image degrade with the decreased spatial feature sizes, eventually leading to the loss of structure whose size approaches that of the encoding pixel (Fig. 4.2d). The effective spatial resolution was thus determined to be 20 μm (Fig. 4.2e). Under these experimental conditions, the minimum power density for the SPLIT system was quantified to be 0.06 W mm^{-2} (detailed in Supplementary Note 6 and Supplementary Fig. 4.4).

To demonstrate SPLIT's ability to distinguish different lifetimes, we imaged the UCNPs with shell thicknesses of 1.9 nm, 3.5 nm, and 5.6 nm, covered by transparencies of letters "C", "A", and "N", respectively, using a single laser pulse. The lifetime maps of these samples are shown in Fig. 4.2f, which reveals the averaged lifetimes for the $^4S_{3/2}$ excited state of samples "C", "A", and "N" to be 142 μs, 335 μs, and 478 μs, respectively (Figs. 4.2g–h). These results were verified by using the standard TCSPC method (detailed in Supplementary Note 7 and Supplementary Fig. 4.5).

SPLIT's reconstruction algorithm shows a superb performance to existing mainstream algorithms popularly used in single-shot compressed ultrafast imaging [97, 144, 150, 152]. By using the experimental data, the comparison demonstrates that the dual-view PnP-ADMM algorithm used by SPLIT is more powerful in preserving spatial features while maintaining a low background, which enables a more accurate lifetime quantification and the ensuing temperature mapping (detailed in Supplementary Note 8 and Supplementary Fig. 4.6).

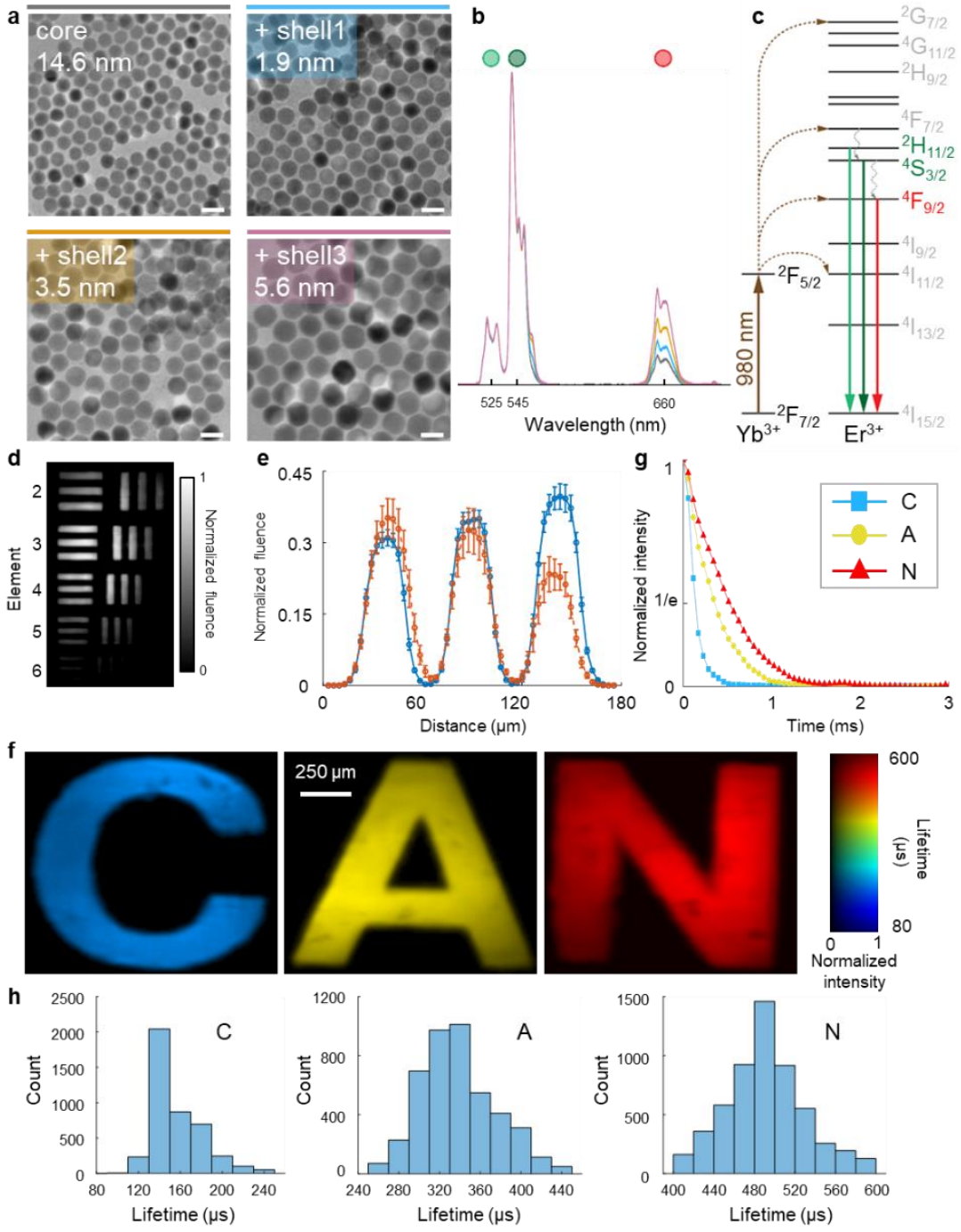


Figure 4.2 Quantification of the performance of the SPLIT system. **a** Images of core/shell UCNPs acquired with a transmission electron microscope. Scale bar: 25 nm. **b** Normalized upconversion spectra of UCNPs shown in (a). **c** Simplified energy level diagram of Yb^{3+}-Er^{3+} energy transfer upconversion excitation and emission. **d** Temporally projected image of photoluminescence intensity decay of the 5.6 nm-thick-shell UCNPs covered by a negative resolution target. **e** Comparison of averaged light fluence distribution along the horizontal bars (blue) and vertical bars (orange) of Element 5 in Group 4 on the resolution target. Error bar: standard deviation. **f** Lifetime images of UCNPs with the shell thicknesses of 1.9 nm, 3.5 nm, and 5.6 nm covered by transparencies of letters “C”, “A”, and “N” in green emission. **g** Time-lapse averaged emission intensities of the samples. h Histograms of photoluminescence lifetimes in the letters shown in (f).

4.2.3 Single-shot temperature mapping using SPLIT

We used the 5.6 nm-thick-shell UCNPs as the temperature indicator for SPLIT. The UCNPs' temperature was controlled by a heating plate placed behind the sample. To image the green ($^4S_{3/2}$) and red ($^4F_{9/2}$) upconversion emissions, the sample was covered by transparencies of a lily flower and a maple leaf, respectively. The temperature of the entire sample was measured with both a Type K thermocouple (Omega, HH306A) and a thermal camera (FLIR, E4) as references. The reconstructed lifetime images in the 20–46 ºC temperature range are shown in Figs. 4.3a–b. Plotted in Figs. 4.3c–d, the time-lapse averaged intensity over the entire FOV shows that the averaged lifetimes of green and red emissions decrease from 489 to 440 μs and from 458 to 398 μs, which is due to their enhanced multi-phonon deactivation at higher temperatures. We further plotted the relationship between the temperatures and lifetimes for both emission channels (Fig. 4.3e). Finally, the temperature sensitivities in the preset temperature range were calculated to be $S_{\mathrm{a}} = -1.90\ \mu\mathrm{s}\ ^{\circ}\mathrm{C}^{-1}$ for the green emission and $S_{\mathrm{a}} = -2.40\ \mu\mathrm{s}\ ^{\circ}\mathrm{C}^{-1}$ for the red emission (see detailed calculation and further analysis in Supplementary Note 9 and Supplementary Fig. 4.7). Compared to the green emission, the higher temperature sensitivity of the red emission results from the greater energy separation between its emitting state and the adjacent lower-laying excited state (Fig. 4.2c). Since multi-phonon relaxation rate depends exponentially on the number of phonons necessary to deactivate an excited state to the one below it, the increase in phonon energies at higher temperatures has greater influence over the states with a larger energy gap between them [155]. These results establish lifetime-temperature calibration curves [i.e., Equation (4.3)] for ensuing thermometry experiments.

To demonstrate SPLIT's feasibility in a biological environment, we conducted longitudinal temperature monitoring under a phantom, made by using the 5.6 nm-thick-shell UCNPs covered by lift-out grids (Ted Pella, 460-2031-S), overlaid by fresh chicken breast tissue. We investigated SPLIT's imaging depth with varied tissue thicknesses of up to 1 mm (Fig. 4.3f, Supplementary Note 10 and Supplementary Fig. 4.8). The chicken tissue of 0.5 mm thickness, where both the green and red emissions produced images with full spatial features of the lift-out grid, was used in the following imaging experiments. Subsequently, we cycled the temperature of the sample between 20 ºC and 46 ºC. The lifetime distributions of both green and red emissions and their corresponding temperature maps were monitored every 20 minutes and 23 minutes, respectively, for ~4 hours (see the full evolution in Supplementary Fig. 4.9). As shown in Fig. 4.3g, the results are in good agreement with the temperature change preset by the heating plate, and decisively

showcase how SPLIT can map 2D temperatures over time with high accuracy beneath biological tissue.

We also demonstrated SPLIT using a fresh beef phantom as a scattering medium, where both light scattering and absorption are present (detailed in Supplementary Note 10 and Supplementary Fig. 4.10). The results reveal better penetration of the red emission over the green counterpart due to its weaker scattering and absorption. More importantly, the results confirm the independence of the measured photoluminescence lifetime of UCNPs to tissue thickness and hence the excitation light power density used in our work (≤0.4 W mm^{-2}).

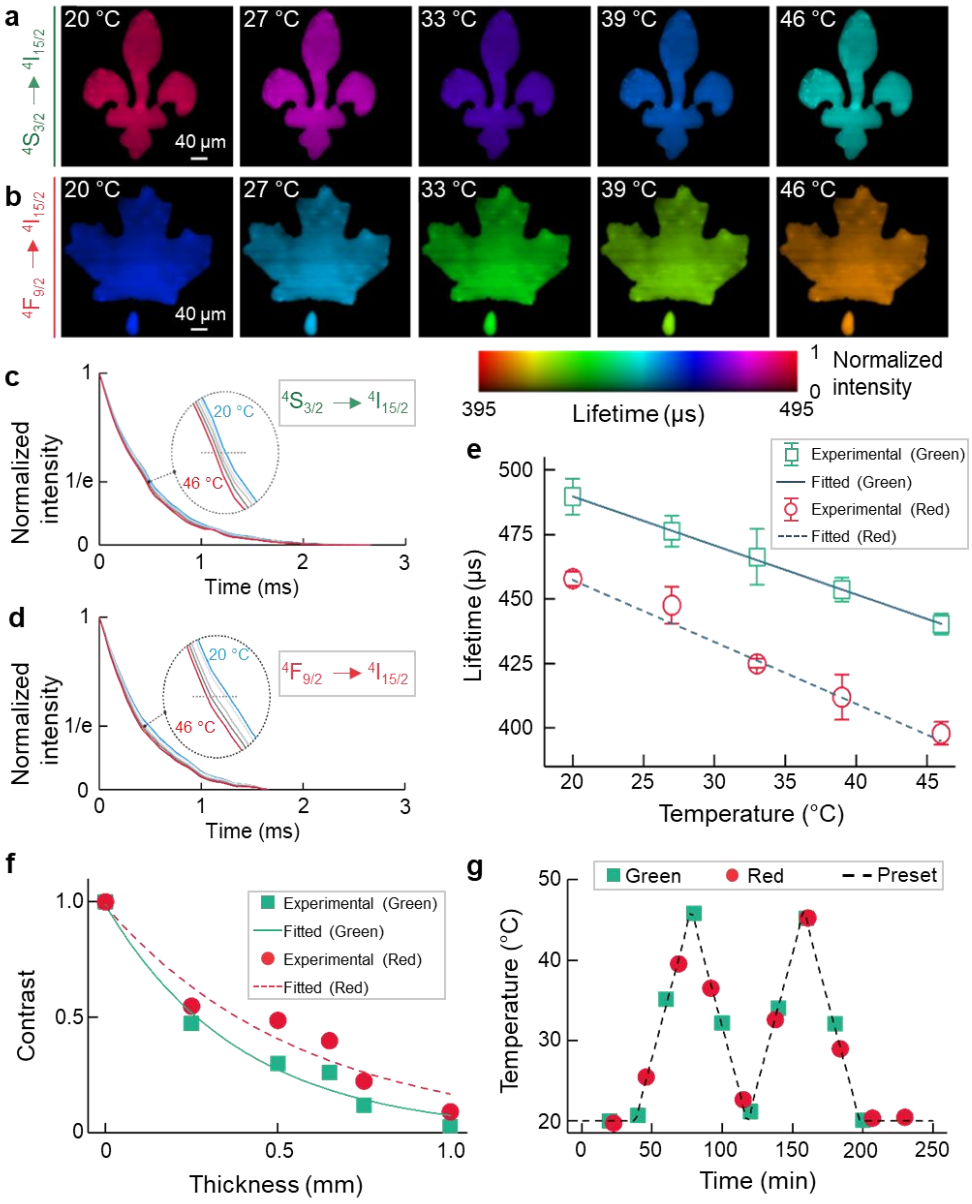


Figure 4.3 Single-shot temperature mapping using SPLIT. **a**–**b** Lifetime images of green (a) and red (b) upconversion emission bands under different temperatures. **c**–**d** Normalized photoluminescence decay curves of green (c) and red (d) emission bands at different temperatures, averaged over the entire field of view. **e** Relationship between temperature and mean lifetimes of green and red emissions with linear fitting. Error bar: standard deviation from three independent measurements. **f** Normalized contrast versus chicken tissue thickness for green and red emission bands with single-component exponential fitting. **g** Longitudinal temperature monitoring through 0.5 mm-thick fresh chicken tissue.

4.2.4 Single-cell dynamic temperature tracking using SPLIT

To apply SPLIT to dynamic single-cell temperature mapping, we tested a single-layer onion epidermis sample labeled by the 5.6 nm-thick-shell UCNPs (detailed in Supplementary Note 11 and Supplementary Fig. 4.11). Furthermore, to generate non-repeatable photoluminescent dynamics, the sample was moved across the FOV at a speed of 1.18 mm s^{-1} by a translation stage. In the 3-second measurement window, the SPLIT system continuously recorded 60 lifetime/temperature maps. Four representative time-integrated images and their corresponding lifetime maps are shown in Figs. 4.4a–b. Figure 4.4c shows intensity decay curves from four selected regions with varied intensities in the onion cell sample at 0.05 seconds. The photoluminescence lifetimes and hence the temperatures remain stable, showing SPLIT's resilience to spatial intensity variation. We also tracked the time histories of the averaged emitted fluence and lifetime-indicated temperatures of these four regions during the sample's translational moving (Fig. 4.4d). In this measurement time window, the emitted photoluminescence fluences have varied in each selected region. In contrast, the measured temperatures show a small fluctuation of ±0.35 °C, which validates the advantage of PLI thermometry in handling temporal intensity variation.

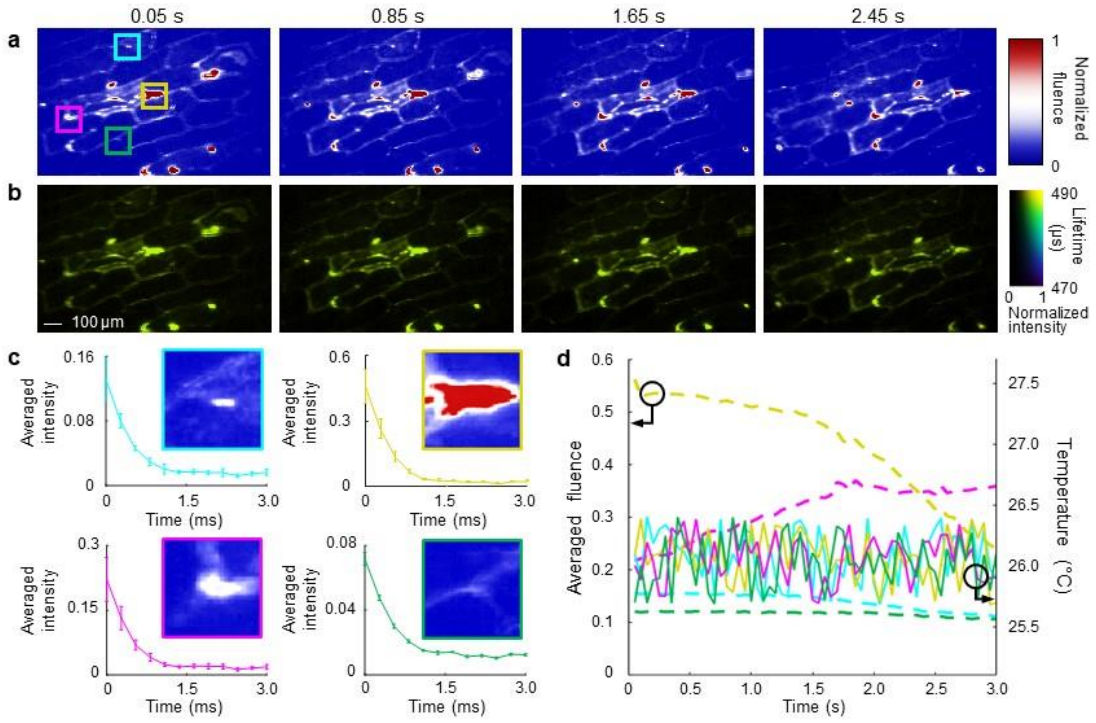


Figure 4.4 Dynamic single-cell temperature mapping using SPLIT. **a** Representative time-integrated images of a moving onion epidermis cell sample labeled by UCNPs. **b** Lifetime images corresponding to (a). **c** Photoluminescence decay profiles at four selected areas [marked by the solid boxes in the first panel of (a)] with varied intensities. **d** Time histories of averaged fluence and corresponding temperature in the four selected regions during the sample's translational motion.

4.3 Discussion

In summary, we have developed SPLIT for wide-field dynamic temperature sensing in real time. In data acquisition, SPLIT compressively records the photoluminescence emission over a 2D FOV in two views. Then, the dual-view PnP-ADMM algorithm reconstructs spatially resolved intensity decay traces, from which a photoluminescence lifetime distribution and the corresponding temperature map are extracted. Used with core/shell $NaGdF_4:Er^{3+},Yb^{3+}/NaGdF_4$ UCNPs, SPLIT has enabled temperature mapping with high sensitivity for both green and red upconversion emission bands with a 20-μm spatial resolution in a 1.5×1.5 mm^2 FOV at a video rate of 20 Hz. SPLIT is demonstrated in longitudinal temperature monitoring of a phantom beneath fresh chicken tissue. SPLIT is also applied to dynamic single-cell temperature mapping of a moving single-layer onion epidermis sample.

SPLIT advances the technical frontier of optical instrumentation in PLI. The high parallelism in SPLIT's data acquisition drastically improves the overall light throughput. The resulting system, featuring single-shot temperature sensing over a 2D FOV, solves the long-

standing issue in scanning-based techniques (see Supplementary Note 12 and Supplementary Figs. 4.12-4.13). In particular, SPLIT improves the measurement accuracy by avoiding artifacts generated from the scanning-induced motion blur and the excitation intensity fluctuation. More importantly, as shown in Fig. 4, SPLIT extends the application scope of PLI to observing non-repeatable 2D temperature dynamics. Its high tunability of imaging speeds also accommodates a variety of UCNPs with a wide lifetime span (from hundreds of nanoseconds to milliseconds). Among existing single-shot 2D ultrafast imaging modalities based on streak cameras, SPLIT is well suited for dynamic PLI of UCNPs in terms of the targeted imaging speed, detection sensitivity, spatial resolution, and cost efficiency (detailed in Supplementary Note 12 and Supplementary Table 4.1). Finally, the SPLIT system by itself records only the lifetime images; yet, when using UCNPs as contrast agents, those images carry temperature information *in situ*, where the UCNPs reside. Thus, compared to thermal imaging cameras, SPLIT supplies superior temperature mapping results with higher image contrast and better resilience to background interference (detailed in Supplementary Note 13 and Supplementary Fig. 4.14).

From the perspective of system design, both the dual-view data acquisition and the PnP-ADMM algorithm support high imaging quality in SPLIT. In particular, View 1 preserves the spatial information in the dynamic scene [156]. Meanwhile, View 2 retains temporal information by optical streaking via time-to-space conversion. Altogether, both views maximally keep rich spatiotemporal information. In software, the dual-view PnP-ADMM algorithm provides a powerful modular structure, which allows separated optimization of individual sub-optimization problems with an advanced denoising algorithm to generate high-quality image restoration results.

SPLIT offers a versatile PLI temperature-sensing platform. In materials characterization, it could be used in the stress analysis of metal fatigue in turbine blades [157]. In biomedicine, it could be implemented for accurate sub-cutaneous temperature monitoring for theranostics of skin diseases (e.g., micro-melanoma) [158, 159]. SPLIT's microscopic temperature mapping ability could also be exploited for the studies of temperature-regulated cellular signaling [160]. Finally, the operation of SPLIT could be extended to Stokes emission in lanthanide-doped nanoparticles and spectrally resolved temperature mapping. All of these topics are promising research directions in the future.

4.4 Methods

4.4.1 Synchronization of the SPLIT system

The optical chopper outputs a transistor-transistor logic (TTL) signal that is synchronized with the generated optical pulses. This TTL signal is input to a delay generator (Stanford Research Systems, DG 645), which then generates three synchronized TTL signals at 20 Hz. The first two signals are used to trigger the 3-ms exposure of the EMCCD and CMOS cameras. The last one is used to trigger a function generator (Rigol, DG1022Z) that outputs a 20-Hz sinusoidal waveform under the external burst mode to control the rotation of the galvanometer scanner (GS).

4.4.2 Calculation of SPLIT's key parameters

The GS, placed at the Fourier plane of the $4f$ imaging system consisting of lenses L4 and L5 (Fig. 4.1), deflects temporal information to different spatial positions. Rotating during the data acquisition, the GS changes the reflection angles of the spatial frequency spectra of individual frames with different time-of-arrival. After the Fourier transformation by Lens 5, this angular difference is converted to the lateral shift in space on the EMCCD camera, which results in temporal shearing. An illustration with a simple example is provided in Supplementary Fig. 4.15.

The imaging speed is determined by the data acquisition for View 2. In particular, the reconstructed movie has a frame rate of [161]

$$\mathrm{r} = \frac{\gamma_\mathrm{a} V_\mathrm{g} f_5}{t_\mathrm{s} d}. \tag{4.4}$$

Here, V_g is the voltage added onto the GS. γ_a is a constant that links V_g with GS's deflection angle with the consideration of the input waveform. f_5=100 mm is the focal length of lens L5, $t_\mathrm{s} = 50\ \mathrm{ms}$ is the period of the sinusoidal voltage waveform added to the GS, and d = 13 μm is the EMCCD sensor's pixel size. In this work, we used the voltage from $V_\mathrm{g} = 0.2$–1.7 V. The imaging speed of SPLIT ranged from 4 to 33 kfps. In addition, we used $t_\mathrm{e} = 3\ \mathrm{ms}$ as the exposure time of the EMCCD and CMOS cameras. The sequence depth, N_t, is determined by

$$N_t = \mathrm{r} t_\mathrm{e}. \tag{4.5}$$

In the experiments presented in this work, N_t ranged from 12 to 100 frames.

Acknowledgements

The authors thank Professor Aycan Yurtsever and Wanting He for experimental assistance and fruitful discussion.

Funding

Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05959, RGPAS-2017-507845, I2IPJ-555593-20, RGPIN-2018-06217, RGPAS-2018-522650); Canada Foundation for Innovation and Ministère de l'Économie et de l'Innovation du Québec (37146); Canadian Cancer Society (707056); New Frontier in Research Fund (NFRFE-2020-00267); Fonds de Recherche du Québec–Nature et Technologies (2019-NC-252960); Fonds de Recherche du Québec–Santé (267406, 280229).

Competing interests

The authors disclose the following patent applications: WO 2020/154806 A1 (J.Liang, F.V., and X.L.) and US Provisional 63/260,511 (J.Liang, F.V., X.L., and A.S.).

Data availability

All data needed to evaluate the findings of this study are present in the paper and Supplementary Information. The raw data for Fig. 2 can be downloaded via the following link: https://figshare.com/articles/figure/SPLIT_Fig2/16703413. All other raw data are available from the corresponding authors upon reasonable request.

Code availability

The image reconstruction algorithm is described in detail in Supplementary Information. The custom computer code is not publicly available because it is proprietary and included in a patent application.

Supplementary materials

Fast Wide-field Upconversion Luminescence Lifetime Thermometry Enabled by Single-shot Compressed Ultrahigh-speed Imaging

Xianglei Liu[1,3], Artiom Skripka[1,2,3], Yingming Lai[1], Cheng Jiang[1], Jingdan Liu[1], Fiorenzo Vetrone[1,*], Jinyang Liang[1,*]

[1]Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650 boulevard Lionel-Boulet, Varennes, Québec J3X1S2, CANADA

[2]Present address: Nanomaterials for Bioimaging Group, Departamento de Física de Materiales, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, 28049, SPAIN and The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[3]These authors contributed equally to this work

*Corresponding authors: Fiorenzo.Vetrone@inrs.ca (F.V.) and Jinyang.Liang@inrs.ca (J.Liang)

Supplementary Note 1: Two-view image registration of the SPLIT system

To conduct the image registration between the two views, we used an established procedure [162] to calibrate the single-shot photoluminescence lifetime imaging thermometry (SPLIT) system. In particular, a static target of upconverting nanoparticles (UCNPs) was imaged by the SPLIT system to form View 1 and View 2. No optical shearing was performed in the recording of View 2. The projective transformation was then quantified by using the registration estimator toolbox in MATLAB R2019b [163], which supplied a feature-based registration operator to automatically detect distinct local features, such as sharp corners, blobs, or regions of images. The transformation matrix $\mathbf{P}_{\mathrm{tm}}$ is defined as

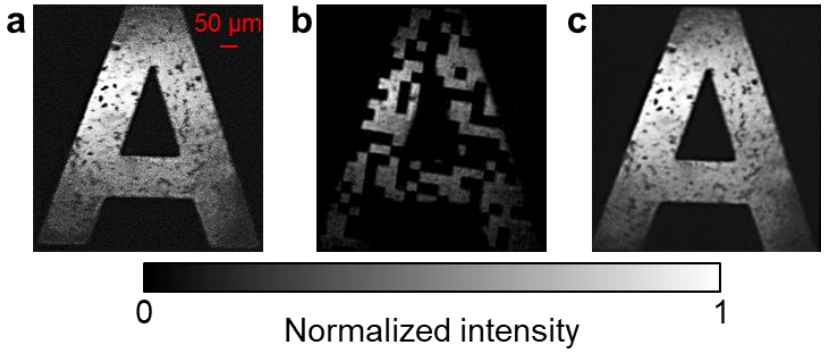
$$\mathbf{P}_{\mathrm{tm}} = \begin{bmatrix} s_x \cos\theta & -s_y \sin\theta & l_x \\ s_x \sin\theta & s_y \cos\theta & l_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{S4.1}$$

Here s_x and s_y are the scaling factors in the x direction and the y direction. θ is the rotation angle. l_x and l_y represent translation factors in the x direction and the y direction, respectively. Each

pixel in View 1 with a homogeneous coordinate $[h_x\ h_y\ 1]$ is transformed to the corresponding point $[h_{xc}\ h_{yc}\ 1]$ by

$$\left[h_{xc}\ h_{yc}\ 1\right]^T = \mathbf{P}_{\mathrm{tm}}\left[h_x\ h_y\ 1\right]^T. \tag{S4.2}$$

In practice, $\mathbf{P}_{\mathrm{tm}}$ was computed by using the static letter "A" pattern. Supplementary Figs. 1a–b show the acquired images in View 1 and View 2. The co-registered View 1 image (Supplementary Fig. 4.1c) and the View 2 image were used for SPLIT's image reconstruction.



Supplementary Figure 4.1 Image registration in SPLIT's dual-view data acquisition. **a** Image acquired in View 1. **b** Image acquired in View 2 without using optical shearing. **c** Co-registered image of View 1.

Supplementary Note 2: Derivation of the SPLIT's reconstruction algorithm

In image reconstruction, the datacube of the dynamic scene is recovered by solving the minimization problem aided by regularizers [164]. In particular, the inverse problem [i.e., Equation (4.2)] is first written as

$$\begin{aligned}\hat{I} = \underset{I,v,u,w\in\mathbf{A}}{\operatorname{argmin}} \left\{\frac{1}{2}\|\mathbf{T}v - E\|_2^2 + R(u) + \mathbf{I}_+(w)\right\} \\ \text{subject to } v = \mathbf{M}I, u = I, w = I,\end{aligned} \tag{S4.3}$$

where v, u, and w are primal variables. $\mathbf{A}$ is the set of possible solutions in compliance with the spatial constraint [47], which is generated by binarizing the image of View 1 (i.e., E_1) with an appropriate intensity threshold that is determined by the Otsu's method [165]. Then, Supplementary Equation (S4.3) is further written in the augmented Lagrangian arguments [52, 55, 166]:

$$\hat{I} = \underset{I,v,u,w\in\mathbf{A}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{T}v - E\|_2^2 + R(u) + \mathbf{I}_+(w) + \frac{\mu_1}{2}\|\mathbf{M}I - v + \frac{\gamma_1}{\mu_1}\|_2^2 + \frac{\mu_2}{2}\|I - u + \frac{\gamma_2}{\mu_2}\|_2^2 + \frac{\mu_3}{2}\|I - w + \frac{\gamma_3}{\mu_3}\|_2^2 \right\}. \tag{S4.4}$$

Here, γ_1, γ_2, and γ_3 are dual variables. μ_1, μ_2, and μ_3 are penalty parameters [167, 168]. The block-matching and 3D (BM3D) filtering [169] is used as the plug-and-play (PnP) denoiser in the implicit regularizer $R(\cdot)$. The ramp function [170] is used in the non-negative indicator function $\mathbf{I}_+(\cdot)$.

To retrieve the dynamic scene, the reconstruction algorithm sequentially updates primal variables, estimated solution I^{k+1} (k denotes the iteration time), dual variables and penalty parameters as well as evaluates the pre-set criteria, as following five steps.

Step 1: update primal variables (i.e., v, u, and w) by

$$\begin{aligned} v^{k+1} &= (\mathbf{T}^T \cdot \mathbf{T} + \mu_1^k \mathrm{D})^{-1} \cdot \left(\mathbf{T}^{\mathrm{T}} E + \mu_1^k \mathbf{M} I^k + {\gamma_1}^k\right), \\ u^{k+1} &= \mathrm{D}_{\mathrm{BM3D}}(I^k + \frac{{\gamma_2}^k}{\mu_2^k}), \quad \text{and} \\ w^{k+1} &= \max\{0, I^k + \frac{{\gamma_3}^k}{\mu_3^k}\}. \end{aligned} \tag{S4.5}$$

Here, D is the identity matrix. $\mathrm{D}_{\mathrm{BM3D}}(\cdot)$ stands for the BM3D filtering [169].

Step 2: update the estimated datacube of the dynamic scene [i.e., $I(x, y, t)$] by

$$I^{k+1} = \left(\mu_1^k \mathbf{M}^T \cdot \mathbf{M} \cdot \mathrm{D} + \mu_2^k \mathrm{D} + \mu_3^k \mathrm{D}\right)^{-1} \left[\mu_1^k \mathbf{M}^T (v^{k+1} - \frac{{\gamma_1}^k}{\mu_1^k}) + \mu_2^k (u^{k+1} - \frac{{\gamma_2}^k}{\mu_2^k}) + \mu_3^k (w^{k+1} - \frac{{\gamma_3}^k}{\mu_3^k})\right]. \tag{S4.6}$$

Step 3: update dual variables (i.e., $\gamma_1, \gamma_2,$ and γ_3) by

$$\begin{aligned} {\gamma_1}^{k+1} &= {\gamma_1}^k + \mu_1^k \left(\mathbf{M} I^{k+1} - v^{k+1}\right), \\ {\gamma_2}^{k+1} &= {\gamma_2}^k + \mu_2^k \left(I^{k+1} - u^{k+1}\right), \quad \text{and} \\ {\gamma_3}^{k+1} &= {\gamma_3}^k + \mu_3^k \left(I^{k+1} - w^{k+1}\right). \end{aligned} \tag{S4.7}$$

Step 4: update the penalty parameters (i.e., μ_1, μ_2, and μ_3) by

$$\mu_i^{k+1} = \begin{cases} \varphi\mu_i^k, & \text{if } p > \sigma q \\ \dfrac{\mu_i^k}{\varphi}, & \text{if } \sigma p < q \\ \mu_i^k, & \text{otherwise} \end{cases} \quad (i = 1, 2, 3). \tag{S4.8}$$

Here, $p = \left\| I^{k+1} - v^{k+1} \right\|_2$ is the primal residual, and $q = \mu_i^k \left\| I^{k+1} - I^k \right\|_2$ is the dual residual. φ ($\varphi > 1$) is the balancing factor, and σ ($\sigma > 1$) is the residual tolerance [171]. In our experiments, we chose $\varphi = 1.1$ and $\sigma = 1.5$.

Step 5: judge the relative change in results and the parameters μ_1^{k+1}, μ_2^{k+1}, and μ_3^{k+1} in adjacent iterations by

$$\text{if } \eta = \frac{\left\| I^{k+1} - I^k \right\|_2}{\left\| I^{k+1} \right\|_2} < \rho \text{ and } \mu_i^{k+1} = \mu_i^k \ (i = 1, 2, 3). \tag{S4.9}$$

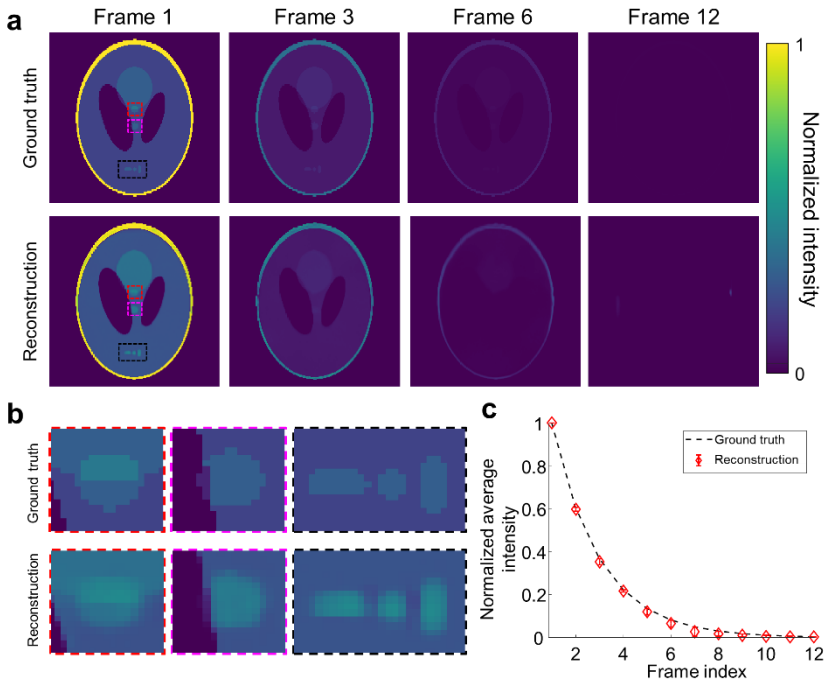
Here, ρ ($0 < \rho < 10^{-3}$) is the pre-set tolerance value. These steps are repeated until both criteria in Step 5 are satisfied. The image reconstruction recovers the datacube of the dynamic scene.

Supplementary Note 3: Simulation results of the dual-view PnP-ADMM algorithm

To test the proposed dual-view PnP alternating direction method of multipliers (ADMM) algorithm, we reconstructed a simulated dynamic scene—the intensity decay of a static Shepp-Logan phantom. This dynamic scene contains 12 frames, each with a size of 200×200 pixels. The intensity in each frame is determined by a single exponential function of $I_{n_t} = \exp[-(n_t - 1)/2]$, where $n_t = 1, \ldots, 12$ denotes the frame index.

Then, this dynamic scene was fed into SPLIT's forward model [i.e., Equation (4.1)] to generate E_1 and E_2. To mimic the experimental conditions, we added Gaussian noise (0.01 variance and 0 mean value) into E_1 and E_2, respectively. Finally, these two images were input into the dual-view PnP-ADMM algorithm to retrieve the datacube of this dynamic scene. Representative reconstructed frames and their corresponding ground truth frames are compared side by side in Supplementary Fig. 4.2a. The averaged peak signal-to-noise ratio (SNR) and the averaged structural similarity index over all reconstructed images were calculated to be 34.6 dB and 0.96, respectively. The reconstructed three local features in Frame 1 are compared to their ground truths (Supplementary Fig. 4.2b). Supplementary Fig. 4.2c presents the reconstructed

normalized average intensity versus time, which has a good agreement with the pre-set intensity decay (black dashed line).



Supplementary Figure 4.2 Simulation of the dual-view PnP-ADMM reconstruction algorithm. **a** Comparison of representative frames of the reconstructed result with the ground truth. **b** Comparison of three local features in Frame 1 of the reconstructed result with the ground truth (marked by the red, magenta, and black dashed boxes). **c** Normalized average intensity of the reconstructed result versus the frame index. Error bar: standard deviation.

Supplementary Note 4: Details on the relationship between temperature and lifetime

The normalized area integration method is commonly used for calculating lifetime based on pulsed excitation [172]. Photoluminescence lifetime of UCNPs following pulsed excitation can be expressed by

$$\mathrm{L_t} = \int_0^\infty f(t) * g(t) dt. \tag{S4.10}$$

Here $f(t) = \frac{1}{\sqrt{\pi} t_\mathrm{w}} \exp\left(-\frac{t^2}{t_\mathrm{w}^2}\right)$ represents the Gaussian excitation pulse with a pulse width of t_w. $g(t) = \sum \varepsilon_i \exp\left({}^{-t}/_{\tau_i}\right)$ is used to represent the photoluminescence with multiple exponential

decays, each of which has a lifetime τ_i and a proportion ε_i. "$*$" denotes convolution. Then, Supplementary Equation (S4.10) becomes

$$\mathrm{L_t} = \sum \varepsilon_i \tau_i \exp\left(\frac{t_\mathrm{w}^2}{4\tau_i^2}\right). \tag{S4.11}$$

When t_w approaches to zero, which denotes the case of an ultrashort pulse, Supplementary Equation (S4.11) becomes

$$\mathrm{L_t} = \sum \varepsilon_i \tau_i. \tag{S4.12}$$

Following the established theory [173], we defined the photoluminescence lifetime as $\tau = \sum \varepsilon_i \tau_i / \sum \varepsilon_i$. Considering that $\sum \varepsilon_i = 1$, we have $\tau = \mathrm{L_t}$.

The lifetime is linearly linked to the temperature by

$$T = c_\mathrm{t} + \frac{\mathrm{L_t}}{S_\mathrm{a}}. \tag{S4.13}$$

Here S_a denotes the absolute temperature sensitivity, and c_t denotes a constant. This derivation produces Equation (4.3).

In the SPLIT system, we used a continuous-wave laser and an optical chopper to generate excitation pulses. Although the chopper blade's slit width could approach zero for generating an ultrashort pulse duration, it demands a high laser power. Thus, a finite pulse width needs to be chosen to provide sufficient SNRs in measurements while still maintaining accurate lifetime calculation. In practice, we chose $t_\mathrm{w} = 50\ \mu\mathrm{s}$, which was comparable to the values used in the literature [172]. Our calculation also showed that this pulse width induced a <0.3% calculation error for the 5.6-mm-thick-shell UCNPs that were mainly used in our experiments. Thus, 50-μs pulse width allowed SPLIT to produce accurate temperature mapping results.

Supplementary Note 5: Preparation and characterization of UCNPs

Synthesis of UCNPs

Core/shell $NaGdF_4$: 2 mol% Er^{3+}, 20 mol% Yb^{3+}/$NaGdF_4$ UCNPs were synthesized via the previously reported thermal decomposition method, with minor modifications to the synthesis procedure [174]. Core precursors were prepared by mixing 0.025 mmol of Er_2O_3 (REacton 99.99%), 0.250 mmol Yb_2O_3 (REacton 99.99+%), and 0.975 mmol Gd_2O_3 (REacton 99.99+%) with 5 mL trifluoroacetic acid (99%) and 5 mL of distilled water in a 50 mL three-neck round bottom flask. Shell precursors were prepared separately by mixing 1.5 mmol of Gd_2O_3 with 5 mL of trifluoroacetic acid and 5 mL of distilled water in a 50 mL three-neck round bottom flask. Mixtures were refluxed under vigorous stirring at 80 °C until each solution turned from turbid to clear, at which point the temperature was decreased to 60 °C to slowly evaporate the excess trifluoroacetic acid and water. All precursors were obtained as solid dried materials and were used for the UCNPs synthesis without further purification. All materials involved in the precursor synthesis (obtained from Alfa Aesar) were used without further purification.

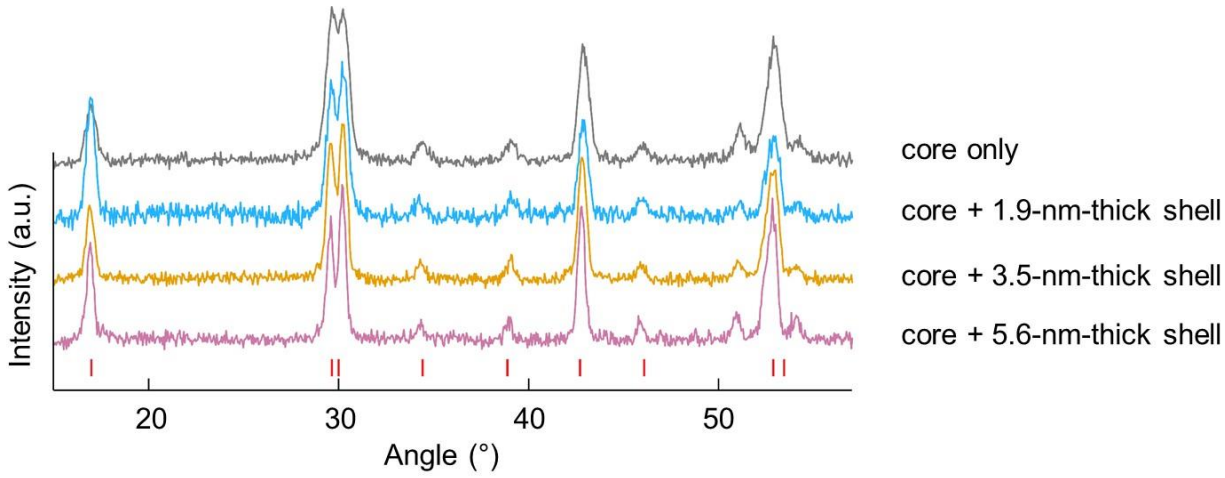
The first step was to synthesize the core UCNPs. An initial mixture of 12.5 mL each of oleic acid (OA; 90%, Alfa Aesar) and 1-octadecene (ODE; 90%, Alfa Aesar) was prepared in a 100 mL three-neck round bottom flask (Solution A). Aside, 2.5 mmol of sodium trifluoroacetate (98%, Alfa Aesar) was added to the dried core precursor together with 7.5 mL each of oleic acid and 1-octadecene (Solution B). Both Solutions A and B were degassed at 145 ºC under vacuum with magnetic stirring for 30 minutes. After degassing, Solution A was placed under an inert Ar atmosphere and the temperature was slowly raised to 315 ºC. Solution B was then injected into the reaction vessel containing Solution A using a syringe and pump system (Harvard Apparatus, Pump 11 Elite) at a 1.5 mL min^{-1} injection rate. The mixture was left at 315 ºC under vigorous stirring for 60 minutes. The synthesized core UCNPs were stored in Falcon centrifuge tubes (50 mL) under Ar for the further shelling step. Due to the evaporation of impurities in starting materials (e.g., OA and ODE) and reaction byproducts, as well as minor losses accrued from intermediate steps of liquid handling, the final volume of the core mixture was around 36 mL.

In the second step, core/shell UCNPs of different shell thicknesses were prepared by epitaxial growth of the shell on the preformed cores via a multi-step hot-injection approach. First, we prepared Solution A by mixing approximately 1.5 mmol of core UCNPs (~21.6 mL) in a 100 mL three-neck round bottom flask together with 9.2 mL each of OA and ODE. Separately, we prepared Solution B by mixing 3 mmol of gadolinium trifluoroacetate (shelling) precursors with 3

mmol of sodium trifluoroacetate, and 10.5 mL each of OA and ODE. Both solutions were degassed under vacuum and magnetic stirring at 110 ºC for 30 minutes. After degassing, Solution A was back-filled with argon gas and the temperature was raised to 315 ºC. Solution B was then injected into the reaction vessel containing Solution A using a syringe and pump system at a 0.75 mL min^{-1} injection rate in three steps. After each ~7 mL injection step, the mixture was allowed to react for 60 minutes. A portion of core/shell UCNPs would be extracted before the next injection step: 15.6 mL after the first injection step for core/shell UCNPs with a 1.9 nm-thick shell and 19.2 mL after the second injection step for core/shell UCNPs with a 3.5 nm-thick shell. Extractions were allowed to cool down to room temperature before transfer from glass syringe to Falcon centrifuge tube for subsequent washing. After the final injection step and a total of 180 minutes of reaction, the mixture (core/shell UCNPs with a 5.6 nm-thick shell) was cooled to room temperature under argon gas and magnetic stirring. All core/shell UCNPs were precipitated with ethanol and washed three times with hexane/acetone (1/4 v/v in each case), followed by centrifugation (with 5400 x g). Finally, all UCNPs were re-dispersed in hexane for further structural and optical characterization.

Structural characterization

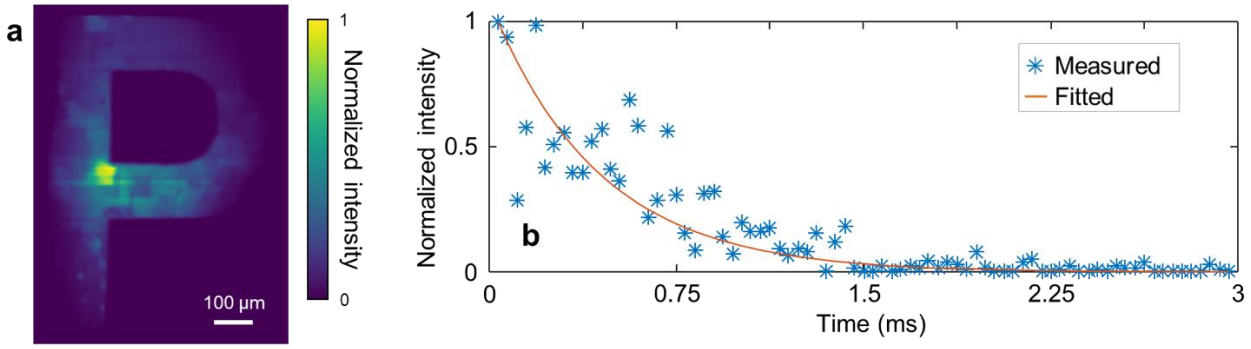
The morphology and size distribution of the core/shell UCNPs were investigated by transmission electron microscopy (TEM, Philips, Tecnai 12). The particle size was determined from TEM images using ImageJ software with a minimum set size of 280 individual UCNPs per sample. The results are shown in Figure 2a. The crystallinity and phase of the core-only and core/shell UCNPs were determined via X-ray powder diffraction (XRD) analysis using a diffractometer (Bruker, D8 Advance) with CuKα radiation (Supplementary Fig. 4.3). The peaks in measured XRD spectra match the reference tabulated data (PDF# 01-080-8787). Along with the TEM images (i.e., Figure 4.2a), this result ensured that the fabricated UCNPs were of the hexagonal crystal phase.



Supplementary Figure 4.3 X-ray powder diffraction patterns of UCNPs. The core-only and core/shell NaGdF4:Er3+, Yb3+/NaGdF4 UCNPs following their growth by increasing the shell thickness. Red lines: Diffraction peaks of pure hexagonal NaGdF4 (data are taken from PDF# 01-080-8787).

Supplementary Note 6: Characterization of SPLIT's system sensitivity

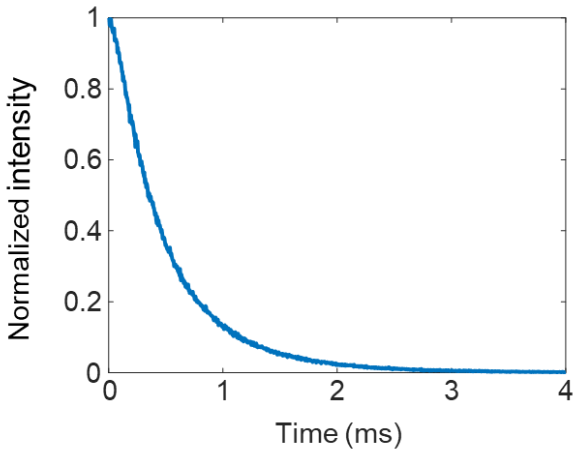
To test the sensitivity of SPLIT, we monitored the reconstructed image quality while decreasing the laser power. The detection sensitivity of the SPLIT system was characterized by imaging photoluminescence intensity decay with various excitation power densities (Supplementary Fig. 4.4). Transparency of the letter "P" covered the sample of UCNPs with a shell thickness of 5.6 nm. The laser power density was varied from 0.4 to 0.04 W mm^{-2}. All other experimental parameters, such as exposure time, camera gain, and temperature, were kept the same. The quality of reconstructed images kept degrading with decreased laser power density until partially losing spatial structure at <0.06 W mm^{-2}. In addition, lower SNRs in measurements deteriorate the image reconstruction, manifested by the increase in noise levels in the intensity decay curves and the deviation of the calculated photoluminescence lifetime from the correct values. Thus, the SPLIT's sensitivity under single-shot imaging for this UCNP sample was quantified to be 0.06 W mm^{-2}.



Supplementary Figure 4.4 Characterization sensitivity of the SPLIT system. **a** Temporally integrated reconstructed image at the excitation laser power density of 0.06 W mm-2. **b** Normalized intensity as a function of time with a fitting curve.

Supplementary Note 7: Measurement of lifetimes of UCNPs using the TCSPC technique

To ascertain our results, we used the standard TCSPC method (Edinburgh Instruments, FLS980, 70-μs excitation pulse) to measure photoluminescence decay of the 5.6 nm-thick-shell UCNPs dispersed in hexane. The measured intensity decay curve is shown in Supplementary Fig. 4.5. Lifetime values acquired from the SPLIT and TCSPC measurements yielded a 6.9% mismatch. This difference is attributed to different environments in which UCNPs were measured (dried powder for SPLIT and solution for TCSPC), different excitation pulse widths (50-μs for SPLIT and 70-μs for TCSPC), and different instrumental responses.

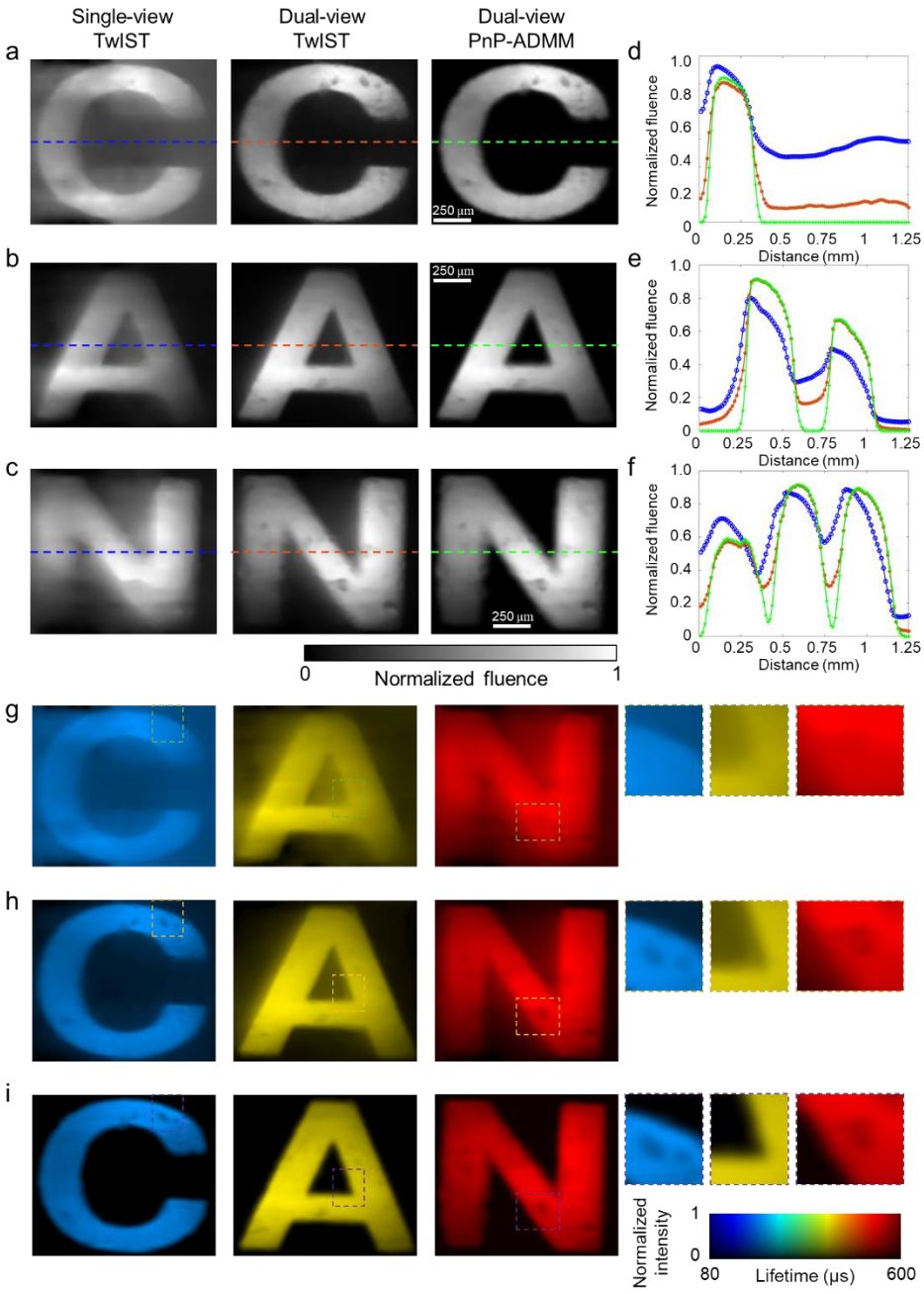


Supplementary Figure 4.5 Measurement of the green upconversion emission lifetime of the 5.6 nm-thick-shell UCNPs using the TCSPC method.

Supplementary Note 8: Comparison of reconstructed image quality

To quantitatively demonstrate the superiority of the dual-view PnP-ADMM algorithm employed in SPLIT's image reconstruction, we compared it with two other algorithms dominantly used in existing streak-camera-based single-shot ultrafast imaging—the single-view two-step iterative thresholding/shrinkage (TwIST) algorithm and the dual-view TwIST algorithm. Specifically, we used the experimental data of the green emission of UCNPs with shell thicknesses of 1.9 nm, 3.5 nm, and 5.6 nm, covered by transparencies of letters "C", "A", and "N", respectively. Both View 1 and View 2 were used for the dual-view TwIST algorithm and the dual-view PnP-ADMM algorithm. Only View 2 was used for the single-view TwIST algorithm. All the reconstructed datacubes had the same size. Supplementary Figs. 4.6a–c show the time-integrated images by projecting datacubes reconstructed by the three algorithms along the time axis. Among them, the result from the dual-view PnP-ADMM is duplicated from Figure 4.2f to better illustrate this comparison. We selected one line from each letter and compared their profiles in Supplementary Figs. 4.6d–f. From these results, the single-view TwIST algorithm gives the worst contrasts of 0.41 for "C", 0.88 for "A", and 0.76 for "N". Dual-view TwIST improves the contrast to 0.82 for "C", 0.99 for "A", and 0.93 for "N", respectively. Dual-view PnP-ADMM gives the best result—producing contrasts of 1 for all three cases.

The better quality in the reconstructed images translated to higher accuracy in lifetime quantification. Supplementary Figs. 4.6g–i show two-dimensional (2D) lifetime maps of these samples with zoom-in-views of three local areas. Both the single-view and dual-view TwIST algorithms yield artifacts, manifesting as false lifetime values on pixels in the background. In contrast, the dual-view PnP-ADMM algorithm eliminates these artifacts with a clean background. Meanwhile, in the selected local areas of letters "C" and "N" (insets in Supplementary Figs. 4.6g–i), single-view TwIST completely wipes out the features induced by the non-uniform distribution of the UCNPs. In contrast, both dual-view TwIST and dual-view PnP-ADMM algorithms preserve these features. Finally, benefitting from the superb denoising capability of the dual-view PnP-ADMM, the noise level in the intensity decay curves as a function of time reduces by 4.6× and 2.5× compared to those of the single-view TwIST and dual-view TwIST algorithms, which contributes to a more accurate lifetime calculation.



Supplementary Figure 4.6 Comparison of quality of images reconstructed by using different algorithms. **a** Letter “C” reconstructed by using the single-view TwIST, dual-view TwIST, and dual-view PnP-ADMM algorithms, respectively. **b**-**c** As (a), but for letters “A” and “N”. **d** Comparison of the selected line profiles of the reconstructed images of letter “C”. **e**-**f** As (d), but for letters “A” and “N”. **g**-**i** Lifetime maps of the three letters produced by the single-view TwIST (g), single-view PnP-ADMM (h), and dual-view PnP-ADMM (i) algorithms. Insets: zoom-in views of three local areas.

Supplementary Note 9: Determination of the absolute temperature sensitivities, relative temperature sensitivities, and thermal uncertainty

Both the absolute temperature sensitivity S_a and the constant c_t in Equation (4.3) are determined by using the curve fitting toolbox in Matlab. Using the data presented in Figure 4.3e with linear fitting, we quantified $S_a = -1.90\ \mu s\ °C^{-1}$ and $c_t = 278\ °C$ for the green emission and $S_a = -2.40\ \mu s\ °C^{-1}$ and $c_t = 210\ °C$ for the red emission.

Moreover, the relative temperature sensitivity can be calculated by [175]

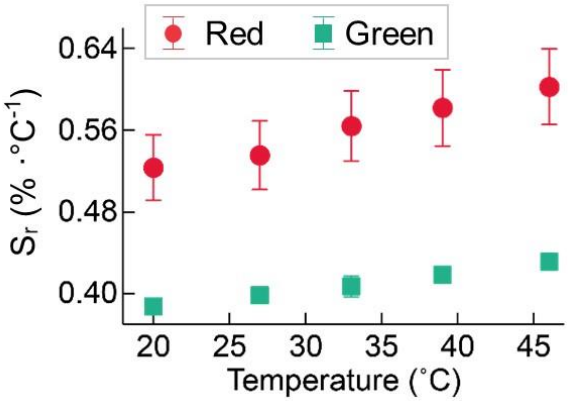
$$S_r = \frac{|S_a|}{\tau}. \tag{S4.14}$$

Using the data shown in Figure 3e, S_r in the pre-set temperature range were quantified to be 0.39–0.43%·ºC^{-1} for the green emission and 0.52–0.60%·ºC^{-1} for the red emission (Supplementary Fig. 4.7).

Finally, the thermal uncertainty [175] in SPLIT is calculated by

$$\delta T = \frac{1}{S_r} \times \frac{\delta\tau}{\tau}, \tag{S4.15}$$

where $\delta\tau$ represents the uncertainty in the measured lifetimes. Supplementary Equation (S4.15) shows that δT depends on both the UCNPs' performance (quantified by the relative sensitivity, S_r) and experimental setup (that limits the normalized fluctuation of lifetimes, $\frac{\delta\tau}{\tau}$). $\delta\tau$ was characterized by repeating measurements using the SPLIT system under the same experimental conditions. Specifically, using the sample of the 5.6 nm shell thickness UCNPs at 20 ºC, we repeated the 2D lifetime measurements 60 times using the excitation power density of 0.4 W mm-2 and 0.06 W mm-2, respectively. These measurements produced $\delta\tau$ of 1.4–2.7 μs for the green emission and 2.2–4.0 μs for the red emission, respectively. With known values of $|S_a|$ and by using Supplementary Equation (S4.15), SPLIT's thermal uncertainty was calculated to be 0.7–1.4 ºC for the green emission and 0.9–1.7 ºC for the red emission.



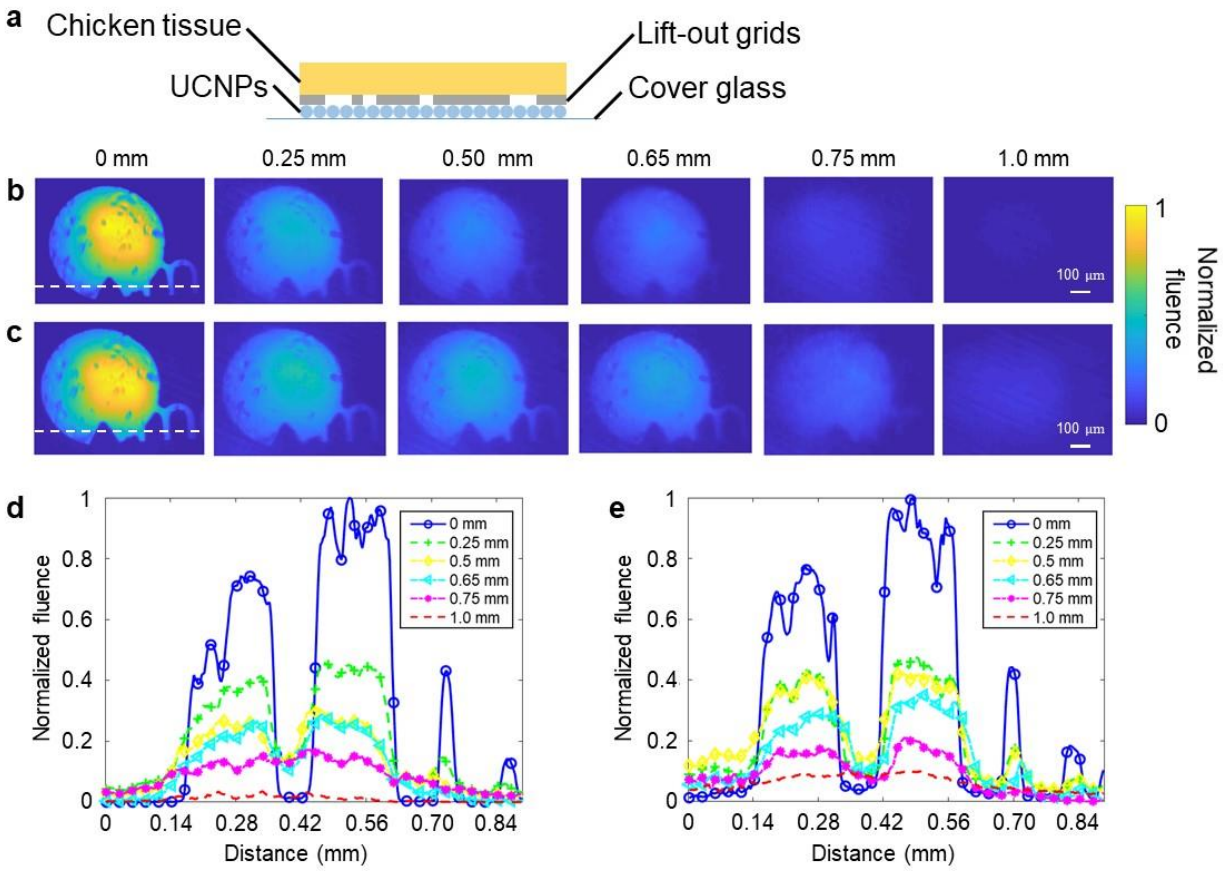
Supplementary Figure 4.7 Quantification of relative temperature sensitivities of the green and red emissions of the core/shell NaGdF4:Er3+,Yb3+/NaGdF4 UCNPs with a 5.6 nm-thick shell. Error bar: standard deviation.

Supplementary Note 10: Demonstration of SPLIT in biological environment

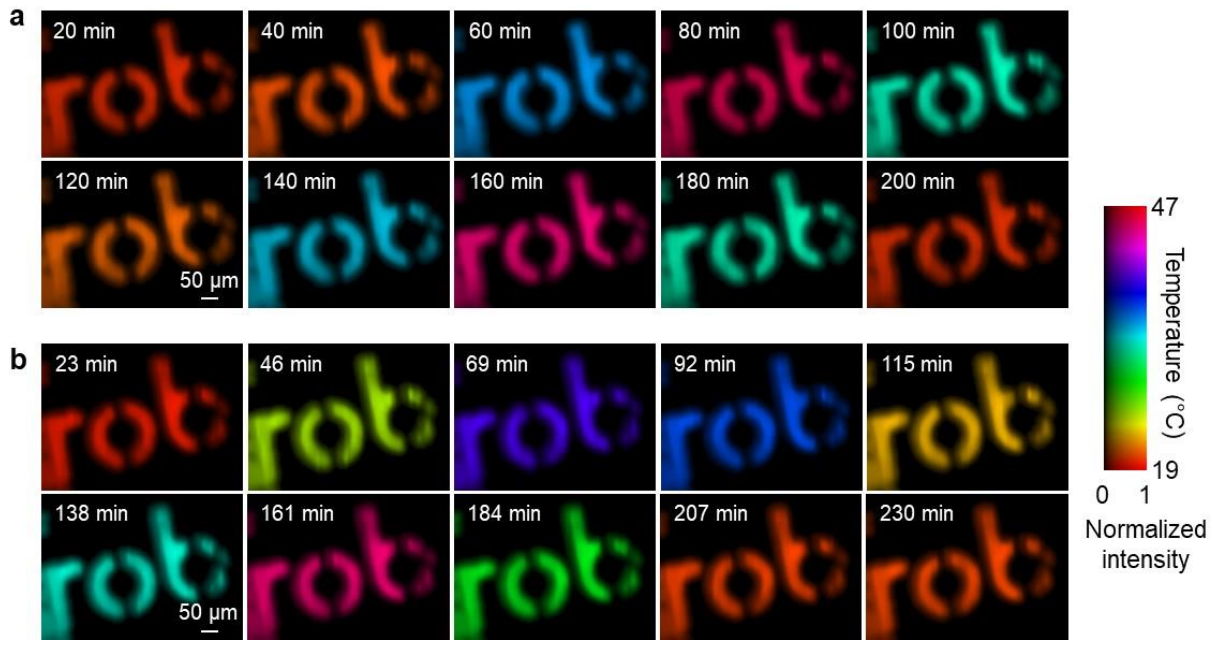
The UCNP sample with the shell thickness of 5.6 nm was covered by lift-out grids (Ted Pella, 460-2031-S), in which we chose the features of the letter “O” with a triangular shape on the bottom and the letter “m”. Then, five pieces of fresh chicken tissue with the thicknesses of 0.25, 0.5, 0.65, 0.75, 1.0 mm were used to cover the sample (Supplementary Fig. 4.8a). SPLIT captured the photoluminescence decay at 20 kfps. The reconstructed datacubes were projected to the $x-y$ plane (Supplementary Figs. 4.8b-c).

The image without chicken tissue, which is referred to as the thickness of “0 mm”, is also included for comparison. With the increased depth, the image intensity and contrast gradually approach zero. Supplementary Fig. 4.8d depicts the normalized fluence profiles across the white dashed line as shown in the first panel of Supplementary Fig. 4.8b. The experimental result was fitted using a single-component exponential function, which yielded a decay coefficient of 26 cm^{-1}. At the depth of 0.65 mm, the triangular feature and the letter “m” cannot be distinguished. Using a similar experimental procedure, we characterized SPLIT’s imaging depth for the red emission (Supplementary Figs. 4.8c and 4.8e). By using the single-component exponential fitting, the red emission had a decay coefficient of 18 cm^{-1}. The spatial features vanished at the depth of 0.75 mm. These results show that the red emission has, as expected, a greater imaging depth than the green upconversion counterpart. These results also show that good contrast can be maintained by using fresh chicken tissue of 0.5 mm thickness, which was selected for the longitudinal temperature monitoring experiments (Supplementary Fig. 4.9).

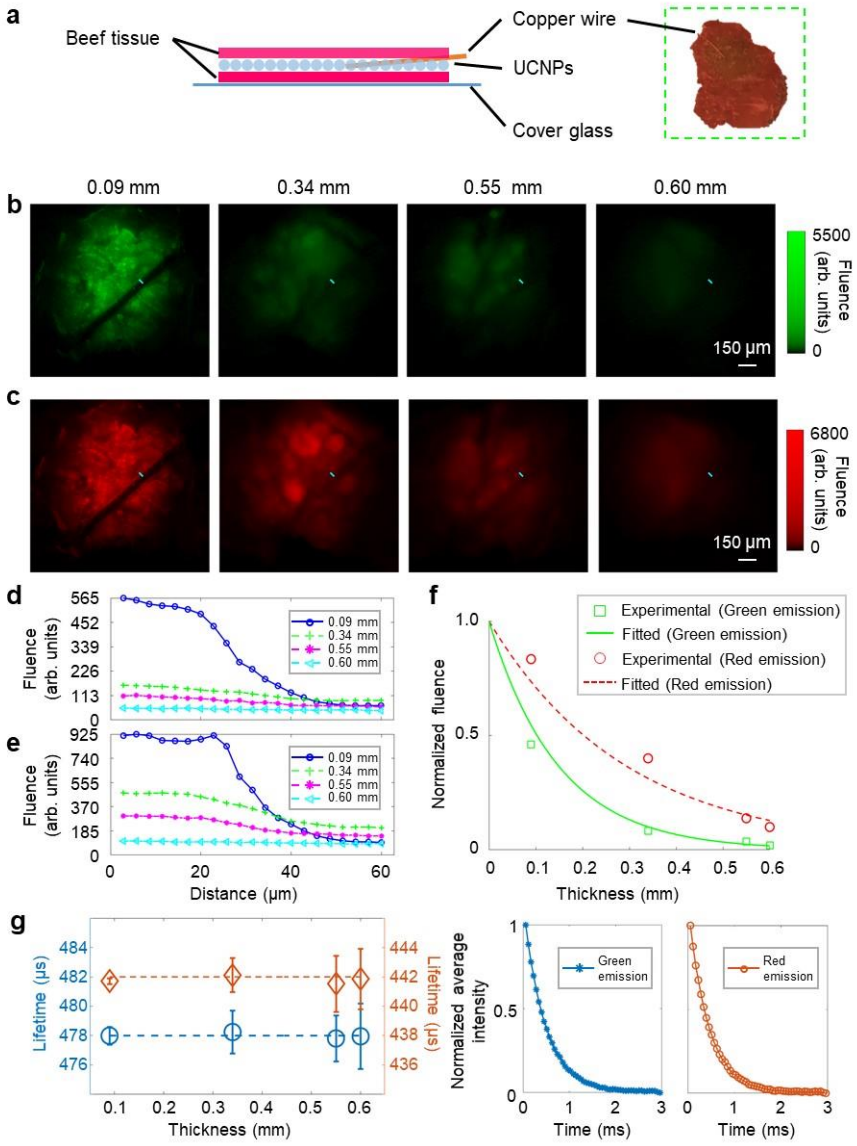
To test SPLIT using a scattering medium with the presence of both light scattering and absorption, the UCNPs with the shell thickness of 5.6 nm were injected into a piece of fresh beef tissue, where we also inserted a 90 μm-diameter copper wire at the depth of 0.09 mm as a spatial feature (Supplementary Fig. 4.10a). Myoglobin in the beef tissue, which has similar optical absorption properties to hemoglobin [176, 177], was used to mimic the absorption by blood. To evaluate the SPLIT's imaging ability at different depths, this phantom was covered by different additional fresh beef slices, so that the thicknesses from the surface to the copper wire were 0.09 mm, 0.34 mm, 0.55 mm, and 0.60 mm. SPLIT performed photoluminescence lifetime imaging at 20 kfps. For both the green and the red emissions, the reconstructed datacubes with the different beef thicknesses were projected temporally, as shown in Supplementary Figs. 4.10b–c. Furthermore, we plotted the profiles of a selected local edge feature of the inserted copper wire under the different thicknesses, as shown in Supplementary Figs. 4.10d–e. We calculated the contrast of these edge profiles. For the green emission, the values are 0.78, 0.27, 0.26, and 0.09 for the four selected curves. As for the red emission, these values are 0.80, 0.38, 0.33, and 0.09. Moreover, these experimental results were fitted by using single-component exponential functions, which yielded decay coefficients of 65 cm^{-1} for the green emission and 33 cm^{-1} for the red emission (Supplementary Fig. 4.10f), which are greater than their counterpart of the chicken tissue of 26 cm^{-1} and 18 cm^{-1}. Because of its longer wavelength, the red emission has weaker scattering and weaker absorption by the myoglobin, which led to deeper penetration over the green emission for both types of scattering media. Finally, we analyzed the photoluminescence lifetimes for different thicknesses, and the results are shown in Supplementary Fig. 4.10g. The measured photoluminescence lifetimes for both emissions do not depend on the tissue thickness and hence excitation power density under the experimental conditions of our work. Lower excitation intensity, however, reduced the SNRs in the captured snapshots, which transfers to a larger standard deviation.



Supplementary Figure 4.8 Demonstration of SPLIT with a fresh chicken tissue phantom. **a** Experimental setup. **b** Temporally projected images of the reconstructed dynamic scene at the depths from 0 to 1 mm with the green emission. **c** As (b), but for the red emission. **d** Comparison of normalized fluence of a representative cross-section [marked by the white dashed line in the first panel in (b)] for various imaging depths. **e** As (d), but for the red emission. The representative cross-section is marked by the white dashed line in the first panel in (c).



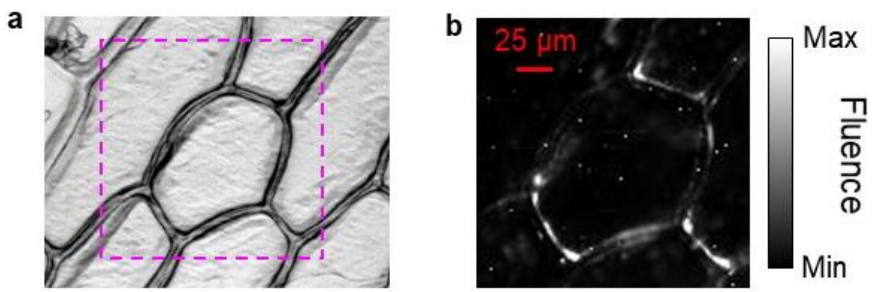
Supplementary Figure 4.9 Longitudinal temperature monitoring using green (a) and red (b) luminescence emissions from the 5.6 nm-thick UCNPs covered by a transmissive mask of letters “rob”.



Supplementary Figure 4.10 Demonstration of SPLIT with a fresh beef tissue phantom. **a** Sample preparation. **b** Temporally projected images of the reconstructed dynamic scenes at the depths from 0.09 to 0.60 mm for the green emission. **c** As (b), but for the red emission. **d**–**e** Cross-sections of a selected spatial feature [marked by the light blue solid line in (b) and (c)] for various depths for the green emission (d) and the red emission (e). **f** Normalized fluence versus tissue thickness for the green and red emissions with single-component exponential fitting. **g** Lifetimes as the function of the thickness for the green emission (blue circles; the mean value is plotted as the blue dashed line) and the red emission (orange diamonds; the mean value is plotted as the orange dashed line). Error bar: standard deviation. Right insets show the decay of normalized average intensity at the depth of 0.09 mm for the green and red emissions, respectively.

Supplementary Note 11: Preparation of the single-layer onion cells doped with UCNPs

For the onion cell experiments, UCNPs with a 5.6 nm-thick shell were first transferred to water via ligand exchange with citrate molecules. In a typical procedure, citrate-coated UCNPs were prepared by mixing 50 mg of oleate-capped UCNPs dispersed in 25 mL of hexane and 25 mL of 0.2 M trisodium citrate (99%; Alfa Aesar) solution (pH 3-4) under vigorous stirring for 3 hours. The two-phase (aqueous/organic) mixture was then poured into the separatory funnel, and the aqueous phase containing the UCNPs was isolated. The UCNPs were precipitated with acetone (1/3 v/v) via centrifugation (5400 x g) for 30 minutes. The obtained pellet was re-dispersed in 25 mL of 0.2 M trisodium citrate solution (pH 7-8) and left under stirring for an additional 2 hours. UCNPs were then precipitated with acetone (1/3 v/v) via centrifugation (5400 x g) for 30 minutes and washed twice with a mixture of water/acetone (1/3 v/v). The citrate-coated UCNPs were re-dispersed in distilled water. The yellow household onion was used to peel single-layer sheets of onion cells, which were incubated in a solution of citrate-coated UCNPs (3 mg mL^{-1}) for 24 hours. After the incubation, single-layer onion cells were rinsed in distilled water and dried by gently tapping with a soft tissue paper, before being placed onto microscope slides for subsequent imaging experiments. Before lifetime imaging, the presence of UCNPs in single-layer onion cells was confirmed (Supplementary Fig. 4.11a) with a bright-field microscope (Nikon, ECLIPSE Ti-S). In addition, a reference photoluminescence intensity image was taken by a custom-built confocal imaging platform (Photon Etc.), equipped with pulsed femtosecond Ti: Sapphire laser (Spectra-Physics, Mai Tai DeepSee). Samples were excited and imaged epi-fluorescently through a 20×/0.40 NA objective lens (Nikon, CFI60 TU Plan Epi ELWD). Photoluminescence intensity was recorded by a low-noise CCD camera (Princeton Instruments, Pixis100). The upconversion emission images of static onion cells (Supplementary Fig. 4.11b) were obtained through raster scanning a 120×120 pixel map, each of which has the size of 2 μm and the integration time 0.2 seconds per pixel. The total time to form one lifetime map was 48 minutes.



Supplementary Figure 4.11 Single-layer onion cell sample. **a** Image of the sample taken by a bright-field microscope. **b** Confocal microscopy of green upconversion emission of UCNPs diffused in an individual onion cell [marked by the magenta dashed box in (a)].

Supplementary Note 12: Comparison between SPLIT and previous imaging modalities for 2D lifetime imaging

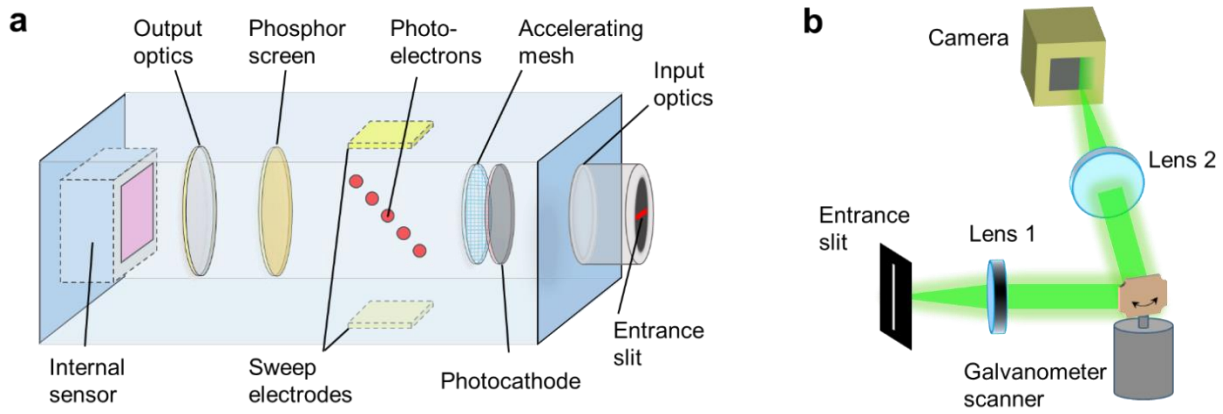
To articulate the difference between SPLIT and previous works on ultrafast imaging that used streak cameras, we summarize their technical specifications and applications in Supplementary Table 4.1. To explain the details included in this table, we first detail the working principles of streak cameras and compressed ultrafast photography (CUP); then, we summarize technical specifications and applications of the existing imaging modalities.

Streak cameras for wide-field lifetime imaging

Streak cameras are highly suitable for 2D lifetime imaging. In its conventional operation, the field of view (FOV) of streak cameras is limited by an entrance slit with typical widths of 50–100 μm. A sweeping unit deflects the time-of-arrival of the incident light signal along the axis perpendicular to the device's entrance slit. Depending on the mechanisms of the sweeping unit, streak cameras can be generally categorized into optoelectronic and mechanical types. In optoelectronic streak cameras (Supplementary Fig. 4.12a), incident photons are first converted to photoelectrons by a photocathode. After acceleration, these photoelectrons are deflected by a time-varying voltage applied on a pair of sweep electrodes. Then, these photoelectrons are converted back to photons on a phosphor screen. Finally, the optical signal is imaged to an internal sensor. The optoelectronic streak camera can achieve a temporal resolution of up to 100 fs. Because of this ultrafast imaging ability, optoelectronic streak cameras have been used for imaging the emission of fluorescence that has lifetimes in the order of picoseconds and nanoseconds [4, 19, 20, 178, 179]. However, due to the photon-to-photoelectron conversion by the photocathode, the quantum efficiency (QE) of the optoelectronic streak cameras is typically <15% for visible light. Besides, the space-charge effect in the electrostatic lens system imposes constraints in the spatial resolution (typically tens to hundreds of micrometers) and the dynamic range (e.g., <10 for certain femtosecond streak cameras). Both weaknesses severely limit the quality of acquired data.

Unlike optoelectronic streak cameras, a mechanical streak camera (Supplementary Fig. 4.12b) usually uses a rotating mirror (e.g., a GS or a polygon mirror) to deflect the light. Since the

mechanical sweeping is much slower than the optoelectronic counterpart, this type of streak camera has tunable temporal resolutions typically from hundreds of nanoseconds to microseconds, which makes them highly suitable for lifetime imaging of luminescence processes on the order of microseconds and milliseconds, such as phosphorescence and parity forbidden 4f-4f transitions in lanthanide ions [180]. Moreover, its all-optical data acquisition allows flexibly implementing many high-sensitivity cameras [e.g., electron-multiplying (EM) CCD and scientific CMOS cameras, whose QEs can be >90% for visible light] to obtain superior SNRs in measurements. The all-optical operation also avoids the space-charge effect, which enables optics-limited spatial resolution and high dynamic range (e.g., >60,000 of the EMCCD camera used in this work). Finally, the mechanical streak camera is considerably more cost-efficient than the optoelectronic streak camera. Therefore, mechanical streak cameras are perfectly suitable for imaging microsecond-level emission from UCNPs.



Supplementary Figure 4.12 Schematics of an optoelectronic streak camera (a) and a mechanical streak camera (b) in their conventional operations.

Single-shot compressed temporal imaging for fast 2D lifetime mapping

Single-shot compressed temporal imaging is a novel computational imaging concept that enables 2D lifetime mapping in one acquisition. In the conventional operation of the streak camera, the entrance slit limits the imaging FOV to be one-dimensional (1D). To lift this limitation, compressed-sensing paradigms have been implemented with optoelectronic streak cameras. The resulted CUP technique [4, 19, 20, 178, 179] allows complete opening of the entrance slit for 2D ultrafast imaging in a single shot. CUP and its variants have been applied to single-shot fluorescence

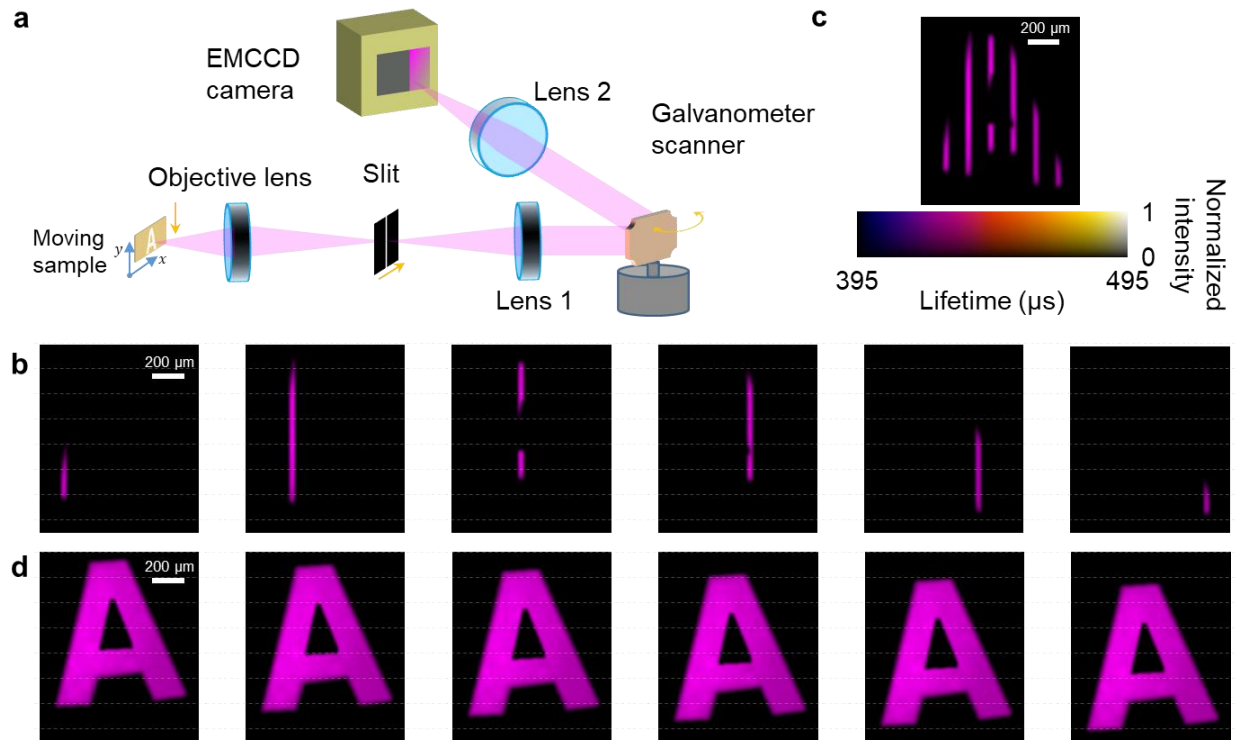
lifetime imaging [4, 19, 178]. In contrast, to our knowledge, single-shot compressed temporal imaging has not yet been applied to 2D imaging of microsecond-to-millisecond scale lifetimes, like those of UCNP emission. SPLIT thus marks the first technique in this category. It is also the first demonstration of single-shot photoluminescence lifetime-based temperature mapping in a 2D FOV. Compared to conventional line-scanning counterpart [181], SPLIT has considerable advantages in light throughput and sample choices.

Performance comparison between SPLIT and line-scanning lifetime imaging

To experimentally demonstrate the advantages of SPLIT to the line-scanning confocal setup [181], we imaged a moving photoluminescent sample (Supplementary Fig. 4.13). The major experimental parameters (e.g., magnification ratio, camera's exposure time, and camera's frame rate) were kept the same as those of SPLIT. The UCNPs with the shell thickness of 5.6 nm were covered by a piece of transparency of letter "A". Loaded onto a translation stage, this sample moved downward at a speed of 0.8 mm s^{-1}. To perform line scanning, we placed a 200-μm-wide slit at the intermediate image plane (i.e., equivalently 50-μm-wide at the sample plane) to limit the FOV to 1D (Supplementary Fig. 4.13a). Attached to another translation stage, the slit was scanned in the x direction at a speed of 2.8 mm s^{-1}. Using the green emission, this line-scanning confocal setup generated six 1D lifetime maps (Supplementary Fig. 4.13b). After stitching these results together, we obtained a 2D lifetime map as shown in Supplementary Fig. 4.13c. However, the stitched result inevitably suffers from the loss of spatial content due to the dark time between adjacent camera exposures. In the meantime, the map is distorted in the vertical direction due to the sample's movement, which proves the incapability of line-scanning-based techniques in measuring dynamic photoluminescent objects. As a comparison, we used SPLIT to image this sample under the same experimental conditions. Because of its single-shot imaging ability, SPLIT produced six 2D lifetime maps (Supplementary Fig. 4.13d). No image produced by the SPLIT system has any loss of spatial content or distortion. The results also clearly illustrate the downward movement of the letter "A". Therefore, SPLIT has unique advantages over the conventional scanning-based lifetime measurement in data throughput, measurement accuracy, and application scope.

It is also worth pointing out that from the perspective of optical instrumentation, SPLIT provides high-sensitivity cameras with ultrahigh imaging speeds in 2D FOV. In this regard, besides the single-shot wide-field photoluminescent lifetime mapping demonstrated in this work,

the SPLIT system offers a generic imaging platform for many other studies. Potential future applications include optical voltage imaging of action potentials in neurons and high-throughput flow cytometry.

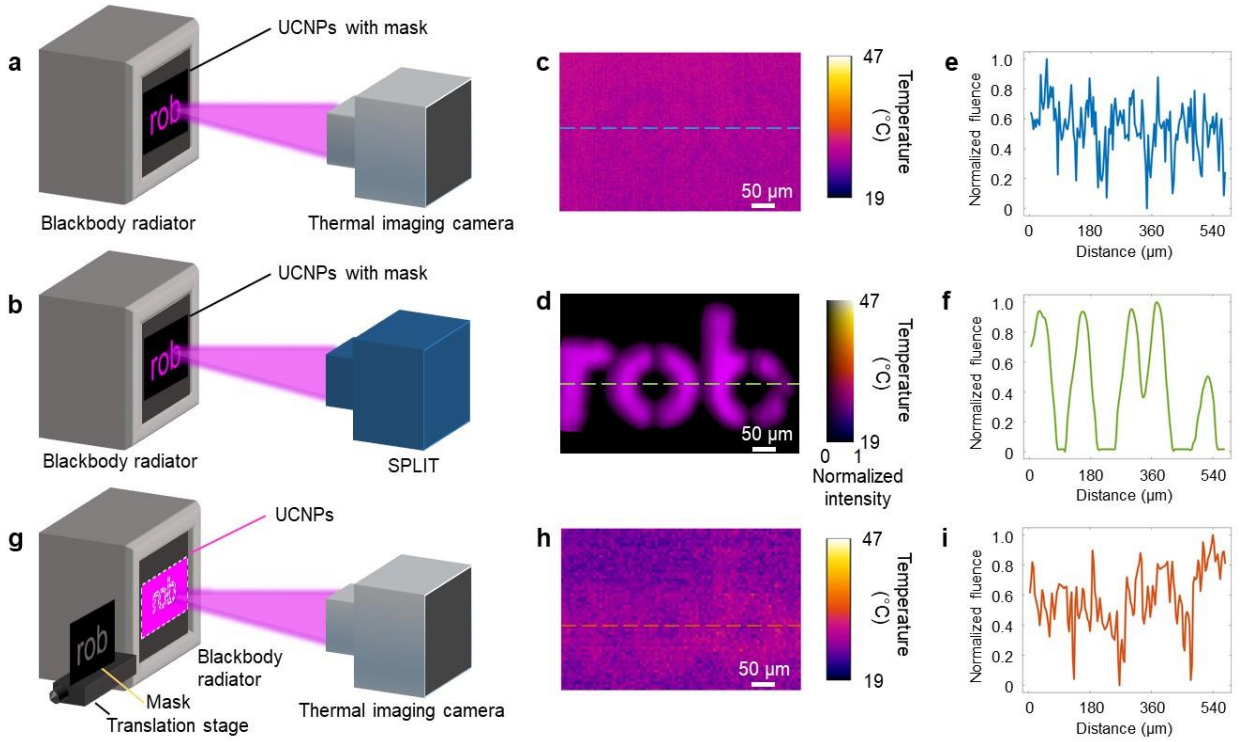


Supplementary Figure 4.13 Comparison between line-scanning microscopy and SPLIT in 2D PLI capability. **a** Experimental setup of line-scanning microscopy. The moving UCNPs sample was loaded onto a translation stage. The moving directions of the sample and the slit are marked by orange arrows. **b** 1D photoluminescence lifetime images produced by using the line-scanning setup. **c** Distorted partial 2D lifetime map synthesized by using the data in (b). **d** Six 2D lifetime maps of the sample moving downward captured by using the SPLIT system.

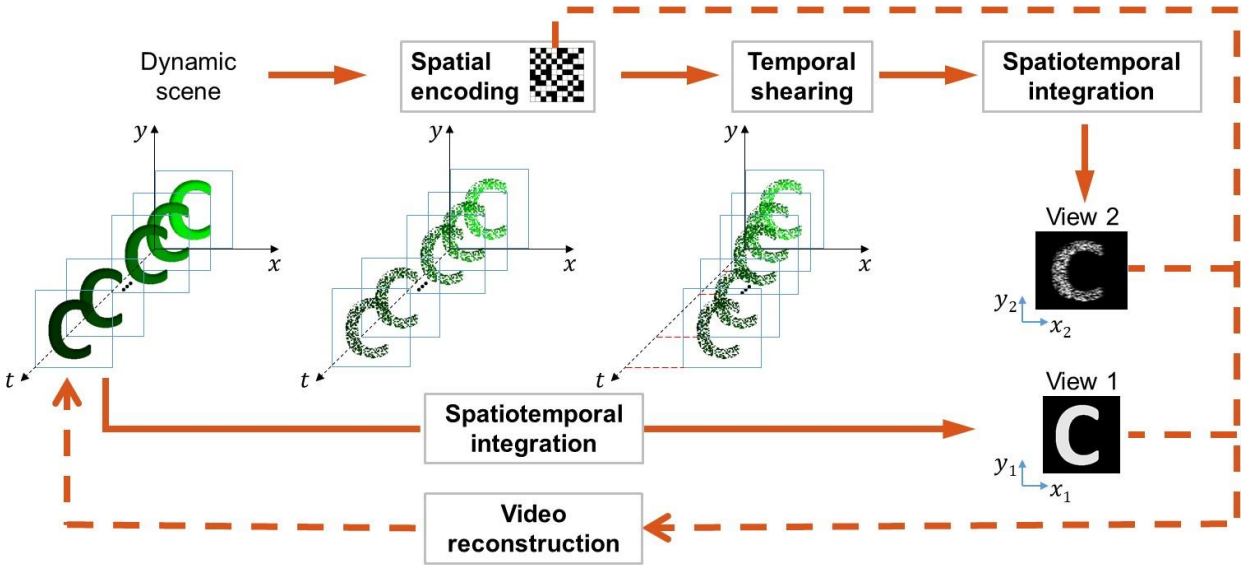
Supplementary Note 13: Comparison between SPLIT and thermal imaging

We used a thermal imaging camera (Yoseen, X384D) (Supplementary Fig. 4.14a) and SPLIT (Supplementary Fig. 4.14b) to image UCNPs covered by a metal mask of letters “rob” in lift-out grids (Ted Pella, 460-2031-S). Akin to the SPLIT system, a 4× magnification ratio was used for the thermal imaging camera. A blackbody radiator (Yoseen, YSHT-35) was used to heat this sample to 27 ºC. The images produced by these two methods are shown in Supplementary Figs. 14c–d and the selected line profiles are shown in Supplementary Figs. 4.14e–f. The edge contrast of the imaged letters using the thermal imaging camera is much worse than that using SPLIT. Moreover, the thermal imaging result presents strong background due to the same temperature of the mask, whereas SPLIT keeps a clean background thanks to its optical sensing ability.

In another experiment, we loaded the metal mask on a translation stage. The mask was kept out of the FOV to keep its temperature at 18 °C (i.e., the room temperature in our laboratory). The UCNPs were still heated up by the blackbody radiator to 27 °C. The mask was quickly moved into the FOV, and the thermal imaging camera captured the images immediately (Supplementary Fig. 4.14g). The thermal image and the selected line profiles are shown in Supplementary Figs. 4.14h–i. Despite the slight improvement in contrast compared to Supplementary Figs. 4.14c and 4.14e, the image quality is still incomparable to the results produced by the SPLIT system (Supplementary Figs. 4.14d and 4.14f). Thus, compared to a thermal imaging camera, SPLIT supplies superior temperature mapping capability.



Supplementary Figure 4.14 Comparison between the thermal imaging camera and SPLIT in temperature imaging. **a**–**b** Experimental setup using thermal imaging camera (a) or SPLIT (b). The sample and mask were heated up by a blackbody radiator. **c** Temperature image captured by using the thermal imaging camera. **d** As (c), but using SPLIT. **e**–**f** Selected line profile from (c) and (d), respectively. **g** As (a), but using a translation stage to move the mask with the room temperature. **h** Temperature image captured by the setup in (g). i Selected line profile from (h).



Supplementary Figure 4.15 Illustration of the working principle of SPLIT.

Supplementary Table

Supplementary Table 4.1 Comparison of representative 2D lifetime imaging modalities using streak cameras.

	CUP [21]	LLE-CUP [22]	CUSP [24]	Sequential 1D ultrafast photography [20]	Sequential 1D ultrahigh-speed photography [25]	SPLIT
Application	Macroscopic FLI	High-speed FLIM [23]	Spectrally resolved FLIM	FLIM [26]	PpLIM	Video-rate PLI thermometry
Imaging scheme	Compressed streak imaging	Compressed streak imaging	Compressed streak imaging	Line-scanning	Line-scanning	Compressed streak imaging
Single-shot 2D mapping	Yes	Yes	Yes	No	No	Yes
Throughput	High	High	High	Low	Low	High
Inter-frame interval	20 ps	40 – 500 ps	2 ps	~6 ps	5.8 ns – 4.7 μs	30 – 250 μs
Recording time window	~5 ns	20 – 100 ns	0.8 ns	6 ns	52 ns – ~1 ms	3 ms
Spatial resolution	1.7 mm	≥ 0.2 μm	50 μm	~0.7 μm	~0.4 μm	20 μm
Applicability to dynamic samples	Yes	Yes	Yes	No	No	Yes
Type of streak camera	Optoelectronic	Optoelectronic	Optoelectronic	Optoelectronic	Mechanical	Mechanical
Cost	High	High	High	High	Low	Low
Size	Big	Big	Big	Big	Small	Small
Quantum efficiency (QE)	Moderate	Moderate	Moderate	Moderate	High	High
Power consumption	High	High	High	High	Low	Low

Note: CUP, compressed ultrafast photography; CUSP, compressed ultrafast spectral photography; FLIM, fluorescence lifetime imaging; LLE, lossless-encoding; PpLIM, phosphorescence lifetime imaging microscopy.

5 DEEP LEARNING ASSISTED CS RECONSTRUCTION

Single-shot Real-time Compressed Ultrahigh-speed Imaging Enabled by a Snapshot-to-video Autoencoder

Imagerie ultra-haute vitesse compressée en temps réel unique activée par un encodeur automatique instantané-vidéo

Authors:

Xianglei Liu[1,3], João Monteiro[1,3], Isabela Albuquerque[1], Yingming Lai[1], Cheng Jiang[1], Shian Zhang[2], Tiago h. Falk[1,4], Jinyang Liang[1,5]

[1]Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650 boulevard Lionel-Boulet, Varennes, Québec J3X1S2, CANADA

[2]State Key Laboratory of Precision Spectroscopy, East China Normal University, 3663 North Zhongshan Road, Shanghai, 200062, PR CHINA

[3]These authors contributed equally to this work.

[4]Corresponding author: falk@emt.inrs.ca

[5]Corresponding author: jinyang.liang@emt.inrs.ca

Publication:

Photonics Research

Vol. 9, Issue 12, pp. 2464-2474 (2021)

https://doi.org/10.1364/PRJ.422179

Contribution of authors: X.L built the system, conducted the experiments, and analyzed the data. J.M. and I.A. developed the reconstruction algorithm, the model architecture, and the training scheme. Y.L. and C.J. prepared the sample and performed some of the experiments. S.Z. and T.H.F. initiated the project. J.L. initiated the project, proposed the concept, contributed to experimental design, and supervised the project. All authors wrote and revised the manuscript.

Abstract

Single-shot two-dimensional (2D) optical imaging of transient scenes is indispensable for numerous areas of study. Among existing techniques, compressed optical-streaking ultrahigh-speed photography (COSUP) uses a cost-efficient design to endow ultra-high frame rates with off-the-shelf CCD and CMOS cameras. Thus far, COSUP's application scope is limited by the long processing time and unstable image quality in existing analytical-modeling-based video reconstruction. To overcome these problems, we have developed a snapshot-to-video autoencoder (S2V-AE)—a new deep neural network that maps a compressively recorded 2D image to a movie. The S2V-AE preserves spatiotemporal coherence in reconstructed videos and presents a flexible structure to tolerate changes in input data. Implemented in compressed ultrahigh-speed imaging, the S2V-AE enables the development of single-shot machine-learning assisted real-time (SMART) COSUP, which features a reconstruction time of 60 ms and a large sequence depth of 100 frames. SMART-COSUP is applied to wide-field multiple-particle tracking at 20 thousand frames-per-second. As a universal computational framework, the S2V-AE is readily adaptable to other modalities in high-dimensional compressed sensing. SMART-COSUP is also expected to find wide applications in applied and fundamental sciences.

5.1 Introduction

Two-dimensional (2D) optical visualization of transient phenomena in the actual time of the event's occurrence plays a vital role to understand many mechanisms in biology, physics, and chemistry [182-184]. To discern spatiotemporal details in these phenomena, high-speed optical imagers are indispensable. Imaging speeds of these systems, usually determined by the frame rates of deployed CCD or CMOS cameras, can be further increased using novel sensor designs [185-187], new readout interfaces [188, 189], and advanced computational imaging methods [190-193].

Among existing approaches, compressed ultrafast photography (CUP) [96, 194-200] is an innovative coded-aperture imaging scheme [17, 201] that integrates video compressed sensing [202] into streak imaging [203]. In data acquisition, a spatiotemporal (x, y, t) scene is compressively recorded by optical imaging hardware to a 2D snapshot. The ensuing reconstruction recovers the datacube of the scene computationally. Despite initially demonstrated using a streak camera, the concept of CUP was soon implemented in CCD and CMOS cameras in compressed optical-streaking ultrahigh-speed photography (COSUP) [204]. Compared to other single-shot ultrahigh-speed imaging modalities [205-208], COSUP is not bounded by the moving speed of piezo-stages [205, 206] or the refreshing rate of spatial light modulators [207, 208]. Exhibiting as a cost-efficient system, COSUP has demonstrated single-shot transient imaging ability with a tunable imaging speed of up to 1.5 million frames-per-second (fps) based on an off-the-shelf CMOS camera with an intrinsic frame rate of tens of hertz.

Despite these hardware innovations, COSUP's video reconstruction has ample room for improvement. Existing reconstruction frameworks can be generally grouped into analytical-modeling-based methods and machine-learning-based methods [209]. Using the prior knowledge of the sensing matrix and the sparsity in the transient scene, the analytical-modeling-based methods reconstruct videos by solving an optimization problem that synthetically considers the image fidelity and the sparsity-promoting regularization. However, demonstrated methods, such as the two-step iterative shrinkage/thresholding (TwIST) algorithm [210], augmented Lagrangian algorithm [211], and alternating direction method of multiplier (ADMM) algorithm [208], undergo time-consuming processing that uses tens to hundreds of iterations. The excessively long reconstruction time strains these analytical-modeling-based methods from real-time (i.e., ≥16 Hz [212]) reconstruction, which excludes COSUP's application scope from tasks that need on-time feedback [213]. Moreover, the reconstructed video quality highly depends on the accuracy of prior knowledge and the empirical tuning of parameters.

To solve these problems, machine learning has become an increasingly popular choice. Instead of relying solely on prior knowledge, large amounts of training data are used for deep neural networks (DNNs) [214] to learn how to map an acquired snapshot back to a video. Upon the completion of training, DNNs can then execute non-iterative high-quality reconstruction during runtime. Thus far, DNNs that employ the architectures of the multi-layer perceptrons (MLPs) [215, 216] and the U-net [54, 74, 217, 218] have shown promise for compressed video reconstruction. Nonetheless, MLPs, with fully connected structures, scale linearly with the dimensionality of input data [219]. Besides, the decomposition in the reconstruction process presumes that all information in the output video block is contained in a patch of the input image, which cannot always be satisfied [215, 216]. As for the U-net, the reconstruction often starts with a pseudo-inverse operation to the input snapshot to accommodate the equality in dimensionality required by the original form of this network [220]. This initial step increases the reconstruction burden in computational time and memory. Moreover, akin to MPLs, U-net-based methods require slicing input data for reconstruction, which could cause the loss of spatial coherence [54]. Finally, inherent temporal coherence across video frames is often unconsidered in the U-net [221]. Because of these intrinsic limitations, videos reconstructed by the U-nets are often subject to spatiotemporal artifacts and a shallow sequence depth (i.e., the number of frames in the reconstructed video) [218].

Here, we propose to overcome these limitations with the use of an autoencoder (AE), whose objective is to learn a mapping from high-dimensional input data to a lower-dimensional representation space, from which the original data is recovered [222]. The implementation of convolutional layers in AE's architecture provides a parameter-sharing scheme that is more efficient than MLPs. Besides, without relying on locality presumptions, deep AEs with convolutional layers can preserve the intrinsic coherence in information content. Furthermore, recent advances in combining AE with adversarial formulations [223] have allowed replacing the loss functions based on pixel-wise error calculation to settings where perceptual features are accounted for, which have enabled more accurate capture of data distribution and increased visual fidelity [224]. In the particular case of training generative models [e.g., generative adversarial networks (GANs)] for natural scenes, recent advances have improved the reconstructed imaging quality by dividing the overall task into sub-problems, such as independent modeling of foreground and background [225], separated learning of motion and frame content [226], and conditioning generation on optical flows [227]. Despite these advances, with popular applications in audio signal enhancement [228] and pattern recognition [229], AEs have been mainly applied to one-dimensional and 2D reconstruction problems [230, 231]. Thus, existing

architectures of AEs cannot be readily implemented for video reconstruction in compressed ultrahigh-speed imaging.

To surmount these problems, we have developed a snapshot-to-video autoencoder (S2V-AE)—a new DNN that directly maps a compressively recorded 2D (x, y) snapshot to a 3D (x, y, t) video. This new architecture splits up the reconstruction process into two sub-tasks, each of which is trained individually to obtain superior quality in reconstructed videos. Implemented in compressed ultrahigh-speed imaging, such a video reconstruction framework enables developing single-shot machine-learning-assisted real-time (SMART) COSUP, which is applied to tracking multiple fast-moving particles in a wide field at 20 thousand fps (kfps).

5.2 Principle of SMART-COSUP

The schematic of the SMART-COSUP system is shown in Fig. 5.1(a). Its operating principle contains single-shot data acquisition and real-time video reconstruction [Fig. 5.1(b)]. A dynamic scene, $I(x, y, t)$, is imaged by front optics onto a printed pseudo-random binary transmissive mask (Fineline Imaging) with encoding pixels of 25 μm × 25 μm in size. This spatial modulation operation is denoted by the operator $\mathbf{C}$. The intensity distribution right after the encoding mask is expressed as

$$I_{\mathrm{c}}(x, y, t) = \sum_{j,k} I\left(\frac{x}{M_{\mathrm{f}}}, \frac{y}{M_{\mathrm{f}}}, t\right) C_{jk} \mathrm{rect}\left(\frac{x}{d_{\mathrm{e}}} - j, \frac{y}{d_{\mathrm{e}}} - k\right). \tag{5.1}$$

Here, M_{f} is the magnification of the front optics. C_{jk} denotes an element of a binary matrix representing the encoding pattern, j and k are matrix element indices. d_{e} is the size of encoding pixels on the mask. $\mathrm{rect}(\cdot)$ represents the rectangular function.

Subsequently, the spatially modulated scene is relayed by a 4*f* imaging system, consisting of a galvanometer scanner (GS, Cambridge Technology, 6220H) and two identical lenses (Lens 1 and Lens 2, Thorlabs, AC254-075-A). The GS is placed at the Fourier plane of this 4*f* imaging system to conduct optical shearing in the x-direction, denoted by the operator $\mathbf{S_o}$. The sheared image can be expressed as

$$I_{\mathrm{s}}(x, y, t) = I_{\mathrm{c}}(x - v_{\mathrm{s}} t, y, t), \tag{5.2}$$

where v_{s}, denoting SMART-COSUP's shearing velocity, is calculated by $v_{\mathrm{s}} = \alpha V_{\mathrm{g}} f_2 / t_{\mathrm{g}}$. Here, $V_{\mathrm{g}} =$0.16–0.64 V is the voltage added onto the GS. α is a constant that links V_{g} with GS's

deflection angle with the consideration of the input waveform. $f_2 = 75\ \mathrm{mm}$ is the focal length of Lens 2 in Fig. 5.1(a). $t_\mathrm{g} = 50$ ms is the period of the sinusoidal signal added to the galvanometer scanner.

Finally, the dynamic scene is spatiotemporally integrated by a CMOS camera (FLIR, GS3-U3-23S6M-C) to a 2D snapshot, denoted by the operator $\mathbf{T}$. The optical energy, $E(m,n)$, measured at pixel (m,n), is

$$E(m,n) = \iiint I_s(x,y,t)\mathrm{rect}\left(\frac{x}{d_\mathrm{c}} - m, \frac{y}{d_\mathrm{c}} - n\right) dxdydt. \tag{5.3}$$

Here, m and n are the pixel indices in the x and y axes on the camera. $d_\mathrm{c} = 5.86\ \mu\mathrm{m}$ is the CMOS sensor's pixel size. From Eqs. 5. (1)–(3), the forward model of SMART-COSUP is expressed by

$$E(m,n) = \mathbf{TS_oC}I(x,y,t). \tag{5.4}$$

In the ensuing real-time video reconstruction, the captured data is transferred to a workstation equipped with a graphic processing unit (NVIDIA, RTX Titan). The S2V-AE retrieves the datacube of the dynamic scene in 60 ms.

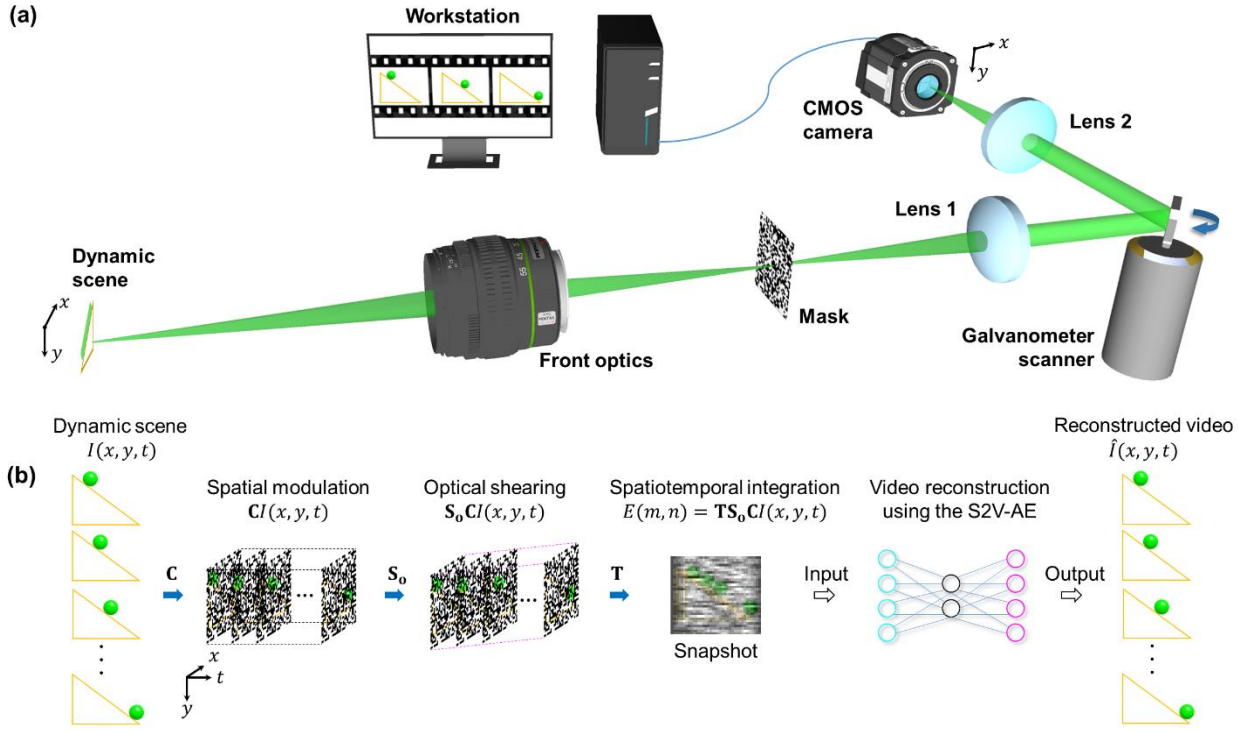


Figure 5.1 Single-shot machine-learning assisted real-time (SMART) compressed optical-streaking ultrahigh-speed photography (COSUP). (a) System schematic. (b) Operating principle. S2V-AE, snapshot-to-video autoencoder.

The frame rate of the SMART-COSUP system is derived from

$$r = \frac{v_s}{d_c}. \tag{5.5}$$

In this work, the reconstructed video has a frame rate of up to $r = 20\ \mathrm{kfps}$, a sequence depth of $N_t = 40 - 100$ frames, and a frame size of up to $N_x \times N_y = 256 \times 256$ pixels.

Compared to the previous hardware configuration [204], SMART-COSUP replaces the digital micromirror device (DMD), which functions as a 2D programmable blazed grating [55], with the transmissive mask for spatial modulation. This arrangement avoids generating a large number of unused diffraction orders, preventing a limited modulation efficiency to unblazed wavelengths, and eliminating intensity loss from the reflection from its cover glass as well as by its inter-pixel gap. In addition, the printed mask is illuminated at normal incidence, making it fully conjugated with both the object and the camera. Thus, the SMART-COSUP system presents a simpler, economical, and compact design with improved light throughput of the system and image quality of the captured snapshot.

5.3 Structure of S2V-AE

The architecture of S2V-AE consists of an encoder and a generator [Fig. 5.2(a)]. The encoder (denoted as $\boldsymbol{\mathcal{E}}$) converts a 2D snapshot to a series of low-dimensional latent vectors that represent particular features of the dynamic scene under study. As shown in Fig. 5.2(b), its architecture consists of five convolutional layers, a bi-directional long short-term memory (Bi-LSTM) recurrent layer [232], and a fully connected layer. In the convolutional layers, each convolution operation is followed by batch normalization (BN) [233] along with rectified linear unit (ReLU) activation [234]. The number of channels of feature maps, denoted by N, decreases from a preset value (512 in our experiments) to N_t. Then, the feature map, output by the last convolutional layer, is reshaped from a tensor into N_t vectors, all of which are fed into the Bi-LSTM recurrent blocks with the fully connected layer to model temporal coherence. The outputs of the encoder, referred as to latent vectors, are then input to the generator (denoted as $\boldsymbol{\mathcal{G}}$). In particular, each latent vector is reshaped back to a tensor, which is fed into the generator to reconstruct one frame in the video. As shown in Fig. 5.2(c), the architecture of the generator consists of seven transposed convolutional layers. Akin to the encoder, BN and ReLU activation are employed after each transposed convolution, whose preset number of channels decreases from 1024 to 1. Each latent

vector is processed by the generator to a frame of $N_x \times N_y$ in size. The composition of N_t such frames produces the reconstructed video.

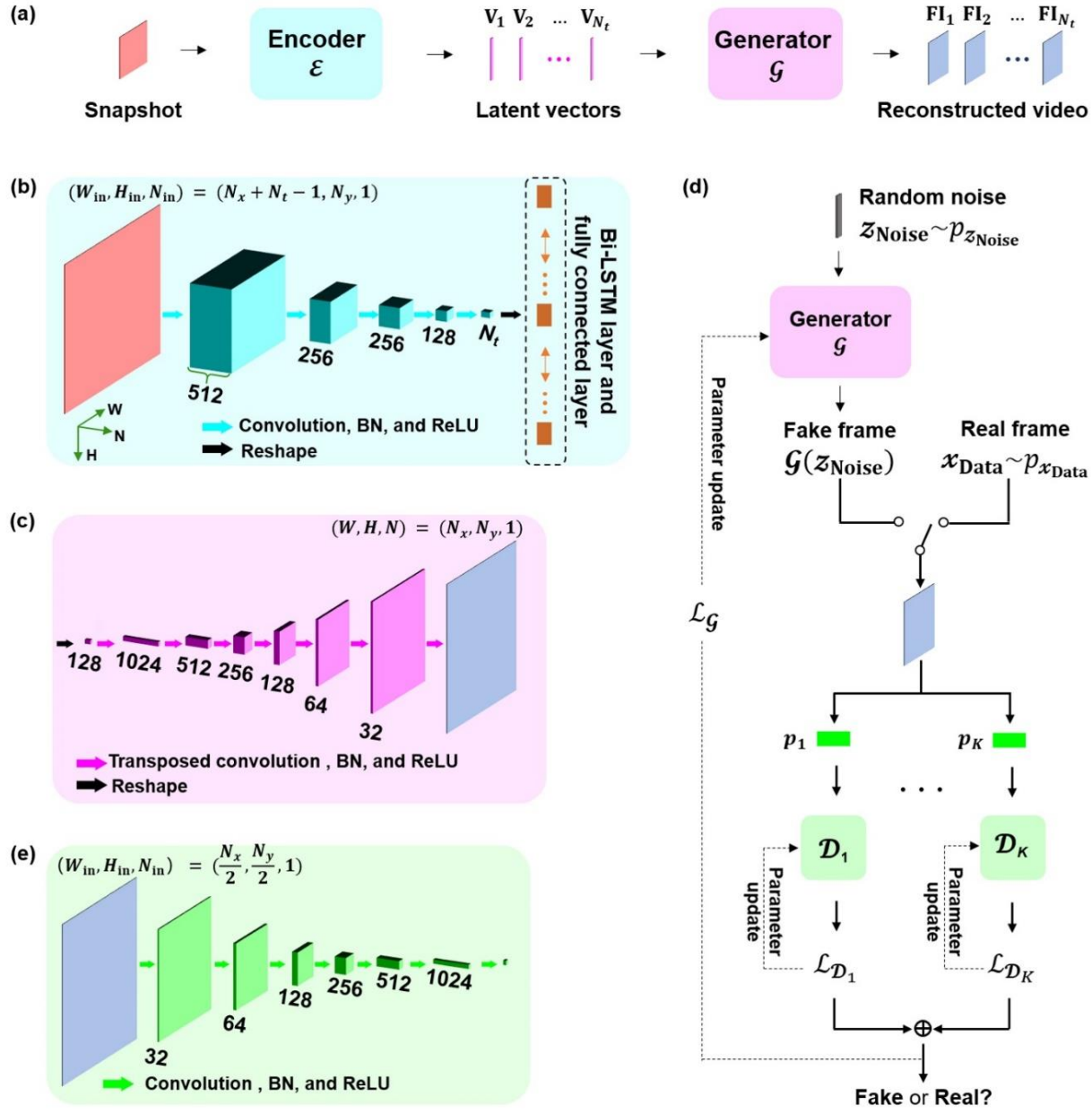


Figure 5.2 Snapshot-to-video autoencoder (S2V-AE). (a) General architecture. FI, frame index. (b) Architecture of encoder showing the generation of latent vectors from a compressively recorded snapshot. Bi-LSTM, Bi-directional long short-term memory; BN, Batch normalization; ReLU, Rectified linear unit; $\boldsymbol{W}$, $\boldsymbol{H}$, and $\boldsymbol{N}$, output dimensions; $\boldsymbol{W}_{\mathbf{in}}$, $\boldsymbol{H}_{\mathbf{in}}$, and $\boldsymbol{N}_{\mathbf{in}}$, input dimensions. (c) Architecture of the generator showing the reconstruction of a single frame from one latent vector. (d) Generative adversarial networks (GANs) with multiple discriminators $\{\boldsymbol{\mathcal{D}_k}\}$. $\boldsymbol{\mathcal{L}_{\mathcal{D}_k}}$, the loss function of each discriminator; $\boldsymbol{\mathcal{L}_{\mathcal{G}}}$, the loss function of the generator; $\{\boldsymbol{p_k}\}$, random projection with a kernel size of [8, 8] and a stride of [2, 2]. (e) Architecture of each discriminator.

The training of the encoder and the generator in the S2V-AE is executed sequentially. Training data are generated on the fly. The details of the training data collection and the training procedure are described in our open-source code (see the link in Disclosure). Additional data, not included in its training phase, are used for evaluation. The generator is first trained under the setting of a GAN with multiple discriminators to ensure its sufficient diversity. In brief, a random noise vector $\boldsymbol{z}_{\mathbf{Noise}}$, sampled from a prior distribution $p_{\boldsymbol{z}_{\mathbf{Noise}}}$ (i.e., $\boldsymbol{z}_{\mathbf{Noise}} \sim p_{\boldsymbol{z}_{\mathbf{Noise}}}$), is input to the generator to produce a fake frame $\boldsymbol{\mathcal{G}}(\boldsymbol{z}_{\mathbf{Noise}})$ that is expected to have visual similarity with the real frame $\boldsymbol{x}_{\mathbf{Data}}$ with an implicit data distribution $p_{\boldsymbol{x}_{\mathbf{Data}}}$ (i.e., $\boldsymbol{x}_{\mathbf{Data}} \sim p_{\boldsymbol{x}_{\mathbf{Data}}}$). The fake or real data are judged by $K = 40$ discriminators [Fig. 5.2(d)]. In each such discriminator, the data are first projected by a random matrix (denoted by $\boldsymbol{p}_k$, where $k = 1, 2, \ldots, K$) to lower dimensions. Then, each discriminator (denoted as $\boldsymbol{\mathcal{D}}_k$) converts the input to a number that is expected to be high for a real frame and low for a fake frame. Each discriminator, corresponding to a binary classifier as schematically shown in Fig. 5.2(e), consists of seven convolutional layers with the numbers of channels ranging from 1024 to 1. The loss functions of each discriminator $\{\boldsymbol{\mathcal{D}}_k\}$ $(k = 1, 2, \ldots, K)$ (denoted by $\mathcal{L}_{\boldsymbol{\mathcal{D}}_k}$) and the generator (denoted by $\mathcal{L}_{\boldsymbol{\mathcal{G}}}$) are calculated by

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\mathcal{D}}_k} = &-\mathbb{E}_{\boldsymbol{x}_{\mathbf{Data}} \sim p_{\boldsymbol{x}_{\mathrm{Data}}}} \left[\log\left(\boldsymbol{\mathcal{D}}_k\left(\boldsymbol{x}_{\mathbf{Data}_k}\right)\right)\right] - \\ &\mathbb{E}_{\boldsymbol{z}_{\mathbf{Noise}} \sim p_{\boldsymbol{z}_{\mathrm{Noise}}}} \left[\log\left(1 - \boldsymbol{\mathcal{D}}_k\left(\boldsymbol{\mathcal{G}}(\boldsymbol{z}_{\mathbf{Noise}})_k\right)\right)\right],\end{aligned} \tag{5.6}$$

$$\mathcal{L}_{\boldsymbol{\mathcal{G}}} = -\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{z}_{\mathbf{Noise}} \sim p_{\boldsymbol{z}_{\mathrm{Noise}}}} \left[\log\left(\boldsymbol{\mathcal{D}}_k\left(\boldsymbol{\mathcal{G}}(\boldsymbol{z}_{\mathbf{Noise}})_k\right)\right)\right]. \tag{5.7}$$

Here, $\mathcal{L}_{\boldsymbol{\mathcal{D}}_k}$ corresponds to the cross-entropy loss [235]. After the random projection $\{\boldsymbol{p}_k\}$, the input to each discriminator is either $\boldsymbol{x}_{\mathbf{Data}_k}$ or $\boldsymbol{\mathcal{G}}(\boldsymbol{z}_{\mathbf{Noise}})_k$. Note that each discriminator is trained on two mini-batches of samples (i.e., real frames and fake frames).The notations $\mathbb{E}_{\boldsymbol{x}_{\mathbf{Data}} \sim p_{\boldsymbol{x}_{\mathbf{Data}}}}[\cdot]$ and $\mathbb{E}_{\boldsymbol{z}_{\mathbf{Noise}} \sim p_{\boldsymbol{z}_{\mathbf{Noise}}}}[\cdot]$ indicate taking the expectations over the distribution $\boldsymbol{x}_{\mathbf{Data}} \sim p_{\boldsymbol{x}_{\mathbf{Data}}}$ and $\boldsymbol{z}_{\mathbf{Noise}} \sim p_{\boldsymbol{z}_{\mathbf{Noise}}}$, respectively. These loss functions are estimated over mini-batches to compute the gradients of losses for each parameter update. Moreover, training iterations are such that each discriminator is first updated in the descent direction of its corresponding loss and then $\mathcal{L}_{\boldsymbol{\mathcal{G}}}$'s gradients are used to update the generator's parameters. The described training game is expected to converge to equilibrium (i.e., no player can improve without changing the other player), which is not guaranteed to occur in highly non-convex cases, such as in the training of neural networks. However, the results found in practice in our setting are satisfactory. Successful

training of the generator will yield parameters that enable its outputs $\boldsymbol{\mathcal{G}}(\boldsymbol{z}_{\mathbf{Noise}})$ to resemble characteristics of the real data. Leveraging this architecture, the goal of each discriminator is to distinguish the real data from the fake ones. The generator, by contrast, aims at fooling all discriminators by learning how to produce frames as close as possible to real data. Parameters in the generator and discriminators are updated according to these loss functions (i.e., Eqs. 5.6 and 5.7), which are minimized via gradient-descent-based optimization.

As the second step, the encoder is trained with the parameters of the generator fixed. The mean square error (MSE) between the recovered video $\boldsymbol{\mathcal{G}}\big(\boldsymbol{\mathcal{E}}(E)\big)$ and the input data I is defined as the loss function denoted by $\mathcal{L}_{\boldsymbol{\mathcal{E}}}$, i.e.,

$$\mathcal{L}_{\boldsymbol{\mathcal{E}}} = \mathrm{MSE}\big[\, \boldsymbol{\mathcal{G}}\big(\boldsymbol{\mathcal{E}}(E)\big), I \,\big]. \tag{5.8}$$

Via minimizing $\mathcal{L}_{\boldsymbol{\mathcal{E}}}$, the encoder learns how to correctly extract the latent vectors with temporal coherence from the 2D snapshot. The training of S2V-AE is finished when the reconstructed video quality stops increasing. Weight decay is employed during the training of the encoder to prevent the weights of the encoder from growing too large [236]. Hyper-parameters to be trained in the encoder are defined through a search over a small grid of candidate values using cross-validation with reconstruction performance measured over a freshly generated batch of data examples.

In the training of both the generator and the encoder, the Adam optimization algorithm [237] was employed with a fixed learning rate, set to 10^{-3} for the training of the generator, and 3×10^{-4} for the training of the encoder. Adam's β_1 and β_2 parameters were set to 0.9 and 0.999 for the training of the generator and 0.5 and 0.9 for the training of the encoder, respectively. Data loading was set at training time such that both scenes and corresponding snapshots were generated on the fly, yielding a virtually infinite amount of training data. Once the completion of both the generator and the encoder training, the S2V-AE was employed to reconstruct dynamic scenes.

5.4 Validation of S2V-AE's reconstruction

To test the feasibility of S2V-AE, we simulated video reconstruction of flying handwritten digits [238]. Each dynamic scene had a size of $(N_x,\ N_y,\ N_t) = (64, 64, 40)$, which produced the snapshot of (103, 64) in size. Snapshots were generated using the forward model of SMART-COSUP (i.e., Eq. 5.1). Simulation results are summarized in Fig. 5.3. For the flying digits corresponding to "3", "5", and "7", six representative frames in the ground truth and the reconstructed videos are shown in Figs. 5.3(a)–(c), respectively. To quantitatively assess the

reconstructed video quality, we analyzed the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [239] frame by frame [Figs. 5.3(d)–(e)]. The average PSNR and SSIM of reconstructed results are 22.9 dB and 0.93, respectively. These results demonstrate that the S2V-AE can accurately reconstruct dynamic scenes from compressively acquired snapshots.

Furthermore, to show that the S2V-AE possesses a more powerful ability of high-quality video reconstruction, we compared its performance to the U-net, which is most popularly used in video compressed sensing [74]. In particular, this U-net featured a convolutional encoder-decoder architecture with residual connection and used the same loss function in Ref. [74]. To implement the optimal specifications of this U-net based technique, we used an approximate inverse operator $\mathbf{\Phi}^{\mathrm{T}}\left(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}\right)^{-\mathbf{1}}$ to alleviate the burden in learning the forward model [54, 74]. In particular to SMART-COSUP, we defined $\mathbf{\Phi} = \mathbf{T}\mathbf{S}_{\mathbf{o}}\mathbf{C}$. Using the compressively recorded snapshot of the scene (i.e., E), the initialized input to the U-net is expressed as $\hat{I}_o = \mathbf{\Phi}^{\mathrm{T}}\left(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{T}}\right)^{-\mathbf{1}}E$, which had the same (x, y, t) dimension to the ground truth. Both the initialized input and its ground truth were used to train the U-net to obtain a good inference ability for new training scenes generated on the fly. To compare the results between the U-net and the S2V-AE, we reconstructed the flying digits of “3”, “5”, and “7”. Despite resembling a close trace of these moving digits to their ground truths, the U-net reconstruction failed to recognize the digits’ spatial information in each frame. The limited feature extraction ability (imposed by the large number of frames in these scenes) and the requirement of high temporal coherence (broken by the fast and randomly moving traces of the digits in these scenes) are the two main reasons that attribute to the unsuccessful reconstruction using the U-net. In contrast, benefiting from its two-step strategy that incorporates spatiotemporal coherence, the S2V-AE has shown superior performance, manifesting in the sharpness of reconstructed digits, the maintenance of high image quality over a large sequence depth, and the capability of handling random moving traces.

5.5 Demonstration of SMART-COSUP

The proof-of-concept experiments of SMART-COSUP were conducted by imaging an animation of three bouncing balls, whose starting positions and moving directions were randomly chosen. This scene had the size of $\left(N_x,\ N_y,\ N_t\right) = (256, 256, 100)$, which produced a snapshot with a size of (355, 256). To improve S2V-AE’s reliability for experimentally captured data, we included various experimental conditions in the forward model to train the S2V-AE. In particular, an experimentally captured mask image was used for the spatial modulation operator. Moreover, with consideration of the noise level in the deployed CMOS camera, Gaussian noise with a

standard deviation randomly selected from 10^{-1} to 10^{-4} was added into the training data to match the signal-to-noise ratios (SNRs) in acquired snapshots. Finally, distortion in the acquired snapshot was corrected by an established procedure [240, 241].

This animation was displayed on a DMD (Ajile Light Industries, AJD 4500) at a pattern refreshing rate of 5 kHz. The trajectories of all three balls were blind to the S2V-AE. A collimated laser beam from a 640-nm continuous-wave laser (CNI, MRL-III-640-50mW) illuminated this DMD at an incident angle of ~24° [Fig. 5.4(a)]. A camera lens (Fujinon, HF75SA1) was used as the front optics. The SMART-COSUP system imaged this event at 5 kfps. A captured 2D snapshot is shown as the inset in Fig. 5.4(a). Video reconstruction using the S2V-AE was compared with those using TwIST and plug-and-play (PnP)-ADMM with the BM3D denoiser [194]. In terms of the reconstruction speed, the execution of algorithms of S2V-AE, TwIST, and PnP-ADMM used 0.06, 5, and 220 seconds, respectively. Thus, the S2V-AE offers speed enhancement of ~80× and ~3,700× to TwIST and PnP-ADMM, respectively. The S2V-AE also provides superior quality in the real-time reconstructed images. Figure 5.4(b) shows five representative frames of ground truth and their corresponding reconstructed results using the three methods. For both TwIST and PnP-ADMM, the reconstructed balls appear blurry and part of the balls are lost in certain frames. In contrast, the S2V-AE provides the best results, in which each ball is fully recovered with a clean background. To quantitatively compare these results, we plotted the PSNRs and SSIMs for all frames [Figs. 5.4(c)–(d)]. The reconstructed frames of S2V-AE have an average PSNR of 25.62 dB, superior to 15.09 dB of TwIST and 16.30 dB of PnP-ADMM. The results from the S2V-AE have an average SSIM of 0.94, considerably better than 0.76 of TwIST and 0.85 of PnP-ADMM. Moreover, we traced the centroids of each ball over time. To further evaluate the reconstruction's accuracy, we calculated the standard deviations of reconstructed centroids (Table 5.1). On average, the S2V-AE improves the accuracy by ~3× compared to the TwIST reconstruction and by ~2× to the PnP-ADMM reconstruction.

Furthermore, the three centroids in each frame were used as vertices to build a triangle. Figures 5.4(e)-(f) show the time histories of the geometric center of this triangle generated from the results of the three reconstruction methods. The standard deviations in the x- and y-directions averaged over time were calculated as (25.4 μm, 17.0 μm), (14.8 μm, 14.5 μm), and (8.3 μm, 6.7 μm) for TwIST, PnP-ADMM, and S2V-AE, respectively. These results show that the S2V- AE has delivered superior performance in image quality and measurement accuracy.

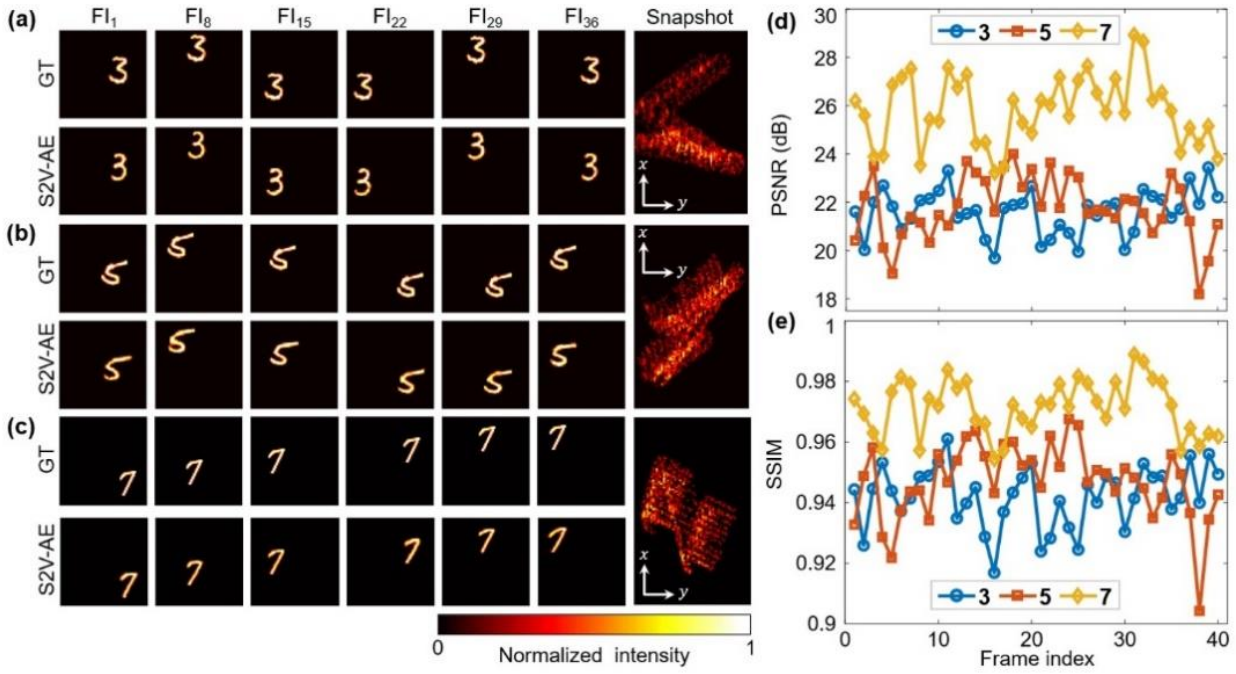


Figure 5.3 Simulation of video reconstruction using the S2V-AE. (a) Six representative frames of the ground truth (GT, top row) and the reconstructed result (bottom row) of the handwritten digit "3". The snapshot is shown in the rightmost column. (b)–(c) As (a), but showing handwritten digits "5" and "7". (d)–(e) Peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) of each reconstructed frame for the three handwritten digits.

Table 5.1 Standard deviations of reconstructed centroids of each ball averaged over time (Unit: μm)

Ball	1		2		3		
Centroid	x	y	x	y	x	y	Mean
TwIST	37.5	36.3	39.4	35.7	43.2	34.9	37.8
PnP-ADMM	27.6	26.2	25.6	25.3	28.6	30.5	27.3
S2V-AE	15.0	12.3	11.0	12.6	15.3	16.0	13.7

5.6 Application of SMART-COSUP to multiple-particle tracking

To show the broad utility of SMART-COSUP, we applied it to tracking multiple fast-moving particles. In the setup, white microspheres were scattered on a surface that rotated at 6800 revolutions per minute [Fig. 5.5(a)]. The 640-nm continuous-wave laser was used to illuminate the rotating microspheres at an incident angle of ~50°. To visualize the beads' continuous motion while capturing a sufficiently long trace, the scattered light was captured by the SMART-COSUP system at 20 kfps. An objective lens (Nikon, CF Achro 4×) was used as the front optics. Figure 5.5(b) shows a static image of three microspheres (marked as M_1–M_3) around the rotation center. Figure 5.5(c) shows a time-integrated image of this dynamic event acquired by using the CMOS

camera in the SMART-COSUP system at its intrinsic frame rate of 20 fps. Due to the low imaging speed, this time-integrated image cannot discern any spatiotemporal details. In contrast, imaging at 20 kfps, SMART-COSUP captures the trajectory of each microsphere. The top image in Figure 5.5(d) provides a color-coded overlay of five reconstructed frames (from 0.55 ms to 4.55 ms with a 1-ms interval), which are shown individually in the bottom row of Fig. 5.5(d). The rotation of M_1 and M_3 at two different radii [i.e., r_{M_1} and r_{M_3} labeled in Fig. 5.5(b)] is evident.

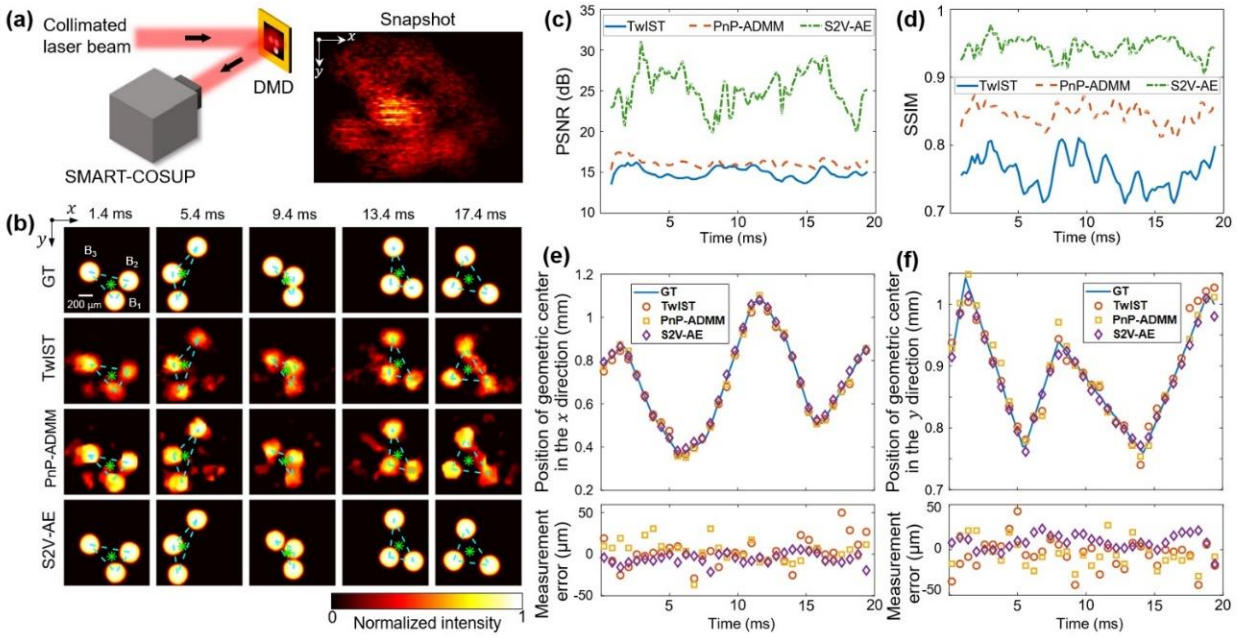


Figure 5.4 SMART-COSUP of animation of bouncing balls at 5 kfps. (a) Experimental setup. DMD, digital micromirror device. Inset: An experimentally acquired snapshot. (b) Five representative frames with 4-ms intervals in the ground truth (GT) and the videos reconstructed by TwIST, PnP-ADMM, and S2V-AE, respectively. Centroids of the three balls are used as vertices to build a triangle (delineated by cyan dashed lines), whose geometric center is marked with a green asterisk. (c)–(d) PSNR and SSIM at each reconstructed frame. (e) Comparison of the positions of the geometric center between the GT and the reconstructed results in the $\boldsymbol{x}$-direction. (f) As (e), but showing the results in the $\boldsymbol{y}$-direction.

To quantitatively analyze these images, we calculated the time histories of x- and y-positions [Fig.5.5(e)] and the corresponding velocities [Fig.5.5(f)] of these microspheres. M_2, sitting at the rotation center, barely changes its position. The time histories of the positions and velocities of M_1 and M_3 follow sinusoidal functions expressed as

$$\boldsymbol{v}_{i(x \text{ or } y)}(t) = \boldsymbol{\omega}_{\mathrm{F}} r_{\mathrm{M}_i} \sin\left(\boldsymbol{\omega}_{\mathrm{F}} t + \alpha_{i(x \text{ or } y)}\right). \tag{5.9}$$

Here, i =1 or 3. $\boldsymbol{\omega}_{\mathrm{F}}$ denotes the angular velocity, whose value was preset at 0.71 rad/ms (i.e., 6800 rounds per minute). r_{M_i} denotes the radius of each microsphere's rotation trajectory. In this

experiment, r_{M_1}= 0.44 mm and r_{M_3}= 0.64 mm. $\alpha_{i(x \text{ or } y)}$ is the initial phase of the i^{th} microsphere in either the x-direction or the y-direction. Thus, the theoretical linear speeds of M_1 and M_3 are 0.31 m/s and 0.45 m/s, respectively.

Based on the above analysis, we used single sinusoidal functions to fit the measured velocities. The fitted maximum velocities in the x-direction and the y-direction are 0.30 m/s and 0.32 m/s for M_1 and 0.46 m/s and 0.45 m/s for M_3, respectively. The fitted angular speeds in the x-direction and the y-direction are 0.71 rad/ms and 0.70 rad/ms for M_1 and 0.71 rad/ms and 0.72 rad/ms for M_3. The experimentally measured values have a good agreement with the preset angular speed of the rotating surface.

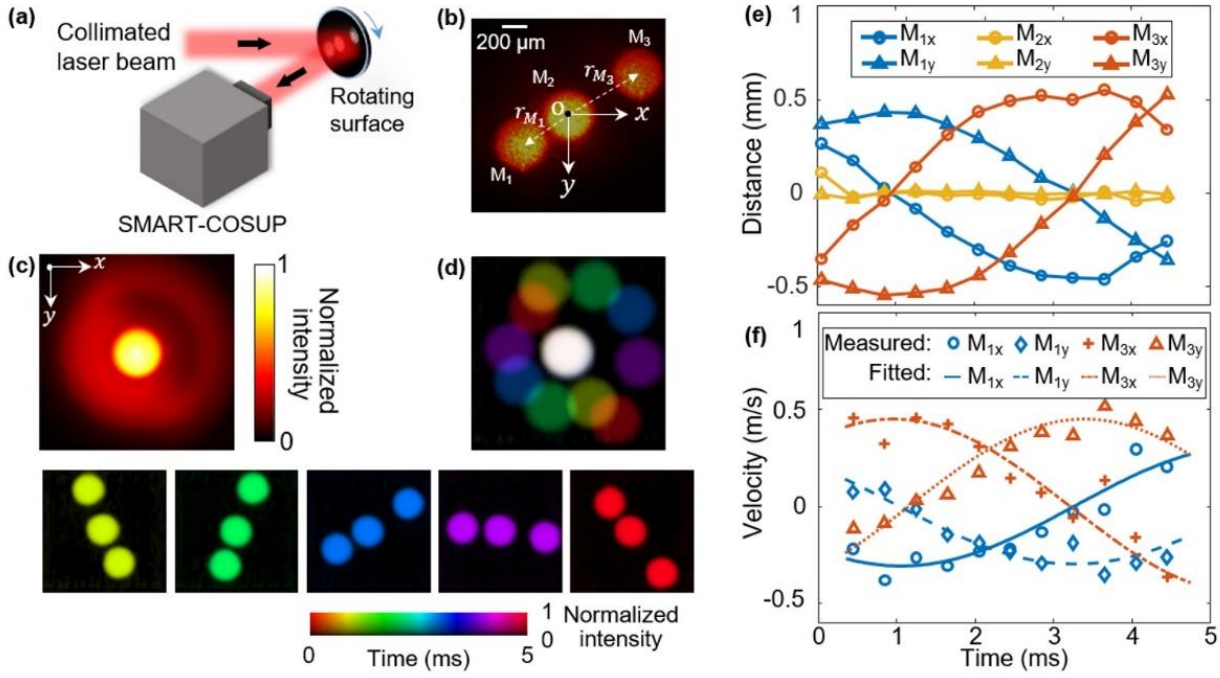


Figure 5.5 SMART-COSUP of multiple-particle tracking at 20 kfps. (a) Experimental setup. (b) Static image of three microspheres (labeled as M1–M3) and the radii (labeled as $\boldsymbol{r}_{\mathbf{M_1}}$ and $\boldsymbol{r}_{\mathbf{M_3}}$). (c) Time-integrated image of the rotating microspheres imaged at the intrinsic frame rate of the CMOS camera (20 fps). (d) Color-coded overlay (top image) of five reconstructed frames (bottom row) with a 1-ms interval. (e) Time histories of the microspheres' centroids. (f) Measured velocities of microspheres with fitting.

5.7 Discussion and conclusions

The S2V-AE offers a new real-time reconstruction paradigm to compressed ultrahigh-speed imaging [Fig. 5.2(a)]. The new architecture of the encoder allows mapping a compressively

recorded snapshot into a set of low-dimensional latent vectors. After that, the GAN-trained generator maps such latent vectors into frames of the reconstructed video. Using this scheme, the training procedure is divided into two distinct phases: to train a generative model of static frames and to train an encoding model aiming to sample from the generator. By doing so, unlike direct reconstruction approaches, high quality in frame-wise reconstruction can be ensured by the initially trained generator, while the encoding model needs to learn only how to query coherently across time. This scheme brings in benefits to the reconstructed videos in terms of both quality and flexibility. The encoder in S2V-AE preserves coherence in both space and time. Different from previous works [54, 215, 216], no artificial segmentation is conducted in the S2V-AE, which avoids generating artifacts due to the loss of spatial coherence. The S2V-AE also explicitly models temporal coherence across frames with the Bi-LSTM. Both innovations ensure artifact-free and high-contrast video reconstruction of sophisticated moving trajectories. Meanwhile, the S2V-AE presents a flexible structure with a higher tolerance for input data. In particular, the generator, used in a PnP setting [147], is independent of the system's data acquisition, which is important for adaptive compressed sensing applications.

The multiple-discriminator framework implemented in the S2V-AE improves training diversity. While able to generate high-quality, natural-looking samples, generators trained under the framework of the GAN have known drawbacks that have to be accounted for at training time. Namely, mode collapse refers to cases where trained generators can generate only a small fraction of the data support [242]. Standard GAN settings do not account for the diversity of the generated data, but instead, the generator is usually rewarded if its outputs are individually close to the real data instances. As such, a large body of recent literature has tackled the mode collapse using different approaches to improve the diversity of the GAN generators [242, 243]. Mode collapse is especially critical in the application we consider herein. The generator in the S2V-AE needs to be able to generate any possible frame, which means being able to output images containing any objects (e.g. balls or digits) in any position. To ensure that the generator is sufficiently diverse, the S2V-AE implements the multiple-discriminator framework [244, 245]. Moreover, each such discriminator is augmented with a random projection layer at its input. More random views of the data distribution aid the generator in producing results that are approximate to the real data distribution.

The S2V-AE enables the development of SMART-COSUP. This new technique has demonstrated the largest sequence depth (i.e., 100 frames) in existing DNNs-based compressed ultrahigh-speed imaging methods [54, 74, 215-218]. The sequence depth, as a tunable

parameter, could certainly exceed 100 frames. In this aspect, the performance of the S2V-AE mainly depends on the encoder [Fig. 5.2(b)] since it needs to extract the same number of latent vectors as the sequence depth. Although a large sequence depth may bring in training instabilities due to vanishing/exploding gradients, our choice of the Bi-LSTM architecture in the S2V-AE could alleviate gradient-conditioning issues relative to standard recurrent neural networks [246]. Thus, we expect the limit of sequence depth to be up to 1000 frames in the current setup. Moreover, although we only experimentally demonstrated the 20-kfps imaging speed in this work, the S2V-AE could be extended to reconstructing videos with much higher imaging speeds. As shown in Eq. 5.5, SMART-COSUP's imaging speed is determined completely by hardware. Regardless of the imaging speed, the operation of the S2V-AE—reconstruction of a 3D datacube from a 2D snapshot—remains the same. Moreover, considering the link between imaging speeds and SNRs, the successful reconstruction of snapshots with different SNRs during the training procedure (Section 5.5) indicates S2V-AE's applicability to reconstruct videos with a wide range of imaging speeds. Furthermore, SMART-COSUP replaces the DMD with a printed transmissive mask. Despite being inflexible, the implemented pseudo-random binary pattern has better compatibility with diverse dynamic scenes, improves light throughput and image quality, as well as offers a simpler and more compact system arrangement. Along with its real-time image reconstruction, the SMART-COSUP system is advancing towards real-world applications.

In summary, we have developed the S2V-AE for fast and high-quality video reconstruction from a single compressively acquired snapshot. This new DNN has facilitated the development of the SMART-COSUP system, which has demonstrated single-shot ultrahigh-speed imaging of transient events in both macroscopic and microscopic imaging at up to 20 kfps with a real-time reconstructed video size of $(N_x, N_y, N_t) = (256, 256, 100)$. This system has been applied to multiple-particle tracking. Despite demonstrated only with the SMART-COSUP system, the S2V-AE could be easily extended to other modalities in compressed temporal imaging [196] and single-shot hyperspectral imaging [247]. Moreover, by implementing the variational AE [248], the dependence of the encoder on the sensing matrix could be further reduced. SMART-COSUP's ability to track multiple fast-moving particles in a wide field may enable new applications on particle imaging velocimetry [249] and flow cytometry [250]. All of these topics are promising research directions in the future.

Funding. This work was supported in part by Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-05959, RGPAS-507845-2017, CRDPJ-532304-18, I2IPJ-

555593-20), Canada Foundation for Innovation and Ministère de l'Économie et de l'Innovation du Québec (37146), Fonds de Recherche du Québec - Nature et Technologies (2019-NC-252960), Fonds de Recherche du Québec - Santé (267406, 280229), Ministère des Relations internationales et de la Francophonie du Québec, Computer Canada, and Calcul Québec.

Acknowledgments. The authors thank Patrick Kilcullen for experimental assistance.

Disclosures. The authors declare no conflicts of interest. The software of the S2V-AE and representative data of SMART-COSUP can be downloaded from https://github.com/joaomonteirof/SMART_COUSP_Reconstruction.

6 CS-BASED ULTRAFAST ELECTRON IMAGING

Single-shot Real-time Sub-nanosecond Electron Imaging Aided by Compressed Sensing: Analytical Modeling and Simulation

Imagerie électronique sub-nanoseconde en une seule prise de vue en temps réel assistée par détection compressée : modélisation analytique et simulation

Authors:

Xianglei Liu[1], Shian Zhang[2], Aycan Yurtsever[1,*], and Jinyang Liang[1,*]

[1]Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650 boulevard Lionel-Boulet, Varennes, Québec J3X1S2, Canada

[2]State Key Laboratory of Precision Spectroscopy, East China Normal University, 3663 North Zhongshan Road, Shanghai, 200062, P. R. China.

*Correspondence should be addressed to

aycan.yurtsever@emt.inrs.ca (A.Y.) and jinyang.liang@emt.inrs.ca (J.L.)

Publication:

Micron

Vol. 117, pp. 47-54 (2019)

https://doi.org/10.1016/j.micron.2018.11.003

Contribution of authors: Xianglei Liu built the system, conducted the experiments, and analyzed the data. Prof. Shian Zhang initiated the project. Prof. Aycan Yurtsever initiated the project and provided the experimental data. Prof. Jinyang Liang initiated the project, proposed the concept, contributed to experimental design, and supervised the project. All authors wrote and revised the manuscript.

Abstract

Bringing ultrafast (nanosecond and below) temporal resolution to transmission electron microscopy (TEM) has historically been challenging. Despite significant recent progress in this direction, it remains difficult to achieve sub-nanosecond temporal resolution with a single electron pulse, in real-time (i.e., duration in which the event occurs) imaging. To address this limitation, here, we propose a methodology that combines laser-assisted TEM with computational imaging methodologies based on compressed sensing (CS). In this technique, a two-dimensional (2D) transient event [i.e. (x, y) frames that vary in time] is recorded through a CS paradigm, which consists of spatial encoding, temporal shearing via streaking, and spatiotemporal integration of an electron pulse. The 2D image generated on a camera is used to reconstruct the datacube of the ultrafast event, with two spatial and one temporal dimensions, via a CS-based image reconstruction algorithm. Using numerical simulation, we find that the reconstructed results are in good agreement with the ground truth, which demonstrates the applicability of CS-based computational imaging methodologies to laser-assisted TEM. Our proposed method, complementing the existing ultrafast stroboscopic and nanosecond single-shot techniques, opens up the possibility for single-shot, real-time, spatiotemporal imaging of irreversible structural phenomena with sub-nanosecond temporal resolution.

Keywords:

transmission electron microscopy; ultrafast imaging; compressed sensing; streak imaging;

6.1 Introduction

Imaging materials with combined high spatial and temporal resolutions merits our attention to understanding the origins of many microscopic phenomena. However, developing such spatiotemporal imaging techniques has been traditionally challenging because of the limitations of the elementary particles used in the imaging processes. Photons and electrons, two fundamental particles that account for the most, if not all, imaging technologies, are dissimilar in terms of the spatiotemporal domains they can access. Photons, with no dispersion in free space and with no elementary charge, excel in forming ultrashort pulses that can propagate long distances and, hence, can be used for extremely high (up to attosecond) temporal studies. Fast electrons, on the other hand, with their picometer wavelengths and strong interaction cross-sections, excel in forming images with the highest spatial resolution (sub-angstrom) available today.

Methodologies based on the former, such as ultrafast optical spectroscopy, have offered access to the first “moments” of fundamental phenomena, including the birth of chemical species and evolution of molecular bonds [251, 252], phonon dynamics [253, 254], and valance-band electronic transitions [255]. While excelling in temporal resolution, photons lack the extreme spatial resolution that electrons can achieve in transmission electron microscopy (TEM).

In the last couple of decades, it was possible to make photons and electrons work together in one microscope [31-33, 256-262] to achieve combined spatiotemporal resolutions that were not possible before. At the California Institute of Technology, a TEM system was modified to merge it with a femtosecond laser [263]. In such an ultrafast TEM, instead of using continuous electron beams, ultrashort electron pulses created by the photoelectric effect were used for imaging, while optical pulses are used to clock the sample synchronously. With such a pump-probe scheme, it was possible to achieve few-nanometers real-space resolution combined with sub-picosecond temporal resolution stroboscopically [261], which represents the current resolution limits in spatiotemporal imaging. Complementarily, another photon-electron imaging methodology was developed at the Lawrence Livermore National Laboratory (LLNL) [264], following the original work of O. Bostanjoglo and colleagues [256-258]. This technique was based on using several intense electron pulses to image a photo-excited sample, and then deflecting these pulses successively to fill a CCD camera with several frames. This single-shot, real-time (i.e., duration in which the event occurs), ultrafast imaging approach enabled to visualize irreversible phenomena, albeit with limited sequence depth (i.e., the maximum number of frames

that can be captured in one acquisition), real-space resolution of tens of nanometers, and temporal resolution in the nanosecond range [264].

Recently, compressed sensing (CS)—a scheme that allows underdetermined reconstruction of sparse scenes [88]—has been introduced into non-laser-assisted TEM [265]. In this proposal, a coded-aperture-based CS method [62] was implemented to increase the TEM's frame rate. After the sample, the spatiotemporally modulated electron beam was encoded by a binary transmissive mask loaded on a moving piezoelectric stage. Different encoding masks were attached to individual temporal frames. These spatially encoded frames were integrated in a single camera exposure during the acquisition process. Their simulation showed that a movie could be reconstructed via CS inversion by using the captured single image. This proposed scheme would allow increasing TEM's frame rate to several kilohertz [62]. However, constrained by the translational speed of the piezoelectric stage, the limited frame rate prevents this scheme from achieving sub-nanosecond temporal resolution. CS has also been implemented in the time domain with a TEM system [266, 267]. In this technique, four replicated images (arranged as a 2×2 array) of the dynamic scene was generated on an image sensor by a high-speed deflector. A dynamic shutter, encoded by a distinct grayscale code sequence for each image, modulated the temporal integration process. The acquired data were fed into a CS-based reconstruction algorithm to generate 10 frames at the kilohertz level. Thus, this method has limitations in the imaging speed and the sequence depth.

To overcome these limitations, here, we introduce a new methodology in laser-assisted TEM, inspired by compressed ultrafast photography (CUP) that has demonstrated breakthroughs in recording optical ultrafast transient scenes [19-21]. CUP, based on streak imaging and CS, has exhibited the fastest receive-only imaging modality at up to 10 trillion frames per second with 0.58 ps temporal resolution. Leveraging a CS-based data acquisition and image reconstruction paradigm, CUP adds another spatial dimension into the conventional streak camera, enabling two-dimensional (2D) ultrafast (i.e., x, y, t) imaging with one camera exposure[90]. CUP has enabled capturing, for the first time, a number of transient optical phenomena, such as real-time fluorescence lifetime mapping [19], propagation of a scattering-induced photonic Mach cone [20], and temporal focusing of a femtosecond laser pulse [21]. Thus far, CUP has been applied only in optical imaging. However, CUP's principle is generic, indicating the possibility of extending it to imaging with electrons.

In this work, we propose two single-shot real-time ultrafast electron imaging configurations, synergizing CUP and laser-assisted TEM. The first configuration, termed

compressed ultrafast TEM (CUTEM), can be integrated into a laser-assisted TEM system without excessive modifications. The second configuration, termed dual-shearing (DS)-CUTEM, can offer improved technical specifications, but with more elaborate modifications to the hardware. For each configuration, we examined how each manipulation of a transient scene during electron propagation and data acquisition affected reconstructed image quality by using experimental data as the ground truth. These quantitative analyses guided us to optimize the reconstructed imaging quality. As a result, we were able to use the proposed configurations to reconstruct spatiotemporal datacubes, which were found to be in good agreement with the ground truth. This study verifies the feasibility of implementing CS-aided ultrafast imaging modalities to TEM. The proposed schemes have potential to achieve single-shot ultrafast imaging in real time with sub-nanosecond temporal resolution and nanometers real-space resolution.

The rest of the paper is organized as follows. In Section 6.2, we describe the system designs with the analytical models of associated data acquisition and imaging reconstruction. In Section 6.3, we present the optimization of reconstructed image quality and proof-of-concept demonstrations using numerical simulation. In Section 6.4, we discuss how the proposed techniques could guide future experiments, followed by a short summary that concludes this paper.

6.2 Method

6.2.1 CUTEM

The proposed CUTEM [Fig. 6.1(a)] technique can be experimentally realized by integrating a mask and shearing electrodes to a laser-assisted TEM machine. In CUTEM, a single nanosecond electron pulse probes a transient event initiated in a sample to generate a spatiotemporally modulated electron density function, $I(x, y, t)$. The transmitted electron pulse, then, experiences several manipulations, each of which is accounted by an operator (detailed in the following paragraph). Specifically, the pulse undergoes spatiotemporal electron distortion during the ensuing propagation within the TEM system. A binary transmission mask, with a specifically designed pattern, is placed on the path of the imaging electrons to encode $I(x, y, t)$. Produced by existing nanofabrication tools, this mask blocks the electrons at certain regions in space and allows them to pass through at others. Following the mask, a time-varying voltage, applied to a pair of electrodes, temporally shears $I(x, y, t)$. Finally, a single streak image, denoted as E, is

formed on a CCD camera via spatiotemporal integration (spatially integrating over each CCD pixel and temporally integrating over the exposure time). The mask and the shearing electrodes can be inserted to the TEM machine using one of the ports after the projection lens system.

Mathematically, the above-described data acquisition process can be expressed as a forward model [Fig. 6.1(a)] by

$$E = \mathbf{TSCD}\, I(x, y, t), \tag{6.1}$$

where the linear operator $\mathbf{D}$ accounts for electron's spatiotemporal distortion induced by all experimental factors. $\mathbf{C}$ represents spatial encoding by the binary transmission mask, $\mathbf{S}$ represents temporal shearing by the time-varying voltage applied to the electrodes, and $\mathbf{T}$ represents spatiotemporal integration by the CCD camera. For simplicity in denotation, we use a linear operator $\mathbf{O}$ to represent the entire data acquisition process (i.e., $\mathbf{O} = \mathbf{TSCD}$). After data acquisition, we retrieve the transient scene computationally. Given the known operator $\mathbf{O}$ and the spatiotemporal sparsity of the transient scene, $I(x, y, t)$ can be recovered by solving the inverse problem of Eq. (6.1). This process can be formulated as

$$\hat{I} = \underset{I}{\operatorname{argmin}} \|E - \mathbf{O}I\|_2^2 + \lambda \Phi(I)\,, \tag{6.2}$$

where λ is a weighting parameter, and $\Phi(\cdot)$ is the regularization function. In Eq. (6.2), the minimization of the first term, $\|E - \mathbf{O}I\|_2^2$, occurs when the actual measurement E closely matches the estimated solution $\mathbf{O}I$, while the minimization of the second term, $\Phi(I)$, encourages I to be piecewise constant (i.e., sparse in the spatial gradient domain). The weighting of these two terms is adjusted by λ to lead to the results that are most consistent with the ground truth. Given the number of rows and columns of the CCD camera to be N_{r} and N_{c}, the number of voxels in the reconstructed datacube (i.e., N_x, N_y, and N_t) must meet the requirement of

$$\begin{aligned} &N_x \le N_{\mathrm{c}}\,, \ \ and \\ &N_y + N_t - 1 \le N_{\mathrm{r}}\,. \end{aligned} \tag{6.3}$$

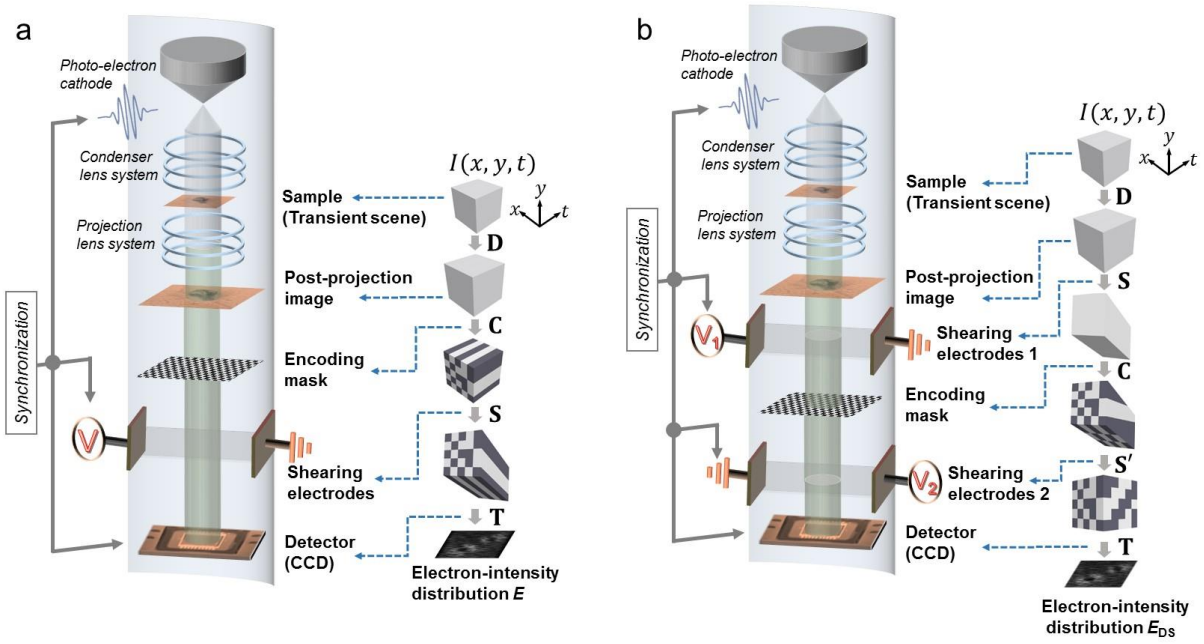


Figure 6.1 Proposed methods for single-shot real-time ultrafast electron imaging. (a), Hardware schematic of CUTEM and its forward model. $\mathbf{C}$, spatial encoding operator; $\mathbf{D}$, electron spatiotemporal distortion operator; $\mathbf{S}$, temporal shearing operator; $\mathbf{T}$, spatiotemporal integration operator. (b), Hardware schematic of DS-CUTEM and its forward model. $\mathbf{S}'$, reverse temporal shearing operator. $\boldsymbol{t}$, time; $\boldsymbol{x}, \boldsymbol{y}$, spatial coordinates.

6.2.2 Dual-shearing (DS)-CUTEM

To increase the size of reconstructed datacube, we further propose a variant referred to as dual shearing (DS)-CUTEM, whose system schematic is shown in Fig. 6.1(b). The difference of DS-CUTEM with respect to CUTEM is that $I(x,y,t)$ is sheared twice in opposite directions by two pairs of electrodes, which sandwich the spatial encoding mask. A similar voltage ramping waveform is used to control both pairs to avoid asymmetric shearing. The delay of shearing onsets is determined by the distance between these two deflectors and the electron pulse's propagation speed. The streak image recorded on the CCD camera is denoted as E_{DS}. Mathematically, the forward model of DS-CUTEM can be expressed as

$$E_{\mathrm{DS}} = \mathbf{TS'CSD}\, I(x,y,t), \tag{6.4}$$

where $\mathbf{S}'$represents temporal shearing in a reverse direction to that of $\mathbf{S}$. After data acquisition, DS-CUTEM uses the similar computational reconstruction method as CUTEM to retrieve the result, expressed by

$$\hat{I} = \underset{I}{\operatorname{argmin}} \|E_{\mathrm{DS}} - \mathbf{O}'I\|_2^2 + \lambda\Phi(I)\,, \tag{6.5}$$

where the linear operator $\mathbf{O}' = \mathbf{TS}'\mathbf{CSD}$.

Compared with CUTEM, the dual-shearing operation implemented in DS-CUTEM allows the temporal course of a specific spatial position in the scene to be integrated at the same pixel of the CCD camera, circumventing the requirement [i.e., Eq. (6.3)] that is imposed by CUTEM. Consequently, each frame in the reconstructed datacube could have a maximum size of $N_x \times N_y = N_\mathrm{c} \times N_\mathrm{r}$, and N_t would be limited by the dynamic range of the CCD camera. In addition, the dual-shearing operation implemented in DS-CUTEM enables designing a more incoherent measurement matrix. In particular, while encoded by the same mask in CUTEM, each frame in $I(x, y, t)$ is imprinted by a different random binary mask in DS-CUTEM due to the first shearing operation prior to the spatial encoding. Consequently, DS-CUTEM has less mutual coherence between the measurement matrix and the sparsity basis of the transient scene than that of CUTEM, which yields an improved reconstructed image quality (detailed in Section 6.3).

6.2.3 Comparison of CUTEM and DS-CUTEM with framing TEM

The proposed two configurations could exceed framing TEM [i.e. where the CCD camera is filled with spatially-separated frames, like in the LLNL TEM system [264]] in the temporal resolution and the sequence depth. Both CUTEM and DS-CUTEM use spatial encoding and temporal shearing to tag each frame with a spatiotemporal "barcode". This prior information, along with the sparsity in the transient scene, allows spatiotemporal mixing of adjacent frames, which can be recovered in image reconstruction. The reconstructed frame rate, r, is thus determined by

$$r = \frac{v}{d}, \tag{6.6}$$

where v is the temporal shearing velocity, and d is the CCD's pixel size along the temporal shearing direction. As a result, for a given recording time window, the proposed configurations can largely improve the temporal resolution, compared with the framing TEM. For instance, using a high-speed electrostatic deflector, the framing TEM [264] can acquire up to 5×5 frames [Fig. 6.2(a)], each of which has 400×400 pixels in size to fill up a sensor with pixels of $N_\mathrm{c} \times N_\mathrm{r} = 2000 \times 2000$. The inter-frame time is 25 ns. In comparison, spatiotemporal mixing allowed in CUTEM could produce an inter-frame time of 62.5 ps. With the consideration of temporal blurring effect [19], the estimated temporal resolution is approximately 0.3 ns. According to Eq. (6.3), in CUTEM, the x axis is used only for recording the spatial information, which allows making full use

of the N_c (i.e., $N_x = 2000$). In the y axis, the pixels are distributed to record both spatial and temporal information. Physically, this restriction means that the size of captured 2D image has to be less than the full size of the sensor [Fig. 6.2(b)]. Therefore, if we keep the number of pixels in the y axis the same as the framing TEM (i.e., $N_y = 400$), the maximum sequence depth is around $N_t = 1600$.

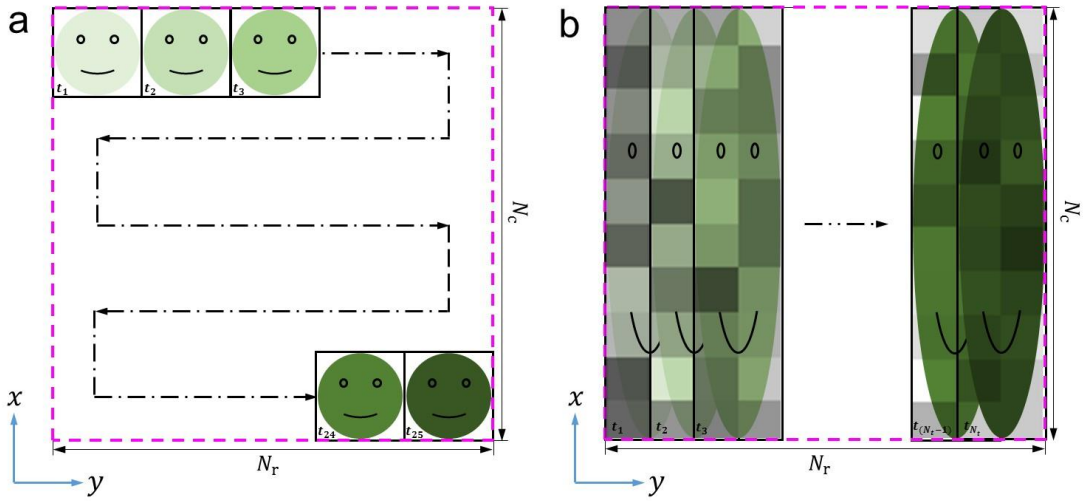


Figure 6.2 Illustrative comparison of the captured images on the camera by (a) framing TEM and (b) CUTEM. The magenta dashed box indicates the camera's sensor area. The dash-dotted arrow lines in (a) and (b) indicate the direction of raster scanning of framing TEM and the temporal shearing direction of CUTEM, respectively. $\boldsymbol{N_r}$ and $\boldsymbol{N_c}$, number of rows and columns of the camera. In (b), the adjacent two frames can be offset by as few as one pixel.

6.2.4 Simulation setup

To test the proposed methods, two transient scenes were used. The first scene [Fig. 6.3 (a)] depicts near-field intensity dynamics of dipoles of two close-by silver nanoparticles excited by an ultrashort laser pulse (referred to hereafter as the 'dipole scene') [268]. While holding fixed spatial positions, the dipoles have fluctuating intensities in time. Second, as shown in Fig. 6.3 (b), a transient scene of a moving nanoscale cantilever was used (referred to as the 'cantilever scene' hereafter) [269]. Complementary to the dipole scene, the cantilever moves in space, while its intensity is almost constant. Both datasets are experimentally acquired and previously published [268, 269]

The general framework of our simulation is as follows. First, the sizes of both the dipole scene and cantilever scene were set to $N_x \times N_y \times N_t = 300 \times 300 \times 30$. Then, both scenes were used for the forward models of the two proposed methods [i.e., Eqs. (6.1) and (6.4)] to generate E and E_{DS}, respectively. To maximally mimic the experimental conditions, no noise filtering was

conducted to any frames in either scene. The sizes of E and E_{DS} were $N_x \times N_{y,t} = 300 \times 329$ and $N_x \times N_{y,t} = 300 \times 300$, respectively. Here, $N_{y,t}$ represents the number of pixels in the axis where the spatial and temporal information are mixed. Finally, E and E_{DS} were fed into the two-step iterative shrinkage/thresholding (TwIST) algorithm [68] to retrieve $I(x, y, t)$ by solving Eqs. (6.2) and (6.5). Total variation (TV) was used as the regularization function [270]. Compared with other regularization functions, such as l_0 or l_1 norms, TV regularization has superior performance in denoising while preserving important details in images [271].

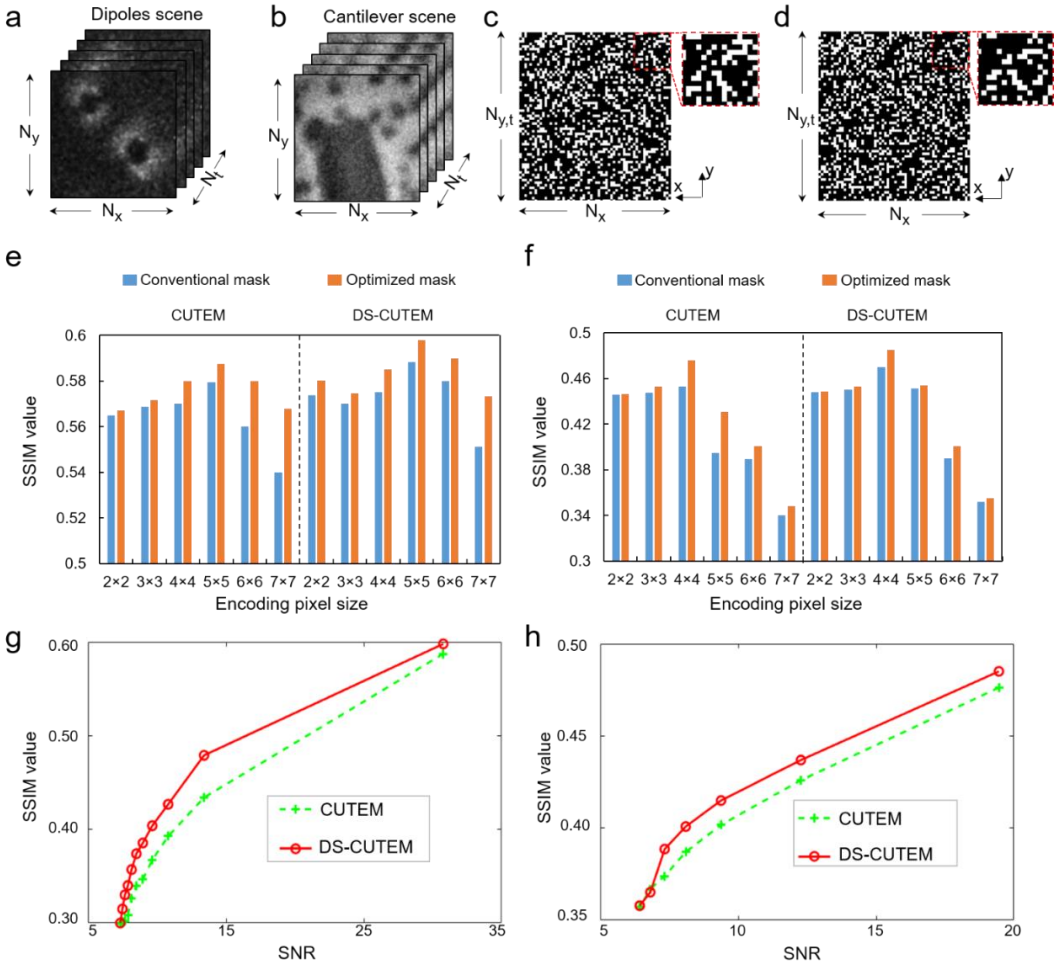


Figure 6.3 Examination of the reconstructed image quality to the types of encoding masks, encoding pixel sizes, and SNRs. (a), Dipole scene. (b), Cantilever scene. (c), Conventional binary mask. (d), Optimized binary mask. Insets in (c) and (d) show the details of local features of the encoding masks; $\boldsymbol{x}, \boldsymbol{y}$, spatial coordinates. (e), SSIM values of the reconstructed dipole scene versus different encoding pixel sizes and the two types of encoding masks for both proposed methods. (f), As (e), but shows the results using the cantilever scene. (g), SSIM values with the respect to SNRs for the dipole scene (using 5×5 encoding pixel size). (h), As (g), but shows the SSIM-SNR relations for the cantilever scene (using 4×4 encoding pixel size). The optimized binary masks were used for obtaining the results in (g) and (h).

6.2.5 Examination of various operators and reconstructed image quality

To investigate how each operator impacts the reconstructed image quality, we used different combinations of encoding masks, encoding pixel sizes (defined as how many pixels of the sensor each encoding pixel contains), and signal-to-noise ratios (SNRs) in both proposed methods. First, the encoding mask provides the necessary prior information to extract the temporal information in image reconstruction. Following the established theories [63, 272, 273], a pseudo-random binary mask [shown in Fig. 6.3(c) and referred to hereafter as the "conventional mask"] was implemented. This mask was constructed by binarizing a random matrix whose element values follow a Gaussian distribution. While incoherent with the most sparse representation of the scene, the conventional mask might not be the optimal choice for a given sparse basis. To optimize the encoding mask, we used a column-normalized random matrix $\tilde{C}$ to generate its Gram matrix [274], defined by $\tilde{C}^T\tilde{C}$. When the Gram matrix approaches to a unit matrix H, the mutual coherence reaches the minimum. Mathematically, this optimization process can be formulated as

$$\hat{C} = \underset{\tilde{C}}{\operatorname{argmin}} \left\| \tilde{C}^T\tilde{C} - H \right\|_F^2 , \tag{6.7}$$

where $\|\cdot\|_F^2$ represents Frobenius norm. In practice, a gradient descent method [275, 276] was implemented to generate the optimized random binary mask [Fig. 6.3(d)]. For simplicity, we denote this mask as the "optimized mask".

Besides the mask pattern, we also optimized the encoding pixel size. On one hand, to satisfy the Nyquist sampling theorem, the smallest feature in both spatial and temporal domains has to be sampled by at least two encoding pixels, meaning that the smaller pixel size aides spatial and temporal resolutions. On the other hand, the smaller pixel size is more vulnerable to noise in the measurement and thus demands higher SNRs for transient scenes [277]. Thus, the optimized encoding pixel size can be chosen by balancing the achievable resolution and noise.

Finally, the SNRs in E and E_{DS}, associated with the operator $\mathbf{T}$ [Eqs. (6.1) and (6.4)], were analyzed. Here, we defined the SNR as the quotient of the maximal intensity to the standard deviation of the noise in the acquired images (i.e., E and E_{DS}). To vary the SNR, we added zero-mean Gaussian noise with variances from 0 to 0.5.

To quantitatively compare the reconstructed image quality, the structural similarity (SSIM) [278] was used as the merit function. SSIM is a perception-based assessment model that considers image degradation as perceived change in structural information while also incorporating significant perceptual phenomena, including both luminance and contrast terms.

Compared with conventional image quality assessment methods, such as mean squared error and peak signal-to-noise ratio, SSIM has a better capability to represent the perceived visual quality.

Figures 6.3(e) and (f) show SSIM values of the reconstructed dipole and cantilever scenes, using both the conventional and optimized encoding masks and with different encoding pixel sizes, for both CUTEM and DS-CUTEM. For both scenes, dual-shearing operation produces better reconstruction image quality than its single-shearing counterpart because of its more incoherent measurement matrix. Figures 6.3(e) and (f) also illustrate that mask optimization improves reconstructed image quality. Although the mask optimization was conducted only in the spatial domain, because C is a part of measurement matrix O and O', this operation still assisted in minimizing the mutual coherence between these measurement matrices and the sparsity basis, leading to a better reconstruction image quality. It is worth noting that although only demonstrated in the spatial domain in this work, mask optimization can be implemented for other sparsity bases, including discrete cosine transformation [279] and discrete wavelet transformation [280]. Finally, we found the optimal encoding pixel size to be 5×5 for the diploe scene and 4×4 for the cantilever scene, respectively. The varied optimal sizes are ascribed to the different imaging contents and characteristics in the two scenes.

Figures 6.3(g) and (h) present the changes in SSIM values of the dipole scene and the cantilever scene versus the SNRs of E and E_{DS}. The optimized encoding patterns and the optimized encoding pixel sizes were used for both scenes in this quantification. With decreasing SNRs, although the reconstructed image quality descends drastically for both proposed methods, DS-CUTEM performs consistently better than CUTEM until the SNR approaches five. There, the SSIM for both proposed methods converge and remain relatively unchanged for lower SNRs, because the noise has dominated the measurement results.

Besides the above-described optimization, we accounted for the influence of the space-charge effect on image distortion by introducing the operator $\mathbf{D}$ in Eqs. (6.1) and (6.4). The space-charge effect degrades both spatial resolution and temporal resolution. First, by distorting the probing electron pulse, it causes deformation of the system's point spread function. The distortion would lead to crosstalk (i.e., electrons that are supposed to be recorded by a certain pixel on the image sensor are actually recorded by a different pixel) in the spatiotemporal integration. Because spatial information and temporal information are mixed during the data acquisition in both CUTEM and DS-CUTEM, this crosstalk would affect the reconstruction accuracy in both the spatial domain

and the time domain. Finally, this inaccurate reconstruction will transfer to the degraded spatial and temporal resolutions.

We modeled electrons' spatial and temporal expansion by a spatiotemporal Gaussian model. The expansion ratio in the model is defined as the quotient between the size of the distorted transient scene and that of the original one. A ratio smaller than one means spatial or temporal shrinkage. Both proposed methods were implemented to account for $\mathbf{D}$ and produced comparable results. For simplicity, only the results of DS-CUTEM are shown. The flow chart of the reconstruction process with electron distortion is shown in the Fig. 6.4(a). The transient scene first went through the operator $\mathbf{O}'$ to generate E_{DS} [see Eq. (6.4)]. Then, E_{DS} was used as an input into the TwIST algorithm to retrieve the estimated measurement $\hat{I}_{\mathrm{u}}(x,y,t)$. The operator $\mathbf{O}'_{\mathbf{u}}$ represents $\mathbf{TS}'\mathbf{CS}$, and its inverse operator is represented as $\mathbf{O}'^{*}_{\mathbf{u}}$. Here, the subscript "u" stands for "uncorrected". Finally, we used the estimated measurement $\hat{I}_{\mathrm{u}}(x,y,t)$ and input transient scene $I(x,y,t)$ to calculate the SSIM value. As a comparison, the flow chart of the reconstruction process with distortion correction is shown in Fig. 6.4(b). $\mathbf{O}'^{*}$ denotes the inverse operator of $\mathbf{O}'$.

Figure 6.4(c) shows the normalized SSIM values of reconstructed dipole scene with spatial and temporal expansions, both with ratios between 0.7 and 1.3. The expansions in both spatial and temporal domains deteriorated the reconstructed image quality, manifesting in the radially decreasing SSIM values from the origin (i.e., no spatial or temporal expansion). Similar degradation trend is observed for the cantilever scene [Fig. 6.4(d)]. The result after correcting for electrons' spatiotemporal distortion through the operator $\mathbf{D}$ is shown in Figs. 6.4(e) and (f). In both cases, corrections significantly alleviate the degradation, demonstrated as the high normalized SSIM values (between 0.95 and 0.96) across various spatiotemporal expansion.

It is worth pointing out that although we used a simple model for the operator $\mathbf{D}$ to compensate for the spatiotemporal distortions of the electron pulse, the reconstruction approach presented here is universal: as long as the operator $\mathbf{D}$ is known (e.g., obtained through a point-by-point calibration), image distortions due to electron repulsions can be accounted for. In a TEM system, the electron pulse can propagate a significant distance, with several crossovers in the imaging/projection lens system. For more precise reconstruction, these effects should be accounted for via elaborate studies of the electron trajectories. Nevertheless, for moderate electron densities in a pulse, the Gaussian model used here, as a first order approximation to model $\mathbf{D}$, should be sufficient.

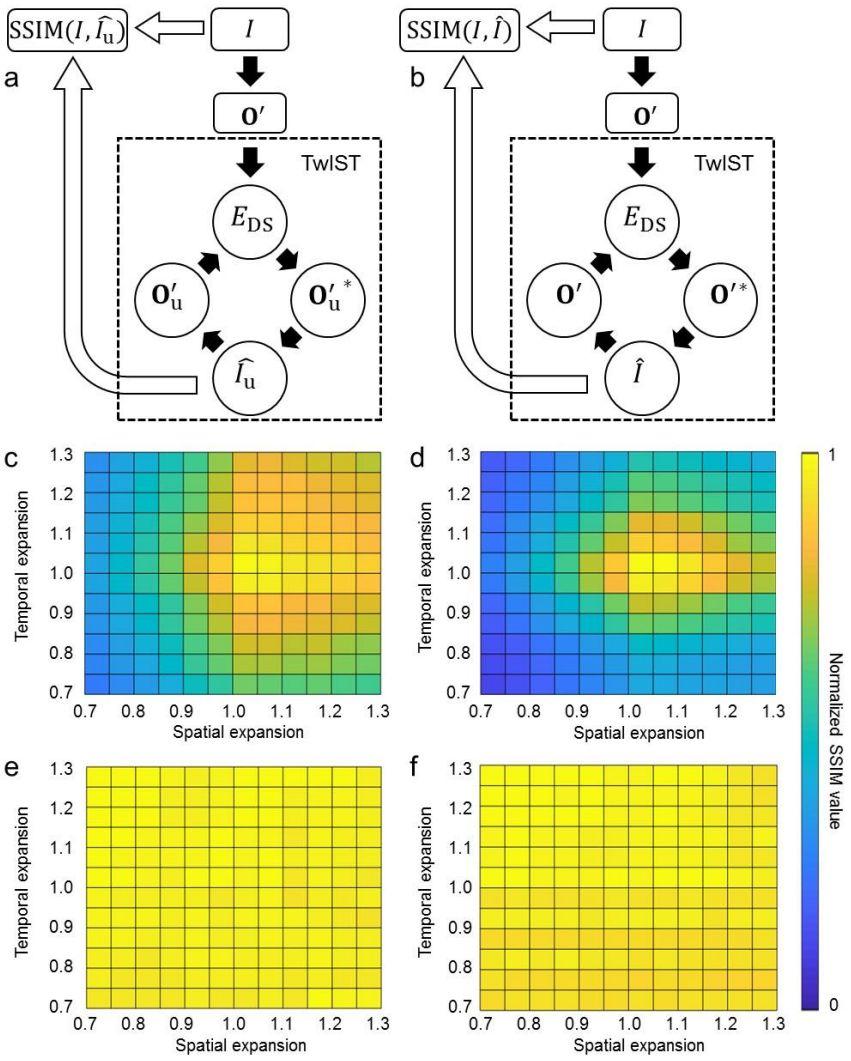


Figure 6.4 Comparison between the distortion and correction in DS-CUTEM. (a), Flowchart of reconstruction without distortion correction. (b), Flowchart of reconstruction with distortion correction. (c), Normalized SSIM values of reconstructed datacubes of the dipole scene without distortion correction. (d), As (c), but shows results for the cantilever scene. (e), Normalized SSIM values of reconstructed datacubes of dipole scene with distortion correction. (f), As (e), but shows results for the cantilever scene.

6.2.6 Feasibility demonstration

We leveraged the proposed methodology in reconstructing the two transient scenes. The represented frames of the dipole scene (as the ground truth) are shown in Fig. 6.5(a). The corresponding frames reconstructed by CUTEM and DS-CUTEM are presented in Figs. 6.5(b) and (c), respectively. A frame-to-frame comparison of the two reconstructed datacubes with the ground truth is shown in Media 1. Both CUTEM and DS-CUTEM significantly removed the noises presented in the dipole scene, demonstrated by a clean background in each reconstructed frame. Figure 6.5(d) illustrates the change in normalized average intensity across all frames. The

reconstruction results of CUTEM and DS-CUTEM are consistent with the ground truth. For a more detailed analysis, in Fig. 6.5(e), we plot the intensity profiles of a diagonal line in the ground truth and the reconstructed results [marked by the red solid line in Frame 15 in Figs. 6.5(a)–(c)]. The reconstruction by both CUTEM and DS-CUTEM well complies with the ground truth.

Six representative frames of the cantilever scene (as the ground truth) and its reconstructed images are shown in Figs. 6.6(a)–(c). Similar to the dipole scene, the noise reduction is clearly represented in Figs. 6.6(b) and (c), and the reconstruction is in good agreement with the ground truth [Fig. 6.6(d)]. Figure 6(e) shows the intensity profiles of a selected horizontal line in the ground truth and reconstructed results [marked by the red solid line in Frame 5 in Figs. 6.6(a)–(c)], which again demonstrates that the good agreement between the reconstructed results with the ground truth.

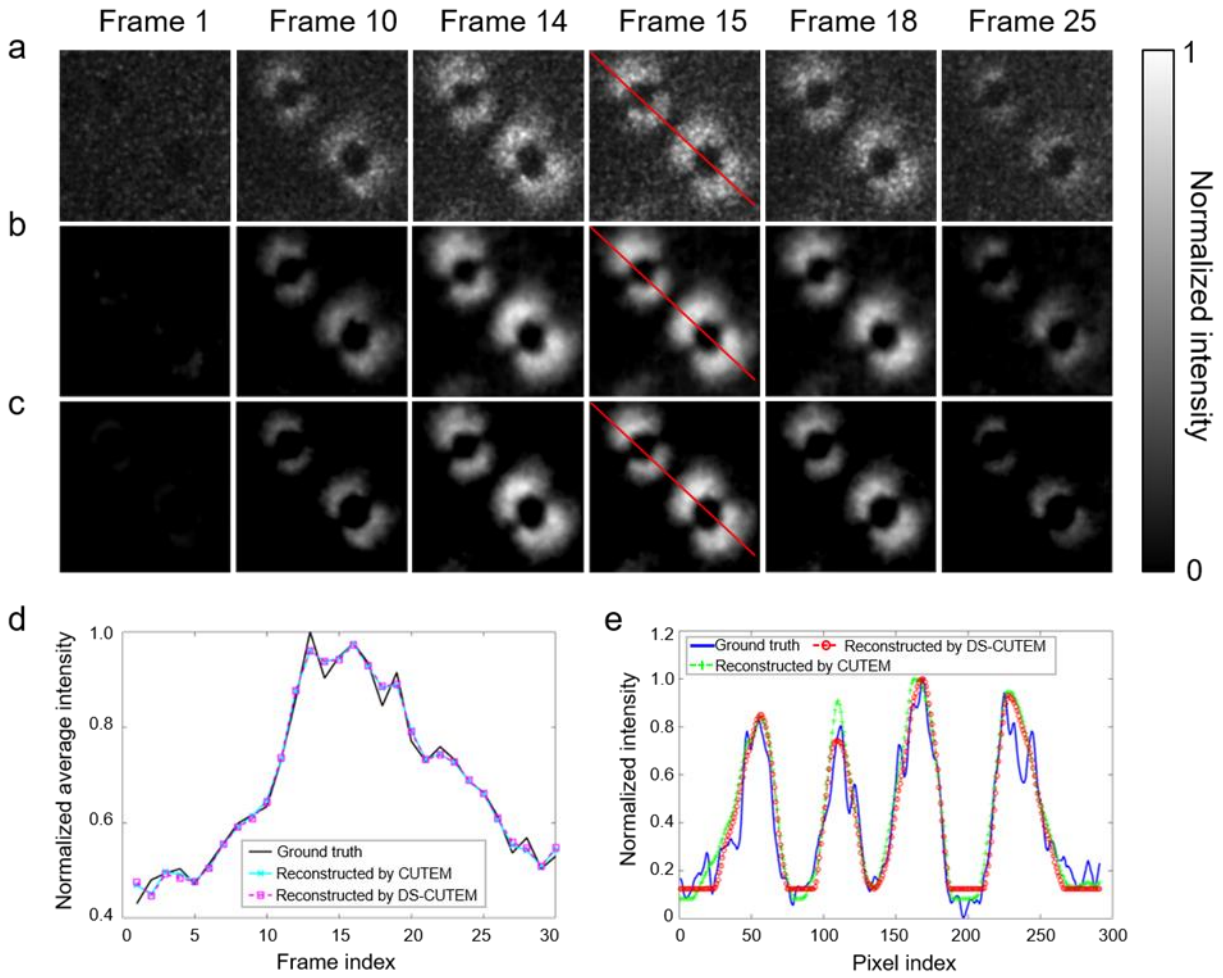


Figure 6.5 Reconstruction of the dipole scene using CUTEM and DS-CUTEM. (a), Representative frames of dipole scene as the ground truth [adapted from [268] with permission]. (b) and (c), Corresponding representative frames of reconstruction using CUTEM and DS-CUTEM, respectively. (d), Normalized average intensity across all frames of dipole scene and the two reconstructed results. (e), Comparison of normalized intensity profiles of a diagonal line in the dipole scene and the reconstruction results [marked by the red solid line in Frame 15 in (a)–(c)].

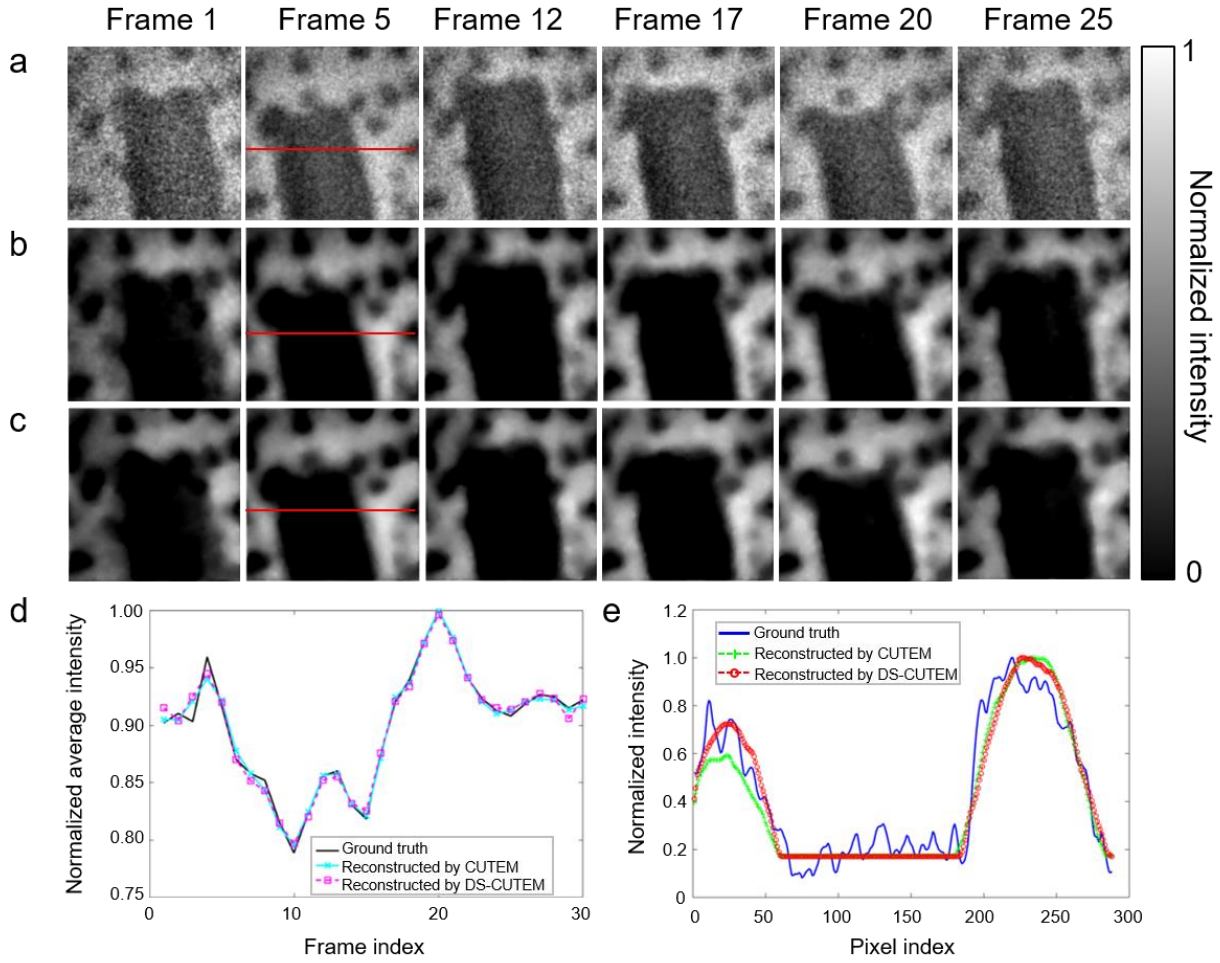


Figure 6.6 Reconstruction of the cantilever scene using CUTEM and DS-CUTEM. (a), Representative frames of the cantilever scene as the ground truth [adapted from [269] with permission]. (b) and (c), Corresponding representative frames of reconstruction using CUTEM and DS-CUTEM, respectively. (d), Change in normalized average intensity across all frames of cantilever scene and the two reconstructed results. (e), Comparison of normalized intensity profiles of a horizontal line in the cantilever scene and the reconstruction results [marked by the red solid line in Frame 5 in (a)–(c)].

6.3 Discussion and Conclusions

Our analytical models and simulation have shown that the single-shot ultrafast electronic imaging with sub-nanosecond temporal resolution could be realized by integrating CS-aided ultrafast imaging modalities to laser-assisted TEM. In principle, through hardware modifications detailed

in Fig. 6.1, CUTEM and DS-CUTEM will be able to record a single spatiotemporally modulated image in measurement and time-resolved frames through reconstruction. Moreover, compared with the previous methods [265, 266], both proposed configurations in our work encode the transient scene with a static mask. Consequently, no mechanical movement will be induced during the image acquisition, which would improve the repeatability and accuracy in calibration and experiments. In practice, to generate the image with high SNRs, nanosecond pulsed electron beams with more than 10 million electrons will be utilized. Such electron pulses can be generated using the photoelectric effect and powerful lasers. Temporal shearing and image recording needed for CUTEM and DS-CUTEM can be achieved with electronics with a rise time in the order of hundreds of picoseconds to nanoseconds [264]. At Institut National de la Recherche Scientifique (INRS), a TEM system using these technologies is under construction.

In summary, we have presented the design and simulation studies of CUTEM and DS-CUTEM—two CS-aided single-shot real-time ultrafast electron imaging methods. Transient scenes experimentally captured by laser-assisted TEM were used to test their fidelity. Different types of mask patterns and encoding pixel sizes were examined to optimize reconstructed image quality for the given SNR. The reconstruction results with both electronic distortion and corresponding correction processes were also provided. Based on these investigations, high-quality reconstruction of both transient scenes have been demonstrated in simulation. Our future work will be to apply both proposed methods in the advanced TEM infrastructure at INRS for imaging non-repetitive ultrafast events, such as structure dynamics of superheated nanoparticles [281] and phase transition in crystals [282, 283] at sub-nanosecond temporal resolution and nanometer real-space resolution. In addition, we plan to graft the concept of computational tomography (CT) into the spatiotemporal domain [90]. By implementing the multiple-angle projections in the proposed configurations and by leveraging various CT reconstruction algorithms, the spatial and temporal resolutions of the CUTEM and DS-CUTEM systems could be further improved.

Acknowledgements

The authors thank Dr. Jingdan Liu for close reading of the manuscript. This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada via programs of Discovery [RGPIN-2017-05959 (to J.L.) and RGPIN-2015-06133 (to A.Y.)], Discovery Accelerator Supplement [RGPAS-507845-2017 (to J.L.)], and Canada Research Chair (to A.Y.),

and by Fonds de Recherche du Québec–Nature et Technologies (FRQNT) via the program of Établissement de Nouveaux Chercheurs Universitaires [2019-NC-252960 (to J.L.) and A.Y.] and Projets de recherché en équipe (to A.Y.).

7 CONCLUSIONS AND PROSPECTIVES

7.1 Conclusions

Higher temporal resolution has been extensively pursued in photography for decades. CUP has attracted the tremendous interest of scientists by its fascinating merits in imaging speed, achieving novel capabilities that one could not with traditional imaging methods. Despite many advantages, the usage of an optoelectronic streak camera makes CUP suitable for niche applications, in which the system cost, size, and the QE of photocathode are negligible. Fortunately, the concept of CUP is generic. Inspired by CUP, we developed COSUP with merits such as cost-efficiency, compactness, and high sensitivity. We further developed COSUP to SPLIT. Using this system, we have demonstrated new applications in optical thermometry with wide-field and fast temperature sensing capability. Moreover, leveraging the rapid advancement of deep learning, we developed the SMART-COSUP system that has achieved fast end-to-end CS reconstruction in the task of high-speed object tracking. Furthermore, bringing the concept of CUP into the electronic imaging domain generates the configurations of CUTEM by combining streaking imaging with TEM. The proposed analytical model and simulation demonstrated the features of CUTEM and its variant possess nanometer spatial resolution and sub-nanosecond temporal resolution simultaneously.

The specific conclusions are as follows. First, COSUP pertains to single-shot 2D ultrahigh-speed passive optical imaging. Unlike the optoelectronic sweeping in CUP, COSUP features the optical streaking using a GS in the 4*f* imaging system. As a result, COSUP achieves tunable imaging speeds of up to 1.5 Mfps, which is approximately three orders of magnitude higher than the state-of-art in imaging speed of CTI [19-21] with an off-the-shelf CMOS camera. Furthermore, the COSUP system is capable of reaching a pixel count of 0.5 megapixels in each frame, which gives the space-bandwidth-time product of up to 750 gigapixels per second. The COSUP system can reconstruct a sequence depth of up to 500 frames in a single measurement, which dramatically increases the on-chip storage capacity. The transmission of single laser pulses through a mask and the shape and position of tracking of a fast-moving object in real time has been used to demonstrate COSUP's ultra-high-speed imaging capability.

Second, the SPLIT system has been developed for wide-field dynamic temperature sensing in real time. SPLIT, adapted from the COSUP system, compressively records the photoluminescence emission over a 2D FOV in two views. Then, the dual-view PnP-ADMM algorithm reconstructs spatially resolved intensity decay traces, from which a photoluminescence

lifetime distribution and the corresponding temperature map are extracted. Used with core/shell $NaGdF_4$:Er^{3+}, Yb^{3+}/$NaGdF_4$ UCNPs, SPLIT has enabled temperature mapping with high sensitivity for both green and red upconversion emission bands with a 20-μm spatial resolution in a 1.5×1.5 mm^2 FOV at a video rate of 20 Hz. SPLIT is demonstrated in longitudinal temperature monitoring of a phantom beneath fresh chicken tissue. SPLIT is also applied to dynamic single-cell temperature mapping of a moving single-layer onion epidermis sample.

Third, we have developed the S2V-AE for fast and high-quality video reconstruction from a single compressively acquired snapshot. This new DNN has facilitated the development of the SMART-COSUP system, which has demonstrated single-shot ultrahigh-speed imaging of transient events in both macroscopic and microscopic imaging at up to 20 kfps with a real-time reconstructed video size of $(N_x, N_y, N_t) = (256, 256, 100)$. This system has been applied to multiple-particle tracking.

Fourth, we have presented the design and simulation studies of CUTEM and DS-CUTEM—two CS-aided single-shot real-time ultrafast electron imaging methods. Transient scenes experimentally captured by laser-assisted TEM were used to test their fidelity. Different types of mask patterns and encoding pixel sizes were examined to optimize reconstructed image quality for the given SNR. The reconstruction results with both electronic distortion and corresponding correction processes were also provided. Based on these investigations, high-quality reconstruction of both transient scenes has been demonstrated in simulation.

7.2 Prospectives

As a universal imaging platform, COSUP can achieve a scalable spatial resolution by coupling with different front optics in microscopes and telescopes. Moreover, although not demonstrated in this work, COSUP can be easily applied to other CCD or CMOS cameras according to specific studies. For instance, the integration of an EMCCD camera in the COSUP system will enable high-sensitivity optical neuroimaging of an action potential propagating at tens of meters per second [102] under microscopic settings [103]. As another example, an infrared-camera-based COSUP system could enable wide-field temperature sensing in deep tissue using nanoparticles [104]. In summary, by leveraging the advantages of off-the-shelf cameras and sensors, COSUP is expected to find widespread applications in both fundamental and applied sciences.

SPLIT offers a versatile PLI temperature-sensing platform. In materials characterization, it could be used in the stress analysis of metal fatigue in turbine blades [157]. In biomedicine, it could be implemented for accurate sub-cutaneous temperature monitoring for theranostics of skin diseases (e.g., micro-melanoma) [158, 159]. SPLIT's microscopic temperature mapping ability could also be exploited for the studies of temperature-regulated cellular signaling [160]. Finally, the operation of SPLIT could be extended to Stokes emission in lanthanide-doped nanoparticles and spectrally resolved temperature mapping.

Despite being demonstrated only with the SMART-COSUP system, the S2V-AE could be easily extended to other modalities in CTI [196] and single-shot hyperspectral imaging [247]. Moreover, by implementing the variational AE [248], the dependence of the encoder on the sensing matrix could be further reduced. SMART-COSUP's ability to track multiple fast-moving particles in a wide field may enable new applications in particle imaging velocimetry [249] and flow cytometry [250].

Single-shot CUTEM could be applied in the advanced TEM infrastructure at INRS for imaging non-repetitive ultrafast events, such as structure dynamics of superheated nanoparticles [281] and phase transition in crystals [282, 283] at sub-nanosecond temporal resolution and nanometer real-space resolution. In addition, we plan to graft the concept of computational tomography (CT) into the spatiotemporal domain [90]. By implementing the multiple-angle projections in the proposed configurations and by leveraging various CT reconstruction algorithms, the spatial and temporal resolutions of the CUTEM and DS-CUTEM systems could be further improved. All of these topics are promising research directions in the future.

REFERENCES

1. Jingzhen, L. Assessing criterion of high speed photography temporal resolution. in 18th Intl Congress on High Speed Photography and Photonics. 1989. International Society for Optics and Photonics.
2. Munn, O. and A. Beach, A horse's motion scientifically determined. Scientific American, 1878. **39**(16): p. 241.
3. Mach, E. and P. Salcher, Photographische Fixirung der durch Projectile in der Luft eingeleiteten Vorgänge. 1887: K. k. Hof-u. Staatsdruckerei.
4. Wang, P., J. Liang, and L.V. Wang, Single-shot ultrafast imaging attaining 70 trillion frames per second. Nature communications, 2020. **11**(1): p. 1-9.
5. Adrian, R.J. and J. Westerweel, Particle image velocimetry. 2011: Cambridge university press.
6. Medhi, M., A. Dandautiya, and J.L. Raheja, Real-time video surveillance based structural health monitoring of civil structures using artificial neural network. Journal of Nondestructive Evaluation, 2019. **38**(3): p. 1-16.
7. Funatsu, R., et al., 8K 240-Hz full-resolution high-speed camera and slow-motion replay server systems. SMPTE Motion Imaging Journal, 2019. **128**(3): p. 44-49.
8. Yang, W. and R. Yuste, In vivo imaging of neural activity. Nature methods, 2017. **14**(4): p. 349-359.
9. Liu, X., et al., Fast wide-field upconversion luminescence lifetime thermometry enabled by single-shot compressed ultrahigh-speed imaging. Nature communications, 2021. **12**(1): p. 1-9.
10. Dong, M., S. Husale, and O. Sahin, Determination of protein structural flexibility by microsecond force spectroscopy. Nature nanotechnology, 2009. **4**(8): p. 514-517.
11. Cocker, T.L., et al., Tracking the ultrafast motion of a single molecule by femtosecond orbital imaging. Nature, 2016. **539**(7628): p. 263-267.
12. Liu, X., et al., Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: Analytical modeling and simulation. Micron, 2019. **117**: p. 47-54.
13. Lau, A.K., et al., Optical time stretch for high-speed and high-throughput imaging—from single-cell to tissue-wide scales. IEEE Journal of Selected Topics in Quantum Electronics, 2015. **22**(4): p. 89-103.
14. Etoh, T.G., et al., Toward one giga frames per second—evolution of in situ storage image sensors. Sensors, 2013. **13**(4): p. 4640-4658.
15. Suzuki, M., et al., A preliminary chip evaluation toward over 50Mfps burst global shutter stacked CMOS image sensor. Electronic Imaging, 2018. **2018**(11): p. 398-1-398-4.
16. Liu, X., et al., Single-shot compressed optical-streaking ultra-high-speed photography. Optics letters, 2019. **44**(6): p. 1387-1390.
17. Liang, J. and L.V. Wang, Single-shot ultrafast optical imaging. Optica, 2018. **5**(9): p. 1113-1127.
18. Chin, C.T., et al., Brandaris 128: A digital 25 million frames per second camera with 128 highly sensitive frames. Review of Scientific Instruments, 2003. **74**: p. 5026-5034.
19. Gao, L., et al., Single-shot compressed ultrafast photography at one hundred billion frames per second. Nature, 2014. **516**(7529): p. 74-77.
20. Liang, J., et al., Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse. Science advances, 2017. **3**(1): p. e1601814.
21. Liang, J., L. Zhu, and L.V. Wang, Single-shot real-time femtosecond imaging of temporal focusing. Light: Science & Applications, 2018. **7**(1): p. 1-10.

22. Eldar, Y.C. and G. Kutyniok, Compressed sensing: theory and applications. 2012: Cambridge university press.
23. Baker, R. and B. Johnson, Sweep circuit design for a picosecond streak camera. Measurement Science and Technology, 1994. **5**(4): p. 408.
24. Xie, C., et al., In-situ diagnostic of femtosecond probes for high resolution ultrafast imaging. arXiv preprint arXiv:2102.05108, 2021.
25. Xie, C., et al., In-situ diagnostic of femtosecond laser probe pulses for high resolution ultrafast imaging. Light: Science & Applications, 2021. **10**(1): p. 1-13.
26. Satoh, D., et al., Ultrafast pump-probe microscopic imaging of femtosecond laser-induced melting and ablation in single-crystalline silicon carbide. Applied Physics A, 2020. **126**(10): p. 1-8.
27. Fan, L., et al., Real-time observation and control of optical chaos. Science advances, 2021. **7**(3): p. eabc8448.
28. Jing, J.C., X. Wei, and L.V. Wang, Spatio-temporal-spectral imaging of non-repeatable dissipative soliton dynamics. Nature communications, 2020. **11**(1): p. 1-9.
29. Lucas, A., et al., Using deep neural networks for inverse problems in imaging: beyond analytical methods. IEEE Signal Processing Magazine, 2018. **35**(1): p. 20-36.
30. Kreizer, M., D. Ratner, and A. Liberzon, Real-time image processing for particle tracking velocimetry. Experiments in fluids, 2010. **48**(1): p. 105-110.
31. Zewail, A.H. and J.M. Thomas, 4D electron microscopy: imaging in space and time. 2009: World Scientific.
32. Musumeci, P., D. Cesar, and J. Maxson, Double-shot MeV electron diffraction and microscopy. Structural Dynamics, 2017. **4**(4): p. 044025.
33. Feist, A., et al., Ultrafast transmission electron microscopy using a laser-driven field emitter: Femtosecond resolution with a high coherence electron beam. Ultramicroscopy, 2017. **176**: p. 63-73.
34. Mait, J.N., G.W. Euliss, and R.A. Athale, Computational imaging. Advances in Optics and Photonics, 2018. **10**(2): p. 409-483.
35. Sliusarev, G.G., Aberration and optical design theory. Bristol, 1984.
36. Lohmann, A.W., et al., Space–bandwidth product of optical signals and systems. JOSA A, 1996. **13**(3): p. 470-473.
37. Ihrke, I., J. Restrepo, and L. Mignard-Debise, Principles of light field imaging: Briefly revisiting 25 years of research. IEEE Signal Processing Magazine, 2016. **33**(5): p. 59-69.
38. Zheng, G., R. Horstmeyer, and C. Yang, Wide-field, high-resolution Fourier ptychographic microscopy. Nature photonics, 2013. **7**(9): p. 739-745.
39. Yuan, X., D.J. Brady, and A.K. Katsaggelos, Snapshot compressive imaging: Theory, algorithms, and applications. IEEE Signal Processing Magazine, 2021. **38**(2): p. 65-88.
40. Gibson, G.M., et al., Real-time imaging of methane gas leaks using a single-pixel camera. Optics express, 2017. **25**(4): p. 2998-3005.
41. Stantchev, R.I., et al., Real-time terahertz imaging with a single-pixel detector. Nature communications, 2020. **11**(1): p. 1-8.
42. Liang, J., et al., Spatially Fourier-encoded photoacoustic microscopy using a digital micromirror device. Optics letters, 2014. **39**(3): p. 430-433.
43. Wu, D., et al., Imaging biological tissue with high-throughput single-pixel compressive holography. Nature Communications, 2021. **12**(1): p. 1-12.
44. Candès, E.J. and M.B. Wakin, An introduction to compressive sampling. IEEE signal processing magazine, 2008. **25**(2): p. 21-30.
45. Xu, Z.-H., et al., 1000 fps computational ghost imaging using LED-based structured illumination. Optics express, 2018. **26**(3): p. 2427-2434.
46. Liang, J., et al., Encrypted three-dimensional dynamic imaging using snapshot time-of-flight compressed ultrafast photography. Scientific reports, 2015. **5**(1): p. 1-10.

47. Zhu, L., et al., Space-and intensity-constrained reconstruction for compressed ultrafast photography. Optica, 2016. **3**(7): p. 694-697.
48. Yang, C., et al., Compressed ultrafast photography by multi-encoding imaging. Laser Physics Letters, 2018. **15**(11): p. 116202.
49. Yang, C., et al., Optimizing codes for compressed ultrafast photography by the genetic algorithm. Optica, 2018. **5**(2): p. 147-151.
50. Lu, Y., et al., Compressed ultrafast spectral-temporal photography. Physical review letters, 2019. **122**(19): p. 193904.
51. Yang, C., et al., Improving the image reconstruction quality of compressed ultrafast photography via an augmented Lagrangian algorithm. Journal of Optics, 2019. **21**(3): p. 035703.
52. Lai, Y., et al., Single-Shot Ultraviolet Compressed Ultrafast Photography. Laser & Photonics Reviews, 2020. **14**(10): p. 2000122.
53. Liang, J., et al., Single-shot stereo-polarimetric compressed ultrafast photography for light-speed observation of high-dimensional optical transients with picosecond resolution. Nature communications, 2020. **11**(1): p. 1-10.
54. Ma, Y., X. Feng, and L. Gao, Deep-learning-based image reconstruction for compressed ultrafast photography. Optics letters, 2020. **45**(16): p. 4400-4403.
55. Qi, D., et al., Single-shot compressed ultrafast photography: a review. Advanced Photonics, 2020. **2**(1): p. 014003.
56. Yang, C., et al., Hyperspectrally compressed ultrafast photography. Physical review letters, 2020. **124**(2): p. 023902.
57. Yao, J., et al., Multichannel-coupled compressed ultrafast photography. Journal of Optics, 2020. **22**(8): p. 085701.
58. Ding, P., et al., Single-shot spectral-volumetric compressed ultrafast photography. Advanced Photonics, 2021. **3**(4): p. 045001.
59. Qi, D., et al., 100-Trillion-Frame-per-Second Single-Shot Compressed Ultrafast Photography via Molecular Alignment. Physical Review Applied, 2021. **15**(2): p. 024051.
60. Reddy, D., A. Veeraraghavan, and R. Chellappa. P2C2: Programmable pixel compressive camera for high speed imaging. in CVPR 2011. 2011. IEEE.
61. Liu, D., et al., Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging. IEEE transactions on pattern analysis and machine intelligence, 2013. **36**(2): p. 248-260.
62. Llull, P., et al., Coded aperture compressive temporal imaging. Optics express, 2013. **21**(9): p. 10526-10545.
63. Koller, R., et al., High spatio-temporal resolution video with compressed sensing. Optics express, 2015. **23**(12): p. 15992-16007.
64. Feng, W., et al., Per-pixel coded exposure for high-speed and high-resolution imaging using a digital micromirror device camera. Sensors, 2016. **16**(3): p. 331.
65. Jalali, S. and X. Yuan, Snapshot compressed sensing: Performance bounds and algorithms. IEEE Transactions on Information Theory, 2019. **65**(12): p. 8005-8024.
66. El Mahdaoui, A., A. Ouahabi, and M.S. Moulay. Image Recovery Using Total Variation Minimization on Compressive Sensing. in 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA). 2019. IEEE.
67. Kai Zhang1, et al., Learning Deep CNN Denoiser Prior for Image Restoration. 2017.
68. Bioucas-Dias, J.M. and M.A. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image processing, 2007. **16**(12): p. 2992-3004.
69. Boyd, S., et al., Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning, 2011. **3**(1): p. 1-122.

70. Iliadis, M., L. Spinoulas, and A.K. Katsaggelos, Deep fully-connected networks for video compressive sensing. Digital Signal Processing, 2018. **72**: p. 9-18.
71. Duran, J., B. Coll, and C. Sbert, Chambolle's projection algorithm for total variation denoising. Image processing on Line, 2013. **3**: p. 311-331.
72. Chan, S.H., Plug-and-play ADMM for Image Restoration: Fixed Point Convergence and Applications. 2016.
73. Dabov, K., et al. BM3D Image Denoising with Shape-Adaptive Principal Component Analysis. in SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations. 2009. Saint Malo, France.
74. Qiao, M., et al., Deep learning for video compressive sensing. APL Photonics, 2020. **5**(3): p. 030801.
75. Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in International Conference on Medical image computing and computer-assisted intervention. 2015. Springer.
76. Liu, X., et al., Single-shot real-time compressed ultrahigh-speed imaging enabled by a snapshot-to-video autoencoder. Photonics Research, 2021. **9**(12): p. 2464-2474.
77. Fuller, P.W.W., An introduction to high speed photography and photonics. Imaging Science Journal, 2009. **57**(6): p. 293-302.
78. Howard, S.S., et al., Frequency-multiplexed in vivo multiphoton phosphorescence lifetime microscopy. Nature Photonics, 2013. **7**(1): p. 33-37.
79. Yang, W. and R. Yuste, In vivo imaging of neural activity (vol 14, pg 349, 2017). Nature Methods, 2017. **14**(7): p. 349-359.
80. Dong, M.D., S. Husale, and O. Sahin, Determination of protein structural flexibility by microsecond force spectroscopy. Nature Nanotechnology, 2009. **4**(8): p. 514-517.
81. Ehn, A., et al., FRAME: femtosecond videography for atomic and molecular dynamics. Light-Science & Applications, 2017. **6**: p. e17045.
82. Goda, K., K.K. Tsia, and B. Jalali, Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. Nature, 2009. **458**(7242): p. 1145-1149.
83. Goda, K., et al., High-throughput single-microparticle imaging flow analyzer. Proc. Natl. Acad. Sci. U.S.A, 2012. **109**(29): p. 11630-5.
84. Goda, K. and B. Jalali, Dispersive Fourier transformation for fast continuous single-shot measurements. Nature Photonics, 2013. **7**(2): p. 102-112.
85. Wu, J.L., et al., Ultrafast laser-scanning time-stretch imaging at visible wavelengths. Light-Science & Applications, 2017. **6**: p. e16196.
86. Tiwari, V., M.A. Sutton, and S.R. McNeill, Assessment of high speed imaging systems for 2D and 3D deformation measurements: Methodology development and validation. Experimental Mechanics, 2007. **47**(4): p. 561-579.
87. Suzuki, M., et al., A Preliminary Chip Evaluation toward Over 50Mfps Burst Global Shutter Stacked CMOS Image Sensor. Electronic Imaging, 2018. **2018**(11): p. 3981-3984.
88. Hunt, J., et al., Metamaterial Apertures for Computational Imaging. Science, 2013. **339**(6117): p. 310-313.
89. Donoho, D.L., Compressed sensing. IEEE Trans. Inf. Theory, 2006. **52**(4): p. 1289-1306.
90. Liang, J.Y. and L.H.W. Wang, Single-shot ultrafast optical imaging. Optica, 2018. **5**(9): p. 1113-1127.
91. Baraniuk, R.G., et al., Compressive Video Sensing. Ieee Signal Processing Magazine, 2017. **34**(1): p. 52-66.
92. Reddy, D., A. Veeraraghavan, and R. Chellappa, P2C2: Programmable pixel compressive camera for high speed imaging. CVPR 2011: p. 329-336.
93. Hitomi, Y., et al., Video from a Single Coded Exposure Photograph using a Learned Over-Complete Dictionary. 2011 Ieee International Conference on Computer Vision (Iccv), 2011: p. 287-294.

94. Llull, P., et al., Coded aperture compressive temporal imaging. Optics express, 2013. **21**(9): p. 10526-10545.
95. Mochizuki, F., et al., Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor. Optics express, 2016. **24**(4): p. 4155-4176.
96. Gao, L., et al., Single-shot compressed ultrafast photography at one hundred billion frames per second. Nature, 2014. **516**(7529): p. 74-77.
97. Liang, J.Y., et al., Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse. Science Advances, 2017. **3**(1): p. e1601814.
98. Liang, J.Y., L.R. Zhu, and L.H.V. Wang, Single-shot real-time femtosecond imaging of temporal focusing. Light-Science & Applications, 2018. **7**: p. 42.
99. Wu, D., et al., Fast frame scanning camera system for light-sheet microscopy. Applied Optics, 2015. **54**(29): p. 8632-8636.
100. Buckner, B.D. and D. L'Esperance, Digital synchroballistic schlieren camera for high-speed photography of bullets and rocket sleds. Optical Engineering, 2013. **52**(8): p. 083105.
101. Bioucas-Dias, J.M. and M.A.T. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE T Image Process, 2007. **16**(12): p. 2992-3004.
102. Chen, T.W., et al., Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature, 2013. **499**(7458): p. 295-300.
103. Mikami, H., L. Gao, and K. Goda, Ultrafast optical imaging technology: principles and applications of emerging methods. Nanophotonics, 2016. **5**(4): p. 497-509.
104. Jaque, D. and F. Vetrone, Luminescence nanothermometry. Nanoscale, 2012. **4**(15): p. 4301-4326.
105. Inada, N., et al., Temperature imaging using a cationic linear fluorescent polymeric thermometer and fluorescence lifetime imaging microscopy. Nat. Protoc., 2019. **14**.
106. Wood, M. and K. Ozanyan, Simultaneous temperature, concentration, and pressure imaging of water vapor in a turbine engine. IEEE Sens. J., 2015. **15**.
107. Obermayer, D. and C. Kappe, On the importance of simultaneous infrared/fiber-optic temperature monitoring in the microwave-assisted synthesis of ionic liquids. Org. Biomol. Chem., 2010. **8**.
108. Zhang, Z., J. Wang, and C. Chen, Near-infrared light-mediated nanoplatforms for cancer thermo-chemotherapy and optical imaging. Adv. Mater., 2013. **25**.
109. Chen, Z., Phosphorescent polymeric thermometers for in vitro and in vivo temperature sensing with minimized background interference. Adv. Funct. Mater., 2016. **26**.
110. Kurokawa, H., High resolution imaging of intracellular oxygen concentration by phosphorescence lifetime. Sci. Rep., 2015. **5**.
111. Datta, R., Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. J. Biomed. Opt., 2020. **25**.
112. Childs, P., J. Greenwood, and C. Long, Review of temperature measurement. Rev. Sci. Instrum., 2000. **71**.
113. Snyder, W., et al., Classification-based emissivity for land surface temperature measurement from space. Int. J. Remote Sens., 1998. **19**.
114. Gao, L., Single-cell photoacoustic thermometry. J. Biomed. Opt., 2013. **18**.
115. Suhling, K., Fluorescence lifetime imaging (FLIM): basic concepts and some recent developments. Med. Photonics, 2015. **27**.
116. Labrador-Páez, L., Reliability of rare-earth-doped infrared luminescent nanothermometers. Nanoscale, 2018. **10**.
117. Pickel, A., Apparent self-heating of individual upconverting nanoparticle thermometers. Nat. Commun., 2018. **9**.

118. Chelushkin, P.S. and S.P. Tunik, Phosphorescence Lifetime Imaging (PLIM): State of the Art and Perspectives, in Progress in Photon Science. 2019, Springer. p. 109-128.
119. Shen, Y., In vivo spectral distortions of infrared luminescent nanothermometers compromise their reliability. ACS nano, 2020. **14**.
120. Becker, W., Fluorescence lifetime imaging–techniques and applications. J. Microsc., 2012. **247**.
121. Zhang, H., Dual-emissive phosphorescent polymer probe for accurate temperature sensing in living cells and zebrafish using ratiometric and phosphorescence lifetime imaging microscopy. ACS Appl. Mater. Interfaces, 2018. **10**.
122. Gao, H., A simple yet effective AIE-based fluorescent nano-thermometer for temperature mapping in living cells using fluorescence lifetime imaging microscopy. Nanoscale Horiz., 2020. **5**.
123. Tan, M., Accurate in vivo nanothermometry through NIR-II lanthanide luminescence lifetime. Small, 2020. **16**.
124. Bolek, P., Ga-modified YAG: Pr3+ dual-mode tunable luminescence thermometers. Chem. Eng. J., 2021. **421**.
125. Maciejewska, K., A. Bednarkiewicz, and L. Marciniak, NIR Luminescence lifetime nanothermometry based on phonon assisted Yb3+-Nd3+ energy transfer. Nanoscale Adv., 2021. **3**.
126. Schlegel, G., Fluorescence decay time of single semiconductor nanocrystals. Phys. Rev. Lett., 2002. **88**.
127. Allison, S.W., Nanoscale thermometry via the fluorescence of YAG: Ce phosphor particles: measurements from 7 to 77 °C. Nanotechnology, 2003. **14**.
128. Benninger, R., Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. Anal. Chem., 2006. **78**.
129. Graham, E.M., Quantitative mapping of aqueous microfluidic temperature with sub-degree resolution using fluorescence lifetime imaging microscopy. Lab Chip, 2010. **10**.
130. Auzel, F., Upconversion and anti-stokes processes with f and d ions in solids. Chem. Rev., 2004. **104**.
131. Skripka, A., Spectral characterization of LiYbF4 upconverting nanoparticles. Nanoscale, 2020. **12**.
132. Jacques, S., Optical properties of biological tissues: a review. Phys. Med. Biol., 2013. **58**.
133. Vetrone, F., Temperature sensing using fluorescent nanothermometers. ACS Nano, 2010. **4**.
134. Brites, D.S.C., Thermometry at the nanoscale. Nanoscale, 2012. **4**.
135. Rostami, I., et al., Breakthroughs in medicine and bioimaging with up-conversion nanoparticles. Int. J. Nanomed., 2019. **14**.
136. Zhou, J., et al., Advances and challenges for fluorescence nanothermometry. Nat. Methods, 2020. **17**.
137. Qin, H., Tuning the upconversion photoluminescence lifetimes of NaYF4: Yb3+, Er3+ through lanthanide Gd3+ doping. Sci. Rep., 2018. **8**.
138. Howard, S., et al., Frequency-multiplexed in vivo multiphoton phosphorescence lifetime microscopy. Nat. Photonics, 2013. **7**.
139. Suhling, K., Wide-field TCSPC-based fluorescence lifetime imaging (FLIM). Microsc. SPIE proc., 2016. **9858**.
140. Hirvonen, L., et al., Photon counting phosphorescence lifetime imaging with TimepixCam. Rev. Sci. Instrum., 2017. **88**.
141. Sen, R., New luminescence lifetime macro-imager based on a Tpx3Cam optical camera. Biomed. Optics express, 2020. **11**.
142. Franke, R. and G. Holst, Frequency-domain fluorescence lifetime imaging system (pco. flim) based on a in-pixel dual tap control CMOS image sensor. SPIE proc., 2015. **9328**.

143. Xiong, B. and Q. Fang, Luminescence lifetime imaging using a cellphone camera with an electronic rolling shutter. Opt. Lett., 2020. **45**.
144. Liang, J., Punching holes in light: recent progress in single-shot coded-aperture optical imaging. Rep. Prog. Phys., 2020. **83**.
145. Liang, J., Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse. Sci. Adv., 2017. **3**.
146. Liu, Y., et al., Rank minimization for snapshot compressive imaging. IEEE Trans. Pattern Anal. Mach. Intell., 2018. **41**.
147. Yuan, X., et al. Plug-and-play algorithms for large-scale snapshot compressive imaging. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
148. Yuan, X., Compressive hyperspectral imaging with side information. IEEE J. Sel. Top. Signal Process, 2015. **9**.
149. Chan, S., X. Wang, and O. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. IEEE Trans. Comput. Imag., 2016. **3**.
150. Liang, J., L. Zhu, and L. Wang, Single-shot real-time femtosecond imaging of temporal focusing. Light Sci. Appl., 2018. **7**.
151. Liang, J., Single-shot stereo-polarimetric compressed ultrafast photography for light-speed observation of high-dimensional optical transients with picosecond resolution. Nat. Commun., 2020. **11**.
152. Wang, P., J. Liang, and L. Wang, Single-shot ultrafast imaging attaining 70 trillion frames per second. Nat. Commun., 2020. **11**.
153. May, P. and M. Berry, Tutorial on the acquisition, analysis, and interpretation of upconversion luminescence data. Methods Appl. Fluoresc., 2019. **7**.
154. Dos Santos, P., et al., Optical temperature sensing using upconversion fluorescence emission in Er 3+/Yb 3+-codoped chalcogenide glass. Applied physics letters, 1998. **73**(5): p. 578-580.
155. Miller, M. and J. Wright, Multiphonon and energy transfer relaxation in charge compensated crystals. J. Chem. Phys., 1979. **71**.
156. Liu, X., et al., Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: analytical modeling and simulation. Micron, 2019. **117**.
157. Wang, R., Thermomechanical fatigue experiment and failure analysis on a nickel-based superalloy turbine blade. Eng. Fail. Anal., 2019. **102**.
158. Jung, H., Organic molecule-based photothermal agents: an expanding photothermal therapy universe. Chem. Soc. Rev., 2018. **47**.
159. Shen, Y., Ag2S nanoheaters with multiparameter sensing for reliable thermal feedback during in vivo tumor therapy. Adv. Funct. Mater., 2020. **30**.
160. Wang, C., Determining intracellular temperature at single-cell level by a novel thermocouple method. Cell Res., 2011. **21**.
161. Liu, X., et al., Single-shot compressed optical-streaking ultra-high-speed photography. Opt. Lett., 2019. **44**.
162. Jiang, C., et al., High-speed dual-view band-limited illumination profilometry using temporally interlaced acquisition. Photonics Research, 2020. **8**(11): p. 1808-1817.
163. Inc., T.M. Register Images Using Registration Estimator App. 2020; Available from: https://www.mathworks.com/help/images/register-images-using-the-registration-estimator-app.html.
164. Antipa, N., et al., DiffuserCam: lensless single-exposure 3D imaging. Optica, 2018. **5**(1): p. 1-9.
165. Otsu, N., A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 1979. **9**(1): p. 62-66.

166. Boggs, P.T. and J.W. Tolle, Sequential quadratic programming. Acta numerica, 1995. **4**: p. 1-51.
167. Boyd, S., N. Parikh, and E. Chu, Distributed optimization and statistical learning via the alternating direction method of multipliers. 2011: Now Publishers Inc.
168. Dabov, K., et al. BM3D image denoising with shape-adaptive principal component analysis. in SPARS'09-Signal Processing with Adaptive Sparse Structured Representations. 2009.
169. Burger, H.C., C.J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D? in 2012 IEEE conference on computer vision and pattern recognition. 2012. IEEE.
170. function, R.
171. Chan, S.H., Performance analysis of plug-and-play ADMM: A graph signal processing perspective. IEEE Transactions on Computational Imaging, 2019. **5**(2): p. 274-286.
172. May, P.S. and M. Berry, Tutorial on the acquisition, analysis, and interpretation of upconversion luminescence data. Methods and applications in fluorescence, 2019. **7**(2): p. 023001.
173. Sillen, A. and Y. Engelborghs, The correct use of “average” fluorescence parameters. Photochemistry and photobiology, 1998. **67**(5): p. 475-486.
174. Boyer, J.-C., et al., Synthesis of colloidal upconverting NaYF4 nanocrystals doped with Er3+, Yb3+ and Tm3+, Yb3+ via thermal decomposition of lanthanide trifluoroacetate precursors. Journal of the American Chemical Society, 2006. **128**(23): p. 7444-7445.
175. Brites, C., A. Millán, and L. Carlos, Lanthanides in luminescent thermometry, in Handbook on the Physics and Chemistry of Rare Earths. 2016, Elsevier. p. 339-427.
176. Lin, L., et al., In vivo photoacoustic tomography of myoglobin oxygen saturation. Journal of biomedical optics, 2015. **21**(6): p. 061002.
177. Arakaki, L.S., D.H. Burns, and M.J. Kushmerick, Accurate myoglobin oxygen saturation by optical spectroscopy measured in blood-perfused rat muscle. Applied spectroscopy, 2007. **61**(9): p. 978-985.
178. Ma, Y., et al., High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells. Proceedings of the National Academy of Sciences, 2021. **118**(3).
179. Liu, L., et al. Fluorescence lifetime imaging microscopy using a streak camera. in Multiphoton Microscopy in the Biomedical Sciences XIV. 2014. International Society for Optics and Photonics.
180. Liu, X., et al., Fast fluorescence lifetime imaging techniques: A review on challenge and development. Journal of Innovative Optical Health Sciences, 2019. **12**(05): p. 1930003.
181. Wang, C., et al., Line scanning mechanical streak camera for phosphorescence lifetime imaging. Optics express, 2020. **28**(18): p. 26717-26723.
182. Kannan, M., et al., Fast, in vivo voltage imaging using a red fluorescent indicator. Nat. Methods, 2018. **15**(12): p. 1108-1116.
183. Sasaki, M., et al., Motion-picture recording of ultrafast behavior of polarized light incident at Brewster’s angle. Sci. Rep. , 2020. **10**(1): p. 1-7.
184. Poulin, P.R. and K.A. Nelson, Irreversible organic crystalline chemistry monitored in real time. Science, 2006. **313**(5794): p. 1756-1760.
185. Kondo, T., et al. A 3D stacked CMOS image sensor with 16Mpixel global-shutter mode and 2Mpixel 10000fps mode using 4 million interconnections. in Proc. IEEE, Symposium on VLSI Circuits. 2015. IEEE.
186. Etoh, T., et al. Toward 1Gfps: Evolution of ultra-high-speed image sensors-ISIS, BSI, multi-collection gates, and 3D-stacking. in Proc. IEEE, IEDM 2014. IEEE.
187. York, T., et al., Bioinspired polarization imaging sensors: from circuits and optics to signal processing algorithms and biomedical applications. Proc. IEEE 2014. **102**(10): p. 1450-1469.

188. Calvet, D., A new interface technique for the acquisition of multiple multi-channel high speed ADCs. IEEE Trans. Nucl. Sci. , 2008. **55**(5): p. 2592-2597.
189. Hejtmánek, M., G. Neue, and P. Voleš, Software interface for high-speed readout of particle detectors based on the CoaXPress communication standard. J. Instrum., 2015. **10**(06): p. C06011.
190. Barbastathis, G., A. Ozcan, and G. Situ, On the use of deep learning for computational imaging. Optica, 2019. **6**(8): p. 921-943.
191. Ehn, A., et al., FRAME: femtosecond videography for atomic and molecular dynamics. Light Sci. Appl., 2017. **6**(9): p. e17045-e17045.
192. Kakue, T., et al., Digital light-in-flight recording by holography by use of a femtosecond pulsed laser. IEEE J. Sel. Top. Quantum Electron., 2011. **18**(1): p. 479-485.
193. Li, Z., et al., Single-shot tomographic movies of evolving light-velocity objects. Nat. Commun., 2014. **5**(1): p. 1-12.
194. Lai, Y., et al., Single-Shot Ultraviolet Compressed Ultrafast Photography. Laser Photonics Rev.: p. 2000122.
195. Liang, J.Y., L.R. Zhu, and L.H.V. Wang, Single-shot real-time femtosecond imaging of temporal focusing. Light Sci. Appl., 2018. **7**: p. 1-10.
196. Liu, X.L., et al., Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: Analytical modeling and simulation. Micron, 2019. **117**: p. 47-54.
197. Wang, P., J.Y. Liang, and L.V. Wang, Single-shot ultrafast imaging attaining 70 trillion frames per second. Nat. Commun., 2020. **11**(1): p. 1-9.
198. Liang, J., et al., Single-shot stereo-polarimetric compressed ultrafast photography for light-speed observation of high-dimensional optical transients with picosecond resolution. Nat. Commun., 2020. **11**(1): p. 1-10.
199. Yang, C., et al., Hyperspectrally Compressed Ultrafast Photography. Phys. Rev. Lett., 2020. **124**(2): p. 023902.
200. Liang, J., et al., Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse. Sci. Adv., 2017. **3**(1): p. e1601814.
201. Jinyang, L., Punching holes in light: Recent progress in single-shot coded-aperture optical imaging. Rep. Prog. Phys. , 2020.
202. Yang, J., et al., Video compressive sensing using Gaussian mixture models. IEEE Trans. Image Process. , 2014. **23**(11): p. 4863-4878.
203. Wang, C., et al., Line scanning mechanical streak camera for phosphorescence lifetime imaging. Optics express, 2020. **28**(18): p. 26717-26723.
204. Liu, X.L., et al., Single-shot compressed optical-streaking ultra-high-speed photography. Opt. Lett., 2019. **44**(6): p. 1387-1390.
205. Llull, P., et al., Coded aperture compressive temporal imaging. Optics express, 2013. **21**(9): p. 10526-10545.
206. Koller, R., et al., High spatio-temporal resolution video with compressed sensing. Optics express, 2015. **23**(12): p. 15992-16007.
207. Reddy, D., A. Veeraraghavan, and R. Chellappa. P2C2: Programmable pixel compressive camera for high speed imaging. in Proc. IEEE, CVPR. 2011. IEEE.
208. Liu, Y., et al., Rank minimization for snapshot compressive imaging. IEEE Trans. Pattern Anal. Mach. Intell., 2018. **41**(12): p. 2990-3006.
209. Lucas, A., et al., Using Deep Neural Networks for Inverse Problems in Imaging Beyond analytical methods. IEEE Signal Process. Mag., 2018. **35**(1): p. 20-36.
210. Bioucas-Dias, J.M. and M.A. Figueiredo, A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Trans. Image Process 2007. **16**(12): p. 2992-3004.
211. Yang, C., et al., Improving the image reconstruction quality of compressed ultrafast photography via an augmented Lagrangian algorithm. J. Opt., 2019. **21**(3): p. 035703.

212. Hui, J., et al., Real-time intravascular photoacoustic-ultrasound imaging of lipid-laden plaque in human coronary artery at 16 frames per second. Scientific reports, 2017. **7**(1): p. 1-11.
213. Kreizer, M., D. Ratner, and A. Liberzon, Real-time image processing for particle tracking velocimetry. Exp. Fluids, 2010. **48**(1): p. 105-110.
214. LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. **521**(7553): p. 436-444.
215. Iliadis, M., L. Spinoulas, and A.K. Katsaggelos, Deep fully-connected networks for video compressive sensing. Digit. Signal Process., 2018. **72**: p. 9-18.
216. Yoshida, M., et al. Joint optimization for compressive video sensing and reconstruction under hardware constraints. in Proceedings of the European Conference on Computer Vision (ECCV). 2018.
217. Yang, C., et al., High-fidelity image reconstruction for compressed ultrafast photography via an augmented-Lagrangian and deep-learning hybrid algorithm. Photonics Research, 2021. **9**(2): p. B30-B37.
218. Zhang, A., et al., Single-shot compressed ultrafast photography based on U-net network. Optics express, 2020. **28**(26): p. 39299-39310.
219. Gardner, M.W. and S. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ., 1998. **32**(14-15): p. 2627-2636.
220. Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in Springer, MICCAI. 2015. Springer.
221. Yuan, X., BIRNAT: Bidirectional recurrent neural networks with adversarial training for video snapshot compressive imaging. ECVA, ECCV, 2020.
222. Tschannen, M., O. Bachem, and M. Lucic, Recent advances in autoencoder-based representation learning. arXiv preprint arXiv:1812.05069, 2018.
223. Nguyen, A., et al. Plug & play generative networks: Conditional iterative generation of images in latent space. in Proc. IEEE, CVPR. . 2017.
224. Larsen, A.B.L., et al. Autoencoding beyond pixels using a learned similarity metric. in International conference on machine learning. 2016. PMLR.
225. Vondrick, C., H. Pirsiavash, and A. Torralba. Generating videos with scene dynamics. in Adv. Neural Inf. Process Syst. 2016.
226. Tulyakov, S., et al. Mocogan: Decomposing motion and content for video generation. in Proc. IEEE, CVPR. . 2018.
227. Ohnishi, K., et al., Hierarchical video generation from orthogonal information: Optical flow and texture. arXiv preprint arXiv:1711.09618, 2017.
228. Plchot, O., et al. Audio enhancing with DNN autoencoder for speaker recognition. in Proc. IEEE, ICASSP. . 2016. IEEE.
229. Yu, J., X. Zheng, and S. Wang, A deep autoencoder feature learning method for process pattern recognition. J. Process Control, 2019. **79**: p. 1-15.
230. Ranzato, M.A., et al. Efficient learning of sparse representations with an energy-based model. in NIPS, Adv Neural Inf. Process Syst. 2007.
231. Vincent, P., et al., Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn Res., 2010. **11**(12).
232. Ma, X. and E. Hovy, End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint arXiv:1603.01354, 2016.
233. Ioffe, S. and C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in Proceedings of the 32nd International Conference on Machine Learning, B. Francis and B. David, Editors. 2015, PMLR: Proceedings of Machine Learning Research. p. 448--456.
234. Nair, V. and G.E. Hinton. Rectified linear units improve restricted boltzmann machines. in ICML. 2010.

235. Zhang, Z. and M. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. in NIPS, Adv. Neural Inf. Process Syst. . 2018.
236. Krogh, A. and J.A. Hertz. A simple weight decay can improve generalization. in NIPS, Adv Neural Inf Process Syst. 1992.
237. Kingma, D.P. and J. Ba, Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
238. Deng, L., The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE Signal Process. Mag., 2012. **29**(6): p. 141-142.
239. Wang, Z., et al., Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. , 2004. **13**(4): p. 600-612.
240. MathWorks. Register Images Using Registration Estimator App. Available from: https://www.mathworks.com/help/images/register-images-using-the-registration-estimator-app.html.
241. Hartley, R. and A. Zisserman, Multiple view geometry in computer vision. 2003: Cambridge university press.
242. Lin, Z., et al. Pacgan: The power of two samples in generative adversarial networks. in NIPS, Adv. Neural Inf. Process Syst. . 2018.
243. Jolicoeur-Martineau, A., The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734, 2018.
244. Neyshabur, B., S. Bhojanapalli, and A. Chakrabarti, Stabilizing GAN training with multiple random projections. arXiv preprint arXiv:1705.07831, 2017.
245. Albuquerque, I., et al., Multi-objective training of Generative Adversarial Networks with multiple discriminators, in PMLR, International Conference on Machine Learning C. Kamalika and S. Ruslan, Editors. 2019, PMLR: Proceedings of Machine Learning Research. p. 202--211.
246. Pascanu, R., T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. in International conference on machine learning. 2013. PMLR.
247. Miao, X., X. Yuan, and P. Wilford. Deep learning for compressive spectral imaging. in Proc. OSA, Digital Holography and Three-Dimensional Imaging. 2019. Optical Society of America.
248. Pu, Y., et al. Variational autoencoder for deep learning of images, labels and captions. in NIPS, Adv Neural Inf Process Syst. 2016.
249. Ten Cate, A., et al., Particle imaging velocimetry experiments and lattice-Boltzmann simulations on a single sphere settling under gravity. Physics of Fluids, 2002. **14**(11): p. 4012-4025.
250. Nitta, N., et al., Intelligent image-activated cell sorting. Cell, 2018. **175**(1): p. 266-276. e13.
251. Zewail, A.H., The birth of molecules. Scientific American, 1990. **263**(6): p. 76-83.
252. Zewail, A.H., Femtochemistry: Atomic-scale dynamics of the chemical bond using ultrafast lasers (Nobel Lecture). Angewandte Chemie International Edition, 2000. **39**(15): p. 2586-2631.
253. Elsayed-Ali, H., et al., Time-resolved observation of electron-phonon relaxation in copper. Physical Review Letters, 1987. **58**(12): p. 1212.
254. Cavalleri, A., et al., Evidence for a structurally-driven insulator-to-metal transition in VO 2: A view from the ultrafast timescale. Physical Review B, 2004. **70**(16): p. 161102.
255. Sjodin, T., H. Petek, and H.-L. Dai, Ultrafast carrier dynamics in silicon: A two-color transient reflection grating study on a (111) surface. Physical review letters, 1998. **81**(25): p. 5664.
256. Bostanjoglo, O., J. Kornitzky, and R. Tornow, High-speed electron microscopy of laser-induced vaporization of thin films. Journal of applied physics, 1991. **69**(4): p. 2581-2583.

257. Bostanjoglo, O. and M. Weingärtner, Pulsed photoelectron microscope for imaging laser-induced nanosecond processes. Review of scientific instruments, 1997. **68**(6): p. 2456-2460.
258. Bostanjoglo, O., High-speed electron microscopy. Advances in imaging and electron physics, 2002. **121**: p. 1-51.
259. LaGrange, T., et al., Single-shot dynamic transmission electron microscopy. Applied Physics Letters, 2006. **89**(4): p. 044105.
260. Kim, J.S., et al., Imaging of transient structures using nanosecond in situ TEM. Science, 2008. **321**(5895): p. 1472-1475.
261. Yurtsever, A., J.S. Baskin, and A.H. Zewail, Entangled nanoparticles: Discovery by visualization in 4D electron microscopy. Nano letters, 2012. **12**(9): p. 5027-5032.
262. Bücker, K., et al., Electron beam dynamics in an ultrafast transmission electron microscope with Wehnelt electrode. Ultramicroscopy, 2016. **171**: p. 8-18.
263. Meurig, T.J., 4D electron microscopy: imaging in space and time. 2009: World Scientific.
264. LaGrange, T., et al., Approaches for ultrafast imaging of transient materials processes in the transmission electron microscope. Micron, 2012. **43**(11): p. 1108-1120.
265. Stevens, A., et al., Applying compressive sensing to TEM video: a substantial frame rate increase on any camera. Advanced Structural and Chemical Imaging, 2015. **1**(1): p. 10.
266. Reed, B.W., Temporal compressive sensing systems. 2017, Google Patents.
267. Reed, B.W., et al., Compressively Sensed Video Acquisition in Transmission Electron Microscopy. Microscopy and Microanalysis, 2017. **23**(S1): p. 84-85.
268. Yurtsever, A. and A.H. Zewail, Direct visualization of near-fields in nanoplasmonics and nanophotonics. Nano Lett, 2012. **12**(6): p. 3334-8.
269. Flannigan, D.J., et al., Nanomechanical motions of cantilevers: direct imaging in real space and time with 4D electron microscopy. Nano Lett, 2009. **9**(2): p. 875-81.
270. Vogel, C.R. and M.E. Oman, Iterative methods for total variation denoising. SIAM Journal on Scientific Computing, 1996. **17**(1): p. 227-238.
271. Strong, D. and T. Chan, Edge-preserving and scale-dependent properties of total variation regularization. Inverse problems, 2003. **19**(6): p. S165.
272. Elad, M., Optimized projections for compressed sensing. IEEE Transactions on Signal Processing, 2007. **55**(12): p. 5695-5702.
273. Tsiligianni, E.V., L.P. Kondi, and A.K. Katsaggelos, Construction of incoherent unit norm tight frames with application to compressed sensing. IEEE Transactions on Information Theory, 2014. **60**(4): p. 2319-2330.
274. Bernstein, D.S., Matrix mathematics: Theory, facts, and formulas with application to linear systems theory. Vol. 41. 2005: Princeton University Press Princeton.
275. Abolghasemi, V., et al. On optimization of the measurement matrix for compressive sensing. in Signal Processing Conference, 2010 18th European. 2010. IEEE.
276. Abolghasemi, V., S. Ferdowsi, and S. Sanei, A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing. Signal Processing, 2012. **92**(4): p. 999-1009.
277. Kellman, P. and E.R. McVeigh, Image reconstruction in SNR units: a general method for SNR measurement. Magnetic resonance in medicine, 2005. **54**(6): p. 1439-1447.
278. Wang, Z., et al., Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 2004. **13**(4): p. 600-612.
279. Lustig, M., D. Donoho, and J.M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic resonance in medicine, 2007. **58**(6): p. 1182-1195.
280. Lustig, M., et al., Compressed sensing MRI. IEEE signal processing magazine, 2008. **25**(2): p. 72-82.

281. Gorkhover, T., et al., Femtosecond and nanometre visualization of structural dynamics in superheated nanoparticles. Nature photonics, 2016. **10**(2): p. 93.
282. Baum, P., D.S. Yang, and A.H. Zewail, 4D visualization of transitional structures in phase transformations by electron diffraction. Science, 2007. **318**(5851): p. 788-792.
283. Gedik, N., et al., Nonequilibrium phase transitions in cuprates observed by ultrafast electron crystallography. Science, 2007. **316**(5823): p. 425-429.

SOMMAIRE RÉCAPITULATIF

Développement et applications de l'imagerie ultra-rapide à stries optiques compressées à un seul coup

L’introduction

La résolution temporelle, définissant la proximité temporelle de deux événements adjacents discernables, est un paramètre important de la photographie. Pour évaluer quantitativement la résolution temporelle, les modalités d'imagerie sont caractérisées par le concept de vitesse d'image, qui est quantifiée par la fréquence d'images avec l'unité d'images par seconde (fps). Malgré l'avancée fulgurante du nombre de pixels des caméras jusqu'à des gigapixels, la vitesse d'imagerie des caméras limite fondamentalement la capacité des humains à discerner le monde physique. Les photographes et les scientifiques ont continuellement cherché des méthodes pour capturer des scènes transitoires à une vitesse d'imagerie plus élevée, les premiers exemples étant des enregistrements bien connus en 1878 d'un cheval au galop et la photographie de 1887 d'une balle supersonique.

Les vitesses d'imagerie multi-échelles répondent aux différents besoins de recherche scientifique et d'application. Par exemple, l'imagerie à grande vitesse (par exemple, jusqu'à des milliers d'ips) joue un rôle clé dans la vélocimétrie par images de particules, la vidéosurveillance et la relecture instantanée, pour n'en nommer que quelques-uns. L'imagerie ultra-rapide (par exemple, jusqu'à des millions de fps) est nécessaire pour observer les activités neuronales, l'émission de lumière phosphorescente, les changements conformationnels des protéines, etc. L'imagerie ultra-rapide (par exemple, des milliards de fps) ouvre des recherches scientifiques telles que la surveillance du comportement moléculaire (par exemple , cinéma moléculaire), caractérisant la structure fine de la matière et étudiant des processus chimiques et biologiques jusque-là inaccessibles.

Actuellement, par rapport à la photographie sur film, les technologies de dispositif à couplage de charge (CCD) et de métal-oxyde-semi-conducteur complémentaire (CMOS) ont révolutionné la photographie à haute et ultra-haute vitesse. Les caméras CCD/CMOS à la pointe de la technologie ont atteint des fréquences d'images allant jusqu'à 10^7 ips. Malgré l'impact généralisé de ces capteurs, l'augmentation supplémentaire des fréquences d'images à l'aide de la technologie CCD ou CMOS est fondamentalement limitée par leur bande passante d'électronique et de stockage sur puce. Par exemple, la fréquence d'images est généralement

inversement proportionnelle au nombre total de pixels acquis. Les caméras CCD/CMOS ultra-rapides existantes reposent soit sur des capteurs personnalisés à des prix élevés, soit sur des réseaux de caméras volumineux.

La photographie ultrarapide compressée (CUP) est une modalité émergente d'imagerie computationnelle (CI) qui combine de manière synergique la détection compressée (CS) avec l'imagerie par stries. CUP est réputé pour son imagerie dynamique bidimensionnelle (2D) avec une vitesse d'image pouvant atteindre 10^{13} ips. Par rapport aux méthodes d'imagerie ultrarapides basées sur la pompe-sonde, CUP permet d'enregistrer des événements non répétitifs évoluant dans le temps en un seul instantané. Jusqu'à présent, CUP a mené à une variété d'applications passionnantes en physique telles que l'observation du chaos optique, la dynamique dissipative des solitons et le cône de Mach photonique. Malgré les avantages saillants de la vitesse d'imagerie, CUP ne peut pas enregistrer une dynamique de longue durée (par exemple, des processus de luminescence de conversion ascendante de l'ordre de microsecondes et de millisecondes) en un seul coup, en raison d'un temps de balayage optoélectronique inférieur à des nanosecondes. De plus, en raison de la conversion photon-photoélectron par la photocathode, l'efficacité quantique (QE) des caméras optoélectroniques à balayage est généralement <15 % pour la lumière visible. En outre, l'effet de charge d'espace dans le système de lentille électrostatique impose des contraintes dans la résolution spatiale (généralement des dizaines à des centaines de micromètres) et la plage dynamique (par exemple, <10 pour certaines caméras à balayage femtoseconde). Ces deux faiblesses limitent considérablement la qualité des données acquises.

Une caméra à balayage mécanique, utilisant une fente unidimensionnelle (1D) placée à son entrée, utilise généralement un miroir rotatif [par exemple, un scanner galvanométrique (GS) ou un miroir polygonal] pour dévier la lumière à travers une fente bidimensionnelle (2D) plan d'imagerie d'un capteur CCD/CMOS standard. Étant donné que le balayage mécanique est beaucoup plus lent que l'homologue optoélectronique dans CUP, la caméra à balayage mécanique a un temps d'exposition plus long pour enregistrer une dynamique de longue durée (jusqu'à quelques millisecondes) et a une résolution temporelle réglable généralement de centaines de nanosecondes à microsecondes pour satisfaire ultra-haute- imagerie rapide. De plus, son acquisition de données entièrement optique permet de mettre en œuvre de manière flexible de nombreuses caméras à haute sensibilité [par exemple, des caméras CCD à multiplication d'électrons (EM) et des caméras CMOS scientifiques, dont le QE peut être> 90% pour la lumière visible] pour obtenir des SNR supérieurs dans les mesures. Le fonctionnement

tout optique évite également l'effet de charge d'espace, ce qui permet une résolution spatiale limitée par l'optique et une plage dynamique élevée (par exemple, > 60 000 d'une caméra EMCCD). Ainsi, la caméra à balayage mécanique devient un bon candidat pour surmonter les limitations de taille et de prix des caméras CCD/CMOS ultra-rapides existantes et pour répondre aux limitations de QE, de l'effet de charge d'espace et du temps de balayage en CUP. Cependant, la strie mécanique est une technique d'imagerie 1D en raison de la fente d'entrée.

Objectif de la these

Cette thèse se concentre sur les efforts de développement de matériel d'imagerie ultra-rapide (c'est-à-dire une résolution temporelle inférieure à la microseconde), d'un logiciel de reconstruction d'image haute fidélité et d'applications pertinentes. Tout d'abord, nous prévoyons de proposer la photographie à ultra-haute vitesse à stries optiques compressées (COSUP) en une seule prise. En associant la CS à l'imagerie par traînées optiques, COSUP tente de doter une caméra à traînée mécanique d'une capacité d'imagerie 2D. En particulier, un masque d'encodage 2D est utilisé pour remplacer la fente 1D de la caméra mécanique à balayage. Un GS est placé dans le plan de Fourier d'un système d'imagerie 4f pour effectuer un balayage mécanique. Une caméra CCD/CMOS prête à l'emploi est déployée pour recevoir de manière compressive une scène transitoire. Par rapport aux techniques d'imagerie ultra-rapides existantes, COSUP vise à présenter des avantages en termes de profondeur de séquence (c'est-à-dire le nombre d'images dans une acquisition), de nombre de pixels, de taille et de coût.

Deuxièmement, les instruments optiques existants n'ont pas la capacité d'imagerie de durée de vie de luminescence à conversion ascendante à champ large en temps réel. En utilisant des nanoparticules de conversion ascendante comme indicateurs, COSUP est appliqué à l'imagerie de durée de vie de luminescence et au thermomètre optique avec détection de température de conversion ascendante à débit vidéo en champ large. Développé à partir du système COSUP, le thermomètre enregistre d'abord la décroissance de l'intensité de la luminescence à champ large de manière compressive dans deux vues en une seule prise. Ensuite, un algorithme basé sur CS est utilisé pour reconstruire la vidéo, à partir de laquelle la distribution de durée de vie extraite est convertie en une carte de température. Pour démontrer la large utilité, le thermomètre sera appliqué à la surveillance longitudinale de la température à champ large sous un milieu de diffusion mince.

Troisièmement, dans les systèmes CTI existants, la reconstruction d'image basée sur CS est limitée par le long temps de traitement et la qualité d'image instable dans la méthode existante basée sur la modélisation analytique. Pour surmonter ces problèmes, nous visons à développer un réseau de neurones profonds (DNN) d'instantané à vidéo qui mappe une image 2D enregistrée de manière compressive à un film sans processus d'itération. Par rapport aux méthodes existantes basées sur l'apprentissage, le nouveau DNN vise une grande profondeur de séquence en préservant la cohérence spatio-temporelle des vidéos reconstruites. Combiné à l'imagerie ultra-rapide compressée, le nouveau DNN contribue au développement du COSUP en temps réel assisté par apprentissage automatique (SMART) en un seul coup. SMART-COSUP vise à des applications de rétroaction ponctuelles comme le suivi de particules multiples.

Quatrièmement, apporter une vitesse d'imagerie ultra-rapide à la microscopie électronique à transmission (TEM) a toujours été un défi. Malgré des progrès récents significatifs dans cette direction, il reste difficile d'atteindre une résolution temporelle inférieure à la nanoseconde et une résolution spatiale au niveau du nanomètre avec une seule impulsion électronique, en imagerie en temps réel. Pour remédier à cette limitation, nous prévoyons d'apporter le concept de COSUP en TEM et d'établir une modélisation analytique, c'est-à-dire un TEM ultrarapide (CU) compressé. Par rapport aux méthodes pompe-sonde, CUTEM est compétent pour l'observation de dynamiques transitoires non répétables. De plus, pour obtenir une meilleure résolution spatiale reconstruite, le CUTEM à double cisaillement (DS), en tant que variante de CUTEM, sera également analysé.

Résultats et discussions

COSUP

Le schéma du système COSUP est illustré à la Fig. S1. Une scène transitoire est d'abord imagée sur un dispositif à micromiroir numérique (DMD, AJD-4500, Ajile Light Industries), sur lequel un motif pseudo-aléatoire binaire (avec une taille de pixel d'encodage de 32,4 × 32,4 μm^2) est chargé pour effectuer l'encodage spatial. Par la suite, les trames codées spatialement sont relayées sur une caméra CMOS (GS3-U3-23S6M-C, FLIR) par un système 4f. Un scanner galvanométrique (GS, 6220H, Cambridge Technology), placé dans le plan de Fourier de ce système 4f, cisaille temporellement les images spatialement codées linéairement vers différents emplacements spatiaux le long de l'axe x de la caméra CMOS en fonction de leur temps d'arrivée. La

synchronisation entre la rotation du GS et l'exposition de la caméra est contrôlée par le signal sinusoïdal et le signal rectangulaire d'un générateur de fonctions (DG1022, Rigol Technologies), comme illustré dans l'encadré de la Fig. S1. Enfin, via l'intégration spatio-temporelle, la caméra CMOS enregistre de manière compressive la scène encodée spatialement et cisaillée temporellement sous forme d'une image de strie 2D avec une seule exposition. Il est à noter que notre travail s'inspire des avancées récentes de la photographie ultrarapide compressée. Cependant, au lieu d'utiliser une caméra à balayage, nous implémentons le GS pour le cisaillement temporel et utilisons une caméra CMOS standard pour la détection. Cette conception évite ainsi les inconvénients - tels que l'effet de charge d'espace et le faible rendement quantique de la photocathode - qui sont présentés dans la caméra à balayage. Dans ce qui suit, nous démontrerons que cette approche tout optique peut augmenter la vitesse d'imagerie de la caméra CMOS de quatre ordres de grandeur au niveau Mfps pour enregistrer des impulsions laser uniques éclairant à travers des cibles transmissives et pour suivre un objet en mouvement rapide.

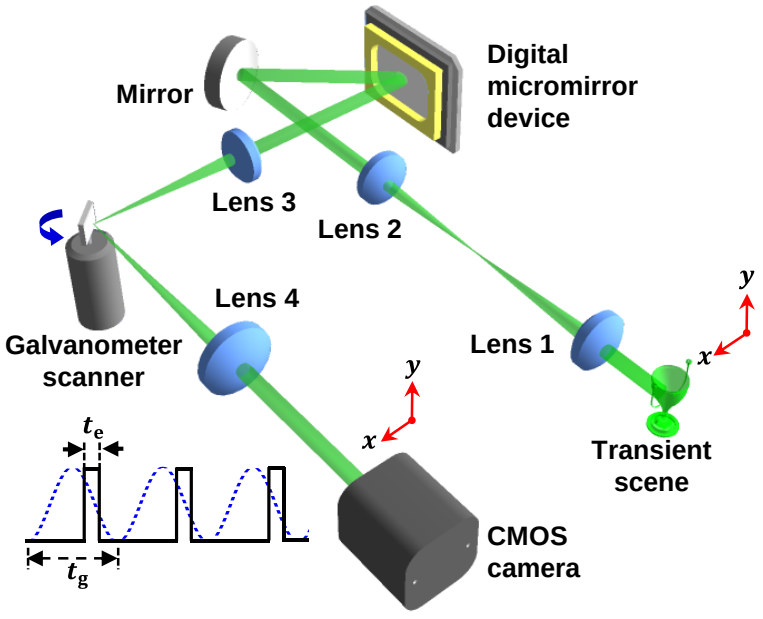


Figure S1 Schéma du système COSUP. En médaillon : synchronisation entre l'exposition de la caméra CMOS (ligne noire pleine) avec un temps d'exposition de $\boldsymbol{t_e}$ et le signal de commande sinusoïdal du scanner galvonométrique (ligne pointillée bleue) avec une période de $\boldsymbol{t_g}$. Objectifs 1 et 4.

Le fonctionnement du système COSUP peut être décrit par le modèle suivant

$$E = \mathbf{T}\mathbf{S}_\mathbf{o}\mathbf{C}\, I(x, y, t), \tag{S1}$$

où $I(x, y, t)$ est l'intensité lumineuse de l'événement transitoire, **C** représente l'encodage spatial par le DMD, $\mathbf{S_o}$ représente le cisaillement linéairement temporel par le GS (l'indice “o” signifie “optique”), et **T** représente Intégration spatio-temporelle par la caméra CMOS. Du fait de la connaissance a priori des opérateurs et de la parcimonie spatio-temporelle de la scène transitoire, $I(x, y, t)$ peut être récupéré à partir de la mesure *E* en résolvant le problème inverse de

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \| E - \mathbf{T}\mathbf{S_o}\mathbf{C}I \|_2^2 + \lambda \Phi_{\mathrm{TV}}(I) \right\} \tag{S2}$$

Ici $\|\cdot\|_2^2$ représente la norme l_2, λ est un coefficient de pondération et $\Phi_{\mathrm{TV}}(\cdot)$ est le régularisateur de la variation totale (TV). En pratique, $I(x, y, t)$ a été récupéré en utilisant un algorithme basé sur CS qui a été développé sur l'algorithme itératif de rétrécissement/seuil en deux étapes.

SPLIT

Le schéma du système SPLIT est illustré à la figure S2. Un laser à ondes continues de 980 nm (BWT, DS3-11312-113-LD) est utilisé comme source lumineuse. Le faisceau laser traverse un système 4f composé de deux lentilles de focale de 50 mm (L1 et L2, Thorlabs, LA1255). Un hacheur optique (Scitec Instruments, 300CD) est placé au niveau du plan focal arrière de la lentille L1 pour générer des impulsions optiques de 50 µs. Ensuite, l'impulsion passe à travers une lentille de focale de 100 mm (L3, Thorlabs, AC254-100-B) et est réfléchie par un miroir dichroïque passe-court (Edmund Optics, 69-219) pour générer une mise au point sur la focale arrière. plan d'un objectif (Nikon, CF Achro 4×, ouverture numérique 0,1, nombre de champ 11 mm). Ce schéma d'éclairage produit un éclairage à grand champ (1,5 × 1,5 mm^2 FOV) aux UCNP au niveau du plan objet.

Les UCNP excités dans le proche infrarouge émettent de la lumière dans le domaine spectral visible. La décroissance de l'intensité lumineuse sur le FOV 2D est une scène dynamique, notée $I(x, y, t)$. La lumière émise est collectée par le même objectif, transmise à travers le miroir dichroïque, et est filtrée par un filtre passe-bande (Thorlabs, MF542-20 ou Semrock, FF01-660/30-25). Ensuite, un diviseur de faisceau (Thorlabs, BS013) divise également la lumière en deux composants. La composante réfléchie est imagée par une caméra CMOS (FLIR, GS3-U3-23S6M-C) avec un objectif de caméra (Fujinon, HF75SA1) via une intégration spatio-temporelle (notée opérateur T) en tant que vue 1, dont la distribution d'énergie optique est notée par $E_1(x_1, y_1)$.

La composante transmise forme une image de la scène dynamique sur un masque de codage transmissif avec un motif binaire pseudo-aléatoire (Fineline Imaging, rapport de transmission de 50 % ; taille de pixel de codage de 60 µm). Ce processus de codage spatial est noté par l'opérateur $\mathbf{C}$. Ensuite, la scène codée spatialement est imagée par une caméra à balayage mécanique. En particulier, la scène est relayée au plan du capteur d'une caméra CCD à multiplication électronique (EM) (Nüvü Camēras, HNü 1024) par un système d'imagerie 4f composé de deux lentilles de focale de 100 mm (L4 et L5, Thorlabs, AC254 -100-A). Un scanner galvanométrique (Cambridge Technology, 6220H), placé dans le plan de Fourier du système d'imagerie 4f, cisaille temporellement les images spatialement codées linéairement vers différents emplacements spatiaux le long de l'axe x_2 de la caméra EMCCD en fonction de leur temps d'arrivée. Ce processus de cisaillement temporel est noté par l'opérateur S. Enfin, la scène dynamique codée spatialement et cisaillée temporellement est enregistrée par l'EMCCD via une intégration spatio-temporelle pour former la Vue 2, dont la distribution d'énergie optique est notée $E_2(x_2, y_2)$.

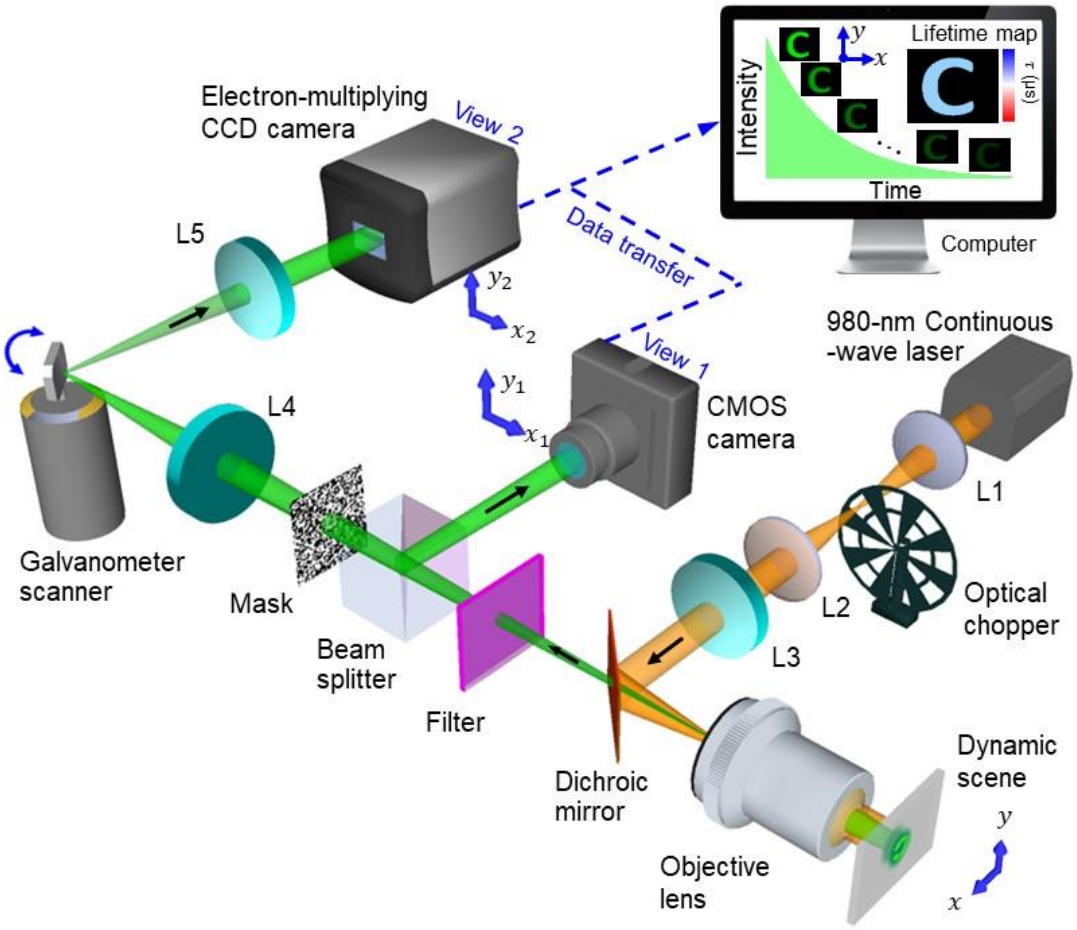


Figure S2 Schéma du système SPLIT. L'illustration montre l'acquisition de données et la reconstruction d'images de la décroissance de l'intensité de la luminescence dans une lettre “C”. L1–L5, objectif.

En combinant la formation d'image de $E_1(x_1, y_1)$ et $E_2(x_2, y_2)$, l'acquisition des données de SPLIT est exprimée par

$$E = \mathbf{TM}\, I, \tag{S3}$$

où E désigne la concaténation des mesures $[E_1, \alpha E_2]^T$ (l'exposant T désigne la transposition), $\mathbf{M}$ désigne l'opérateur linéaire $[\mathbf{1}, \alpha\mathbf{SC}]^T$, et α est un facteur scalaire introduit pour équilibrer l'énergie rapport entre les deux vues pendant la mesure. Le matériel du système SPLIT est synchronisé pour capturer les deux vues (détaillées dans Méthodes) qui sont calibrées avant l'acquisition des données.

Après acquisition des données, E est traité par un algorithme qui récupère le cube de données de la scène dynamique en s'appuyant sur la parcimonie spatio-temporelle de la scène dynamique et la connaissance a priori de chaque opérateur. Développé à partir du cadre de la méthode des multiplicateurs à direction alternée plug-and-play (PnP-ADMM), l'algorithme de reconstruction de SPLIT résout le problème de minimisation de

$$\hat{I} = \underset{I}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{TM}I - E\|_2^2 + R(I) + \mathbf{I}_+(I) \right\}. \quad \text{(S4)}$$

Ici, $\|\cdot\|_2$ représente la norme l_2. Le terme de fidélité, $\frac{1}{2}\|\mathbf{TM}I - E\|_2^2$, représente la similitude entre la mesure et le résultat estimé. $R(\cdot)$ est le régularisateur implicite qui favorise la parcimonie dans la scène dynamique. $\mathbf{I}_+(\cdot)$ représente une contrainte d'intensité non négative. Par rapport aux schémas de reconstruction existants, PnP-ADMM implémente une stratégie de division variable avec un débruiteur de pointe pour obtenir des solutions rapides et fermées à chaque problème de sous-optimisation, ce qui produit une qualité d'image élevée en reconstruction. Le cube de données récupéré de la scène dynamique a une profondeur de séquence (c'est-à-dire le nombre d'images dans un film reconstruit) de 12 à 100 images, chacune contenant 460 × 460 (x, y) pixels. La vitesse d'imagerie est réglable de 4 à 33 000 images par seconde (kfps).

Le cube de données reconstruit est ensuite converti en une carte de durée de vie de photoluminescence. En particulier, pour chaque point (x, y), l'aire sous la courbe de décroissance d'intensité normalisée est intégrée pour rapporter la valeur de la durée de vie de la photoluminescence. Enfin, en utilisant la relation approximativement linéaire entre la durée de vie des UCNP et la plage de température physiologiquement pertinente (20 à 46 ° C dans ce travail), la distribution de température 2D, $T(x, y)$, est calculée par

$$T(x, y) = c_{\mathrm{t}} + \frac{1}{S_{\mathrm{a}}} \int \frac{\hat{I}(x, y, t)}{\hat{I}(x, y, 0)} dt, \quad \text{(S5)}$$

où c_{t} est une constante et S_{a} est la sensibilité absolue à la température. Tirant parti de la fréquence d'images intrinsèque de la caméra EMCCD, le système SPLIT peut générer des cartes de température déterminées sur la durée de vie à une fréquence vidéo de 20 Hz.

SMART-COSUP

Le schéma du système SMART-COSUP est illustré à la figure S3(a). Son principe de fonctionnement contient l'acquisition de données monocoup et la reconstruction vidéo en temps réel [Fig. S3(b)]. Une scène dynamique, $I(x,y,t)$, est imagée par une optique frontale sur un masque transmissif binaire pseudo-aléatoire (Fineline Imaging) avec des pixels de codage de 25 µm × 25 µm. Cette opération de modulation spatiale est notée par l'opérateur $\mathbf{C}$. La distribution d'intensité juste après le masque de codage est exprimée par

$$I_{\mathrm{c}}(x,y,t)=\sum_{j,k} I\left(\frac{x}{M_{\mathrm{f}}},\frac{y}{M_{\mathrm{f}}},t\right) C_{jk}\mathrm{rect}\left(\frac{x}{d_{\mathrm{e}}}-j,\frac{y}{d_{\mathrm{e}}}-k\right). \tag{S6}$$

Ici, M_{f} est le grossissement de l'optique avant. C_{jk} désigne un élément d'une matrice binaire représentant le motif de codage, j et k sont des indices d'éléments matriciels. d_{e} est la taille des pixels d'encodage sur le masque. $\mathrm{rect}(\cdot)$ représente la fonction rectangulaire.

Par la suite, la scène modulée spatialement est relayée par un système d'imagerie 4f, composé d'un scanner galvanométrique (GS, Cambridge Technology, 6220H) et de deux lentilles identiques (Lens 1 et Lens 2, Thorlabs, AC254-075-A). Le GS est placé au plan de Fourier de ce système d'imagerie 4f pour effectuer un cisaillement optique dans la direction x, noté par l'opérateur $\mathbf{S_o}$. L'image cisaillée peut être exprimée comme

$$I_{\mathrm{s}}(x,y,t)=I_{\mathrm{c}}(x-v_{\mathrm{s}}t,y,t), \tag{S7}$$

où v_{s}, désignant la vitesse de cisaillement de SMART-COSUP, est calculé par $v_{\mathrm{s}}=\alpha V_{\mathrm{g}} f_2/t_{\mathrm{g}}$. Ici, V_{g} =0.16–0.64 V est la tension ajoutée sur le GS. α est une constante qui relie V_{g} à l'angle de déviation de GS en tenant compte de la forme d'onde d'entrée. $f_2=75$ mm est la distance focale de l'objectif 2 sur la figure S7(a). $t_{\mathrm{g}}=50$ ms est la période du signal sinusoïdal ajouté au scanner galvanométrique.

Enfin, la scène dynamique est intégrée spatio-temporellement par une caméra CMOS (FLIR, GS3-U3-23S6M-C) à un instantané 2D, noté par l'opérateur $\mathbf{T}$. L'énergie optique, $E(m,n)$, mesurée au pixel (m,n), est

$$E(m,n)=\iiint I_s(x,y,t)\mathrm{rect}\left(\frac{x}{d_{\mathrm{c}}}-m,\frac{y}{d_{\mathrm{c}}}-n\right)dxdydt. \tag{S8}$$

Ici, m et n sont les indices de pixels dans les axes x et y de la caméra. $d_{\mathrm{c}}=5.86$ µm est la taille de pixel du capteur CMOS. À partir des équations. S8-10, le modèle avancé de SMART-COSUP s'exprime par

$$E(m,n) = \mathbf{TS_0C}I(x,y,t). \quad (S9)$$

Dans la reconstruction vidéo en temps réel qui s'ensuit, les données capturées sont transférées vers un poste de travail équipé d'une unité de traitement graphique (NVIDIA, RTX Titan). Le S2V-AE récupère le datacube de la scène dynamique en 60 ms.

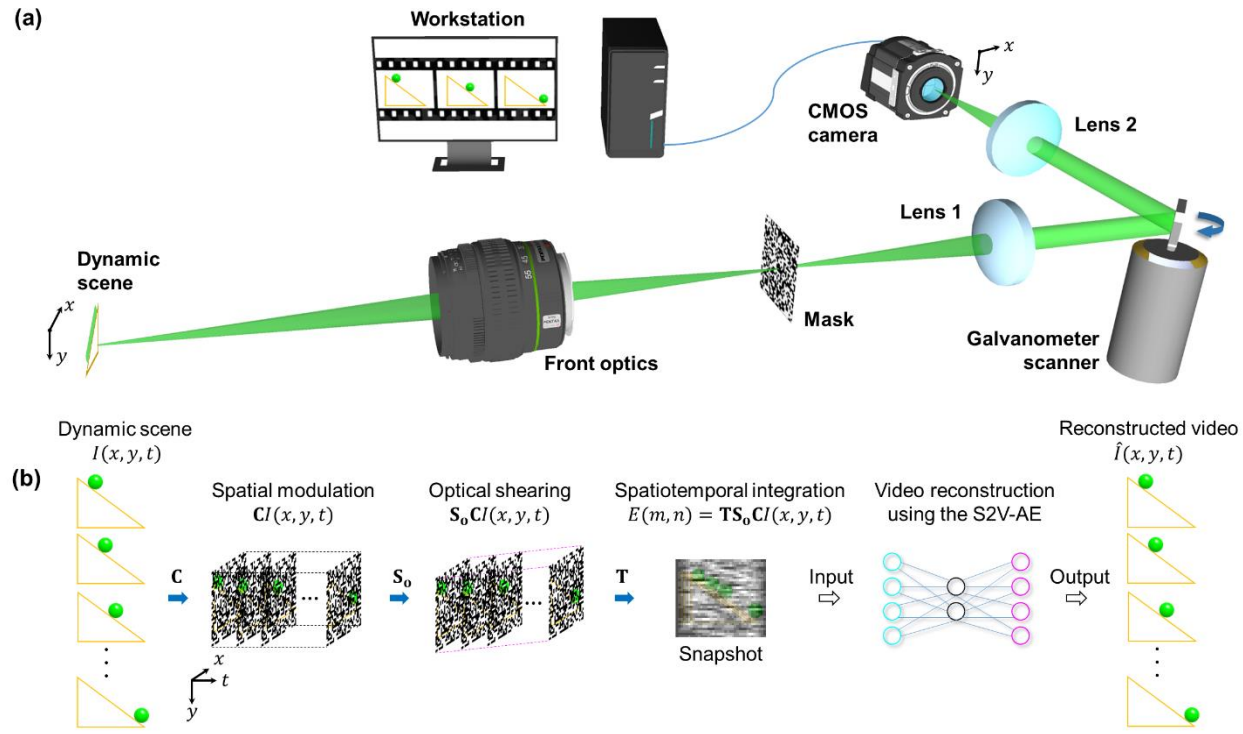


Figure S3 Photographie à ultra-haute vitesse (COSUP) compressée en temps réel assistée par apprentissage automatique (SMART) en une seule prise. (a) Schéma du système. (b) Principe de fonctionnement. S2V-AE, autoencodeur instantané-vidéo.

CUTEM

Le CUTEM proposé [Fig. La technique S4(a)] peut être réalisée expérimentalement en intégrant un masque et des électrodes de cisaillement à une machine TEM assistée par laser. Dans CUTEM, une seule impulsion électronique d'une nanoseconde sonde un événement transitoire initié dans un échantillon pour générer une fonction de densité électronique modulée spatio-temporellement, $I(x,y,t)$. L'impulsion électronique transmise subit alors plusieurs manipulations dont chacune est prise en compte par un opérateur (détaillé dans le paragraphe suivant). Plus précisément, l'impulsion subit une distorsion électronique spatio-temporelle au cours de la propagation qui s'ensuit dans le système TEM. Un masque de transmission binaire, avec un motif

spécialement conçu, est placé sur le trajet des électrons d'imagerie pour coder $I(x, y, t)$. Réalisé par les outils de nanofabrication existants, ce masque bloque les électrons dans certaines régions de l'espace et leur permet de les traverser dans d'autres. Après le masque, une tension variable dans le temps, appliquée à une paire d'électrodes, cisaille temporellement $I(x, y, t)$. Enfin, une seule image strie, notée E, est formée sur une caméra CCD via une intégration spatio-temporelle (intégration spatiale sur chaque pixel CCD et intégration temporelle sur le temps d'exposition). Le masque et les électrodes de cisaillement peuvent être insérés dans la machine TEM en utilisant l'un des ports après le système de lentilles de projection.

Mathématiquement, le processus d'acquisition de données décrit ci-dessus peut être exprimé comme un modèle avancé [Fig. S4(a)] par

$$E = \mathbf{TSCD}\, I(x, y, t), \tag{S10}$$

où l'opérateur linéaire $\mathbf{D}$ rend compte de la distorsion spatio-temporelle de l'électron induite par tous les facteurs expérimentaux. $\mathbf{C}$ représente le codage spatial par le masque de transmission binaire, $\mathbf{S}$ représente le cisaillement temporel par la tension variable dans le temps appliquée aux électrodes, et $\mathbf{T}$ représente l'intégration spatio-temporelle par la caméra CCD. Pour simplifier la dénotation, nous utilisons un opérateur linéaire $\mathbf{O}$ pour représenter l'ensemble du processus d'acquisition de données (c'est-à-dire $\mathbf{O} = \mathbf{TSCD}$). Après l'acquisition des données, nous récupérons informatiquement la scène transitoire. Compte tenu de l'opérateur connu O et de la parcimonie spatio-temporelle de la scène transitoire, $I(x, y, t)$ peut être récupéré en résolvant le problème inverse de l'Eq. (S10). Ce processus peut être formulé comme

$$\hat{I} = \underset{I}{\operatorname{argmin}} \|E - \mathbf{O}I\|_2^2 + \lambda \Phi(I)\,, \tag{S11}$$

où λ est un paramètre de pondération et $\Phi(\cdot)$ est la fonction de régularisation. Dans l'éq. (S14), la minimisation du premier terme, $\|E - \mathbf{O}I\|_2^2$, se produit lorsque la mesure réelle E correspond étroitement à la solution estimée $\mathbf{O}I$, tandis que la minimisation du deuxième terme, $\Phi(I)$, encourage I à être constante par morceaux (c'est-à-dire éparse dans le domaine du gradient spatial). La pondération de ces deux termes est ajustée par λ pour conduire aux résultats les plus cohérents avec la vérité terrain. Étant donné que le nombre de lignes et de colonnes de la caméra CCD est N_r et N_c, le nombre de voxels dans le cube de données reconstruit (c'est-à-dire N_x, N_y, et N_t) doit répondre à l'exigence de

$$\begin{aligned} &N_x \le N_\mathrm{c}\,, \ \ and \\ &N_y + N_t - 1 \le N_\mathrm{r}\,. \end{aligned} \tag{S12}$$

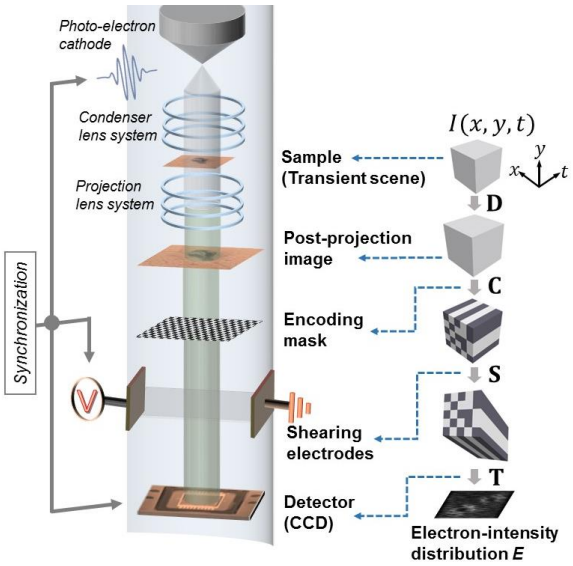


Figure S4 Méthodes proposées pour l'imagerie électronique ultrarapide en temps réel monocoup. Schéma matériel de CUTEM et de son modèle avancé. $\mathbf{C}$, opérateur de codage spatial ; $\mathbf{D}$, opérateur de distorsion spatiotemporelle électronique ; $\mathbf{S}$, opérateur de cisaillement temporel ; $\mathbf{T}$, opérateur d'intégration spatio-temporelle.

Conclusions

Tout d'abord, COSUP concerne l'imagerie optique passive ultra-rapide 2D monocoup. Contrairement au balayage optoélectronique dans CUP, COSUP présente la strie optique à l'aide d'un GS dans le système d'imagerie 4*f*. En conséquence, COSUP atteint des vitesses d'imagerie réglables allant jusqu'à 1.5 Mfps, soit environ trois ordres de grandeur de plus que la vitesse d'imagerie de pointe de CTI avec une caméra CMOS standard. En outre, le système COSUP est capable d'atteindre un nombre de pixels de 0.5 mégapixels dans chaque image, ce qui donne un produit espace-bande passante-temps allant jusqu'à 750 gigapixels par seconde. Le système COSUP peut reconstruire une profondeur de séquence allant jusqu'à 500 images en une seule mesure, ce qui augmente considérablement la capacité de stockage sur puce. La transmission d'impulsions laser uniques à travers un masque et la forme et la position du suivi d'un objet se déplaçant rapidement en temps réel ont été utilisées pour démontrer la capacité d'imagerie ultra-rapide de COSUP.

Deuxièmement, le système SPLIT a été développé pour la détection dynamique de la température à large champ en temps réel. SPLIT, adapté du système COSUP, enregistre en

compression l'émission de photoluminescence sur un FOV 2D en deux vues. Ensuite, l'algorithme PnP-ADMM à double vue reconstruit des traces de décroissance d'intensité résolues dans l'espace, à partir desquelles une distribution de durée de vie de photoluminescence et la carte de température correspondante sont extraites. Utilisé avec les UCNP core/shell NaGdF4:Er^{3+}, Yb^{3+}/$NaGdF_4$, SPLIT a permis une cartographie de la température avec une sensibilité élevée pour les bandes d'émission de conversion ascendante verte et rouge avec une résolution spatiale de 20 µm dans un champ de vision de 1.5 × 1.5 mm^2 à une fréquence vidéo de 20 Hz. SPLIT est démontré dans la surveillance longitudinale de la température d'un fantôme sous un tissu de poulet frais. SPLIT est également appliqué à la cartographie dynamique de la température d'une seule cellule d'un échantillon d'épiderme d'oignon monocouche en mouvement.

Troisièmement, nous avons développé le S2V-AE pour une reconstruction vidéo rapide et de haute qualité à partir d'un seul instantané acquis par compression. Ce nouveau DNN a facilité le développement du système SMART-COSUP, qui a démontré une imagerie à ultra-haute vitesse en un seul coup d'événements transitoires dans l'imagerie macroscopique et microscopique jusqu'à 20 kfps avec une taille de vidéo reconstruite en temps réel de $(N_x, N_y, N_t) = (256, 256, 100)$. Ce système a été appliqué au suivi de particules multiples.

Quatrièmement, nous avons présenté les études de conception et de simulation de CUTEM et DS-CUTEM, deux méthodes d'imagerie électronique ultrarapide en temps réel assistées par CS. Des scènes transitoires capturées expérimentalement par TEM assisté par laser ont été utilisées pour tester leur fidélité. Différents types de modèles de masques et de tailles de pixels de codage ont été examinés pour optimiser la qualité de l'image reconstruite pour le SNR donné. Les résultats de la reconstruction avec à la fois la distorsion électronique et les processus de correction correspondants ont également été fournis. Sur la base de ces investigations, une reconstruction de haute qualité des deux scènes transitoires a été démontrée en simulation.