Weighted multi-scale denoising via adaptive multi-channel fusion for compressed ultrafast photography

CHENGZHI JIN,1,5 DALONG QI,1,5 JIALI YAO,1 YILIN HE,1 PENGPENG DING,1 ZIHAN GUO,1 ZHENGQI HUANG,1 YU HE,1 YUNHUA YAO,1 ZHIYONG WANG,2 ZHENRONG SUN,1 AND SHIAN ZHANG1,3,4,∗

1State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai, 200062, China
2School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China
3Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
4Collaborative Innovation Center of Light Manipulations and Applications, Shandong Normal University, Jinan 250358, China
5These authors contributed equally.
∗sazhang@phy.ecnu.edu.cn

Abstract: Being capable of passively capturing transient scenes occurring in picoseconds and even shorter time with an extremely large sequence depth in a snapshot, compressed ultrafast photography (CUP) has aroused tremendous attention in ultrafast optical imaging. However, the high compression ratio induced by large sequence depth brings the problem of low image quality in image reconstruction, preventing CUP from observing transient scenes with fine spatial information. To overcome these restrictions, we propose an efficient image reconstruction algorithm with multi-scale (MS) weighted denoising based on the plug-and-play (PnP) based alternating direction method of multipliers (ADMM) framework for multi-channel coupled CUP (MC-CUP), named the MCMS-PnP algorithm. By removing non-Gaussian distributed noise using weighted MS denoising during each iteration of the ADMM, and adaptively adjusting the weights via sufficiently exploiting the coupling information among different acquisition channels collected by MC-CUP, a synergistic combination of hardware and algorithm can be realized to significantly improve the quality of image reconstruction. Both simulation and experimental results demonstrate that the proposed adaptive MCMS-PnP algorithm can effectively improve the accuracy and quality of reconstructed images in MC-CUP, and extend the detectable range of CUP to transient scenes with fine structures.

1. Introduction

Aiming at capturing transient scenes in space and time with an ultra-high temporal resolution or frame rate, ultrafast optical imaging (UOI) techniques [1,2] have been widely applied in the fields of photochemistry [3], plasma physics [4] and biophotonics [1], and played irreplaceable roles in the analysis and understanding of their phenomena and mechanisms. Based on the imaging strategy, existing UOI techniques can be generally divided into two categories: multi-shot and single-shot. Multi-shot UOI techniques mainly include various pump-probe imaging methods via temporal scanning by ultrashort probe pulses or spatial scanning using zero-dimensional and one-dimensional ultrafast detectors. However, due to the requirement of repeated measurements, these techniques are not suitable for observing irreducible or destructive events. Complementarily, single-shot UOI techniques [5] can obtain spatiotemporal information of the scene in a snapshot.
without multiple measurements, which can be further divided into active illumination and passive acquisition types. Active illumination types, such as chirped spectral mapping ultrafast photography (CSMUP) [6], sequentially timed all-optical mapping photography (STEAM) [7] and frequency recognition algorithm for multiple exposures (FRAME) [8], require modulations on the illumination, and thus cannot measure self-emission ultrafast events (i.e., fluorescence lifetime). In contrast, passive acquisition types directly receive the photons emitted by the object without the need of illumination. Representative methods include ultrafast all-optical solid-state framing camera (UASFC) [9], time and spatial-frequency multiplexing (TSFM) [10], and compressed ultrafast photography (CUP) [11]. Among them, CUP stands out due to the characteristics of both ultrahigh frame rate and high sequence depth (number of frames per exposure) in a receive-only manner.

CUP is an ultrafast computational imaging technique based on compressed sensing (CS) [12]. It records 3D spatiotemporal information in a 2D compressed image by combining aperture-encoded imaging [13] and deflection imaging techniques, and subsequently recover the scene with reconstruction algorithms. The frame rate of 10 trillion frames per second and sequence depth of 300 frames have been demonstrated in previous studies [11]. Therefore, CUP has great advantages for measuring self-luminous or non-repeatable ultrafast phenomena [14]. However, the large sequence depth also brings the problem of high compression ratio and further results in low image quality in image reconstruction, preventing CUP from observing transient scenes with complex spatial information. Many efforts have been made to improve CUP’s imaging quality, mainly from the two aspects of hardware and algorithm. In terms of hardware, adding channels to increase the sampling rate is an effective strategy, and the most representative methods are complementary dual-channel CUP proposed by Liang et al. [15], and multi-channel coupled CUP (MC-CUP) originated by Yao et al. [16]. Both of them obtain multiple measurements of the same scene by using different encoding masks during a single acquisition. In terms of algorithms, a lot of model-based or learning-based methods are proposed to improve the reconstruction performance to a higher level, such as TV-BM3D [17], DeSCI [18], ALDL [19], U-net [20], alternating direction method of multipliers (ADMM)-based FFDNet [21]. However, all the existing methods in hardware or algorithm aspect are independent, there has not emerged a method making the joint use of hardware and algorithms yet.

Due to its flexibility and efficiency, the plug-and-play (PnP) framework [22,23] has been widely used in the field of computational imaging, including deblurring [24], super-resolution [25,26], hyperspectral imaging [27,28], CUP, and so on. The main idea of the PnP framework is to combine proximal algorithms with advanced image denoisers, which applies the sparse factorizable prior implicitly embedded in a Gaussian denoiser. However, the work by Zhang et al. [29] showed that the noise to be handled in PnP does not follow a fixed Gaussian distribution, and the general-purpose Gaussian denoiser prior and the manual selection of hyper-parameters comes at the cost of loss of efficiency and specialization. The choice of denoiser with optimal denoising parameters is the key to the elevation of image reconstruction performance of PnP-based algorithms, since unsuitable denoising parameters will lead to an increase in the iteration number or even non-converging results. Previous studies [23,30] have shown that different denoisers and different hand-crafted parameter settings have a large impact on the results of the final recovery. Fortunately, Athavale et al. [31] give us a new approach to remove the non-Gaussian distribution noise by Gaussian denoiser with weighted multi-scale (MS) denoising.

Here, we introduce weighted MS denoising into the PnP-based ADMM framework via adaptive MC fusion for CUP reconstruction, named MCMS-PnP algorithm, in which adaptive MS denoising with optimal parameters are obtained. In order to optimize the weights of different scale denoising during each ADMM iteration, MC-CUP technique is fully exploited. By extracting the coupling data from different channels with a least-squares optimization method, adaptive parameters optimization without manual adjustment and complete fusion of hardware
and algorithms is achieved. Considering the high flexibility of the PnP framework, a deep learning-based DRUNet [29] is employed as a denoiser to form an ADMM-DRUNet algorithm for demonstration in this work. The performance of the proposed algorithm is validated by simulations and real experiments from MC-CUP. Compared with the traditional MC processing methods [15,16], the simulated and experimental results demonstrate that the proposed adaptive MS denoising ADMM-DRUNet algorithm outperforms all existed CUP reconstruction algorithms in terms of image quality, and has strong noise robustness as well. Therefore, the proposed MCMS-PnP method is able to achieve the state-of-the-art (SOTA) reconstruction results in CUP by fully extracting the coupling information in different channels.

2. Principles

As a typical computational optical imaging strategy, the overall procedures of MC-CUP are composed of two parts, i.e., the data acquisition and image reconstruction, which integrally collects transient scenes in a coded aperture manner and recovers them with the aid of algorithms for inverse problems, respectively. Figure 1(A) shows the schematic diagram of the data acquisition in MC-CUP. As can be seen, a 3D transient scene, \( I(x, y, t) \), is divided into \( M \) replicas, and each replica is first spatially encoded by an independent random binary pattern generated by a spatial light modulator, e.g., digital micromirror device (DMD) [11] or printed transmissive mask [32]. Successively, the coded replicas are temporally sheared along one of the spatial axes with a temporal deflector, e.g., streak camera [33,34] or electro-optical deflector [35], and integrally measured by an image detector, e.g., complementary metal oxide semiconductor (CMOS) or charged coupled device (CCD). Finally, 2D compressed images \( E_i(x', y')(i = 1, 2, \cdots, M) \) are obtained. Mathematically, the data acquisition through different channels can be expressed as follows

\[
\begin{align*}
E_1(x', y') &= TSC_1 I(x, y, t) + n_1 \\
E_2(x', y') &= TSC_2 I(x, y, t) + n_2 \\
&\vdots \\
E_M(x', y') &= TSC_M I(x, y, t) + n_M
\end{align*}
\]  

(1)

where \( C_i(i = 1, 2, \cdots, M) \) represents the spatial encoding for each acquisition channel with \( M \) denoting the number of channels, \( T \) represents the temporal shearing, \( S \) represents the spatiotemporal integration, \( n_i(i = 1, 2, \cdots, M) \) denotes the measurement and detector noise in the image acquisition for each channel, and \( E_i(x', y') \) is the measured 2D image in the \( i \)th channel. For simplicity, it is set that \( \Phi_i = TSC_i \), and \( I(x, y, t) \) and \( E_i(x', y') \) are abbreviated to \( x \) and \( y_i \), respectively, thus the image acquisition for each channel can be concatenated as

\[
y_i = \Phi_i x + n_i, \ \forall i = 1, 2, \cdots, M,
\]

(2)

where \( y = \text{Vec}(Y) \in \mathbb{R}^{N_xN_y} \), \( n = \text{Vec}(N) \in \mathbb{R}^{N_xN_y} \), \( x = \text{Vec}(X) \in \mathbb{R}^{N_xN_yN_t} \), and \( \Phi_i \in \mathbb{R}^{N_xN_y \times N_xN_yN_t} \). Here, \( X \), \( Y \), and \( N \) represent higher order matrix representation of the corresponding data, and \( N_x \), \( N_y \), and \( N_t \) denote the numbers of discretized pixels in the \( x \), \( y \), and \( t \) coordinates, respectively. The sampling rate of each channel is \( 1/N_t \), so the overall sampling rate of MC-CUP is \( M/N_t \). In general, MC-CUP with multiple independent acquisition channels can effectively increase the sampling rate of the transient scene. The influence of number of channels on the quality of the reconstructed images is investigated by simulations, please refer to Supplemental 1 for the detailed result.

Based on the model derived in Eq. (2), we develop an adaptively weighted MS denoising method for MC-CUP reconstruction using the PnP-ADMM framework. In this process, the original 3D transient scene needs to be reconstructed from the captured 2D compressed image for each
channel, which employs an iterative algorithm via the CS theory to solve the inverse problem given by

$$\hat{x} = \arg \min_x \frac{1}{2} \| y - \Phi x \|_2^2 + \lambda g(x), \quad (3)$$

where $g(x)$ is an employed prior, and $\lambda$ is a regularization parameter to balance the fidelity term and prior. To decouple the fidelity term and prior, an auxiliary parameter, $v$, is introduced, and Eq. (3) is modeled as

$$\hat{x} = \arg \min_{x,v} \frac{1}{2} \| y - \Phi x \|_2^2 + \lambda g(v), \quad \text{subject to } x = v. \quad (4)$$

In the following, we further introduce an updatable auxiliary parameter, $u$, and a manually set penalty parameter, $\rho$, and construct the augmented Lagrangian function

$$\mathcal{L}(x,v,u) = \frac{1}{2} \| y - \Phi x \|_2^2 + \lambda g(v) + u^\top (v - x) + \frac{\rho}{2} \| v - x \|_2^2. \quad (5)$$

By employing the ADMM framework, the minimization of $\mathcal{L}$ in Eq. (5) can be split into the following three sub-problems written as

$$x^{(k+1)} = \arg \min_x \frac{1}{2} \| y - \Phi x \|_2^2 + \frac{\rho}{2} \| x - (v^{(k)} - \frac{1}{\rho} u^{(k)}) \|_2^2, \quad (6)$$

$$v^{(k+1)} = \arg \min_v \lambda g(v) + \frac{\rho}{2} \| v - (x^{(k+1)} + \frac{1}{\rho} u^{(k)}) \|_2^2, \quad (7)$$

$$u^{(k+1)} = u^{(k)} + \rho (x^{(k+1)} - v^{(k+1)}). \quad (8)$$

Here, the superscript $k$ denotes the iteration number within the maximal iterations, $O$. According to Ref. [18], the solution of $x$ sub-problem in Eq. (6) represents a constraint on the estimated
results, and it can be efficiently implemented by

$$x^{(k+1)} = \left( \Phi^T \Phi + \rho I \right)^{-1} \left[ \Phi^T y + \rho \left( u^{(k)} + w^{(k)} \right) \right].$$  \hfill (9)

According to Ref. [23], the \( v \) sub-problem in Eq. (7) can be regarded as a denoising problem by using deep denoising networks such as FFDNNet [36] and DRUNet [29]. Defining \( \sigma = \sqrt{A/\rho} \), the solution of Eq. (7) can be denoted as

$$v^{(k+1)} = D_{\sigma} (x^{(k+1)} - u^{(k)}),$$  \hfill (10)

where \( \sigma \) is the estimated noise level of denoiser \( D \), and the resulting algorithm is called PnP-ADMM.

Considering the efficiency and specialization problems in conventional PnP-ADMM algorithms, the MS denoising method is further derived for MC-CUP to reduce the difficulty of adjusting the denoising parameters and make full use of the coupling information between different channels. Inspired by Ref. [30], we assume that there exists a distribution of the denoiser hyperparameter, \( \sigma \), in each iteration of PnP, which is denoted as \( p(\sigma|x^{(k+1)} - u^{(k)}) \). With the distribution of \( \sigma \) in consideration, we eliminate the hyperparameter via integral as follows

$$p\left( v^{(k+1)} | x^{(k+1)} - u^{(k)} \right) = \int p\left( v^{(k+1)} | x^{(k+1)} - u^{(k)}, \sigma \right) p\left( \sigma | x^{(k+1)} - u^{(k)} \right) d\sigma.$$  \hfill (11)

Therefore, a very straightforward method is to discretize \( \sigma \) for calculation purposes. To discretize \( \sigma \), the hyperparameter is considered obeying the discrete distribution of \( p(\sigma|x - u) \) with \( \sigma \in S \), where \( S \) is a set with finite elements of noise level. For convenience, the superscript \( (k+1) \) is omitted, therefore, Eq. (11) can be rewritten as

$$p\left( v | x - u \right) = \sum_{\sigma \in S} p\left( v | x - u, \sigma \right) p\left( \sigma | x - u \right).$$  \hfill (12)

It is easy to see that Eq. (11) is the posterior we get from the denoiser \( D_{\sigma} \) and

$$p\left( v | x - u, \sigma \right) = D_{\sigma} (x - u).$$  \hfill (13)

By rearranging the above optimization problem and denoting \( w_{\sigma} = p(\sigma|x - u), \sigma \in S \) with , the \((k+1)\)th iterative optimal solution can be obtained by

$$E_{p(v|x-u)}[v] = \sum_{\sigma \in S} w_{\sigma} D_{\sigma} (x - u).$$  \hfill (14)

Since the distribution of noise and computational error during the ADMM iteration is unknown, the distribution of \( w_{\sigma} \) is also unknown, which leads to the fact that Eq. (14) cannot be solved. However, the MC acquisition technique gives us the possibility of solving the distribution of \( w_{\sigma} \). For each channel, the collected data contains two parts of real scene data and noise data, and the real scene data of all channels are coupled with each other while the noise data are irrelevant. It can be simply deduced that after weighted MC denoising, the extracted data should be the same. On the basis of this, we can calculate the distribution of \( w_{\sigma} \), and then extract the coupling data of different channels, which is the \( E_{p(v|x-u)}[v] \) we need. Considering \( x_i, v_i \) with \( i \in \{1, 2, \cdots, M\} \), Eq. (14) can be formulated as

$$E_{p(v|x-u)}[v] = \sum_{\sigma \in S} w_{\sigma} D_{\sigma} (x_i - u_i).$$  \hfill (15)

In the case of an ideal situation without noise, all channels should get the same \( E_{p(v|x-u)}[v] \) and \( w_{\sigma} \). However, considering the noise, we can obtain \( w_{\sigma} \) by solving the following optimization
problem
\[
\min_{w_i, w_j} \left\| \sum_{\sigma \in \mathcal{S}} w_{i,\sigma} D_{\sigma} (x_i - u_i) - \sum_{\sigma \in \mathcal{S}} w_{j,\sigma} D_{\sigma} (x_j - u_j) \right\|_2^2
\]
\[\forall i, j \in \{1, 2, \cdots, M\}, i \neq j\]  
(16)

We use an iterative approach based on the least-squares optimization method to solve the problem via the CVXPY toolbox [37]. Throughout the ADMM calculation, \( S \) is kept constant while \( w_\sigma \) is solved automatically, which omits the process of manually setting the denoising parameters. Finally, the MCMS result is

\[
v = \frac{1}{M} \sum_{i=1}^{M} \sum_{\sigma \in \mathcal{S}} w_{i,\sigma} D_{\sigma} (x_i - u_i)
\]
(17)

Algorithm 1 exhibits the overall workflow of the proposed adaptive MCMS-PnP algorithm for CUP, and the entire computational flow is shown in Fig. 1(B).

---

### Algorithm 1 Adaptive MCMS-PnP algorithm for CUP

Require: \( \Phi_i, y_j, M, S, O \)

1: Initialize \( x_i^0, v^0, u^0 \)
2: for \( k = 0 \) to \( O \) do
3: for \( i = 0 \) to \( M \) do
4: Update \( x_i^{k+1} \) by Eq. (9).
5: for \( \forall \sigma \in \mathcal{S} \) do
6: Update \( v_{i,\sigma} \) by Eq. (10).
7: end for
8: end for
9: Solve optimization problem (16).
10: Update \( v^{k+1} \) by Eq. (17).
11: Output \( v \)

---

3. Results and Discussion

The adaptive MCMS-PnP algorithm is implemented in python, and Pytorch is used to invoke a pretrained network, i.e., DRUNet in this work, as the denoising module on a server with Intel Core i7-12700K CPU and NVIDIA Geforce GTX 3090 GPU. To validate the reconstruction performance of the algorithm, we create two different types of dynamic scenes for simulation in the MC-CUP modality, and take the dual-channel sampling with complementary encoding patterns, i.e., lossless encoding CUP [15], as an example. The first type is composed of 12 kinds of moving image scenarios from publicly available datasets. Each moving image scenario contains 10 pictures with the size of 256×256 pixels, and each picture is shifted right by ten pixels relative to the previous one. The second type is composed of 7 video scenarios captured by ultrahigh-speed cameras, and each of them contains 10 consecutive frames with the same size as it in the former. To create compressed 2D images by the MC-CUP for reconstruction, all scene data of different types are processed following the data acquisition indicated in Fig. 1(A). Specifically, all frames of the dynamic scene are duplicated as a pair of replicas, and each pair is encoded by the same couple of pseudo-random binary masks with elements \( \{0, 1\} \) via element-wise multiplication. Then, the encoded frames are sequentially shifted with each frame moving down one pixel relative to the previous one to simulate the temporal deflection. Finally, both frames are spatially integrated to obtain the final 2D measurements in different channels. To be closer to
Table 1. The averaged results of PSNR (dB) and SSIM by different methods on moving image scenarios.

<table>
<thead>
<tr>
<th>Data</th>
<th>IQA</th>
<th>Methods</th>
<th>GAP-TV</th>
<th>TV-BM3D</th>
<th>DeSCI</th>
<th>FFDNet</th>
<th>DRUNet</th>
<th>DRUNet-MC</th>
<th>DRUNet-MCMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.5637</td>
<td>0.5715</td>
<td>0.6083</td>
<td>0.5678</td>
<td>0.5877</td>
<td>0.7079</td>
<td>0.7607</td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6052</td>
<td>0.6245</td>
<td>0.7883</td>
<td>0.6958</td>
<td>0.7878</td>
<td>0.8867</td>
<td>0.9356</td>
<td></td>
</tr>
<tr>
<td>Boats</td>
<td></td>
<td></td>
<td>23.89</td>
<td>24.98</td>
<td>26.94</td>
<td>25.74</td>
<td>27.1</td>
<td>29.93</td>
<td>32.42</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6862</td>
<td>0.7286</td>
<td>0.7857</td>
<td>0.7489</td>
<td>0.7998</td>
<td>0.8736</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td>Cameraman</td>
<td></td>
<td></td>
<td>22.3</td>
<td>22.8</td>
<td>25.09</td>
<td>25.33</td>
<td>25.8</td>
<td>28.98</td>
<td>30.11</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6061</td>
<td>0.6384</td>
<td>0.7911</td>
<td>0.7894</td>
<td>0.8121</td>
<td>0.8594</td>
<td>0.8878</td>
<td></td>
</tr>
<tr>
<td>Castle</td>
<td></td>
<td></td>
<td>23.81</td>
<td>24.19</td>
<td>25.09</td>
<td>24.47</td>
<td>25.3</td>
<td>28.08</td>
<td>29.46</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6977</td>
<td>0.7131</td>
<td>0.7803</td>
<td>0.748</td>
<td>0.7776</td>
<td>0.8558</td>
<td>0.8975</td>
<td></td>
</tr>
<tr>
<td>Face</td>
<td></td>
<td></td>
<td>30.24</td>
<td>30.86</td>
<td>32.62</td>
<td>29.57</td>
<td>30.89</td>
<td>32.56</td>
<td>34.98</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.7848</td>
<td>0.7851</td>
<td>0.8383</td>
<td>0.7478</td>
<td>0.7756</td>
<td>0.8204</td>
<td>0.8848</td>
<td></td>
</tr>
<tr>
<td>House</td>
<td></td>
<td></td>
<td>26.57</td>
<td>27.95</td>
<td>30.29</td>
<td>29.23</td>
<td>31.47</td>
<td>33.9</td>
<td>36.68</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.7369</td>
<td>0.7762</td>
<td>0.8406</td>
<td>0.8157</td>
<td>0.8543</td>
<td>0.8711</td>
<td>0.9214</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td></td>
<td></td>
<td>23.28</td>
<td>23.57</td>
<td>25.95</td>
<td>24.82</td>
<td>25.75</td>
<td>27.94</td>
<td>29.61</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6508</td>
<td>0.6577</td>
<td>0.7379</td>
<td>0.6908</td>
<td>0.7165</td>
<td>0.7974</td>
<td>0.8558</td>
<td></td>
</tr>
<tr>
<td>Monarch</td>
<td></td>
<td></td>
<td>21.79</td>
<td>22.19</td>
<td>24.77</td>
<td>24.32</td>
<td>25.67</td>
<td>29.12</td>
<td>31.48</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6947</td>
<td>0.714</td>
<td>0.8184</td>
<td>0.8146</td>
<td>0.8591</td>
<td>0.9121</td>
<td>0.9427</td>
<td></td>
</tr>
<tr>
<td>Parrots</td>
<td></td>
<td></td>
<td>24.61</td>
<td>26.07</td>
<td>27.1</td>
<td>26.9</td>
<td>28.4</td>
<td>31.27</td>
<td>33.34</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.7563</td>
<td>0.7884</td>
<td>0.8621</td>
<td>0.8382</td>
<td>0.8733</td>
<td>0.8962</td>
<td>0.9299</td>
<td></td>
</tr>
<tr>
<td>Peppers</td>
<td></td>
<td></td>
<td>23.34</td>
<td>25.49</td>
<td>26</td>
<td>25.75</td>
<td>27.57</td>
<td>30.55</td>
<td>32.92</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.6794</td>
<td>0.7641</td>
<td>0.8117</td>
<td>0.8021</td>
<td>0.8419</td>
<td>0.8828</td>
<td>0.919</td>
<td></td>
</tr>
<tr>
<td>Sunflower</td>
<td></td>
<td></td>
<td>25.36</td>
<td>26.33</td>
<td>27.36</td>
<td>26.37</td>
<td>27.44</td>
<td>30.16</td>
<td>32.83</td>
</tr>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>0.7228</td>
<td>0.7512</td>
<td>0.8131</td>
<td>0.7871</td>
<td>0.8159</td>
<td>0.8732</td>
<td>0.9265</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>24.03</td>
<td>24.88</td>
<td>26.82</td>
<td>25.87</td>
<td>27.20</td>
<td>29.84</td>
<td>31.88</td>
</tr>
</tbody>
</table>

In the actual situation, additive white Gaussian noise (AWGN) with the noise level of $\sigma_0 = 10$, is also considered in the simulations of video scenarios. It is noted that although only AWGN is added to the simulated video scenario, the requirement of removing non-Gaussian distributed noise also exists at the beginning of the iteration [29]. In order to demonstrate the superiority of the proposed adaptive MCMS-PnP algorithm, the pre-trained SOTA DRUNet is integrated into three kinds of PnP-ADMM algorithms (ADMM is omitted in all subsequent contents), who are denoted as DRUNet for the single-channel and single-scale (SCSS) framework, DRUNet-MC for the conventional MC method [16], and DRUNet-MCMS for our proposed one, respectively. Moreover, several kinds of mainstream CUP reconstruction algorithms, including GAP-TV [38], TV-BM3D [17], DeSCI [18] and FFDNet [21], are employed for the comparison with DRUNet in simulations. Besides, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are introduced and averaged by all of the shifted frames as image quality assessment (IQA). In the moving image scenario reconstruction, the initial variables $x_0^0$, $u_0^0$ are set to zero, $v_0$ is set according to the GAP-TV result, $S = \{40, 30, 20, 10, 5\}$ and the maximal iteration number of 100 are also set. Table 1 lists the averaged PSNR and SSIM values for reconstruction results.
of all images, and the best results are highlighted in bold font. Among them, the first five columns in Methods are for SCSS denoising. It can be seen that in this case, DRUNet can achieve very close first-class results to DeSCI, which illustrates the superior performance of DRUNet compared with other regularization methods (i.e., TV, BM3D and FFDNet). Taking MC sampling in consideration, DRUNet-MC demonstrates that enhancing the data sampling rate solely by MC acquisition can significantly improve the reconstruction quality. However, it is still restricted to individual acquisition channels during reconstruction. In contrast, our proposed DRUNet-MCMS algorithm outperforms all the other methods by fully exploiting the coupling of information among multiple channels. According to the obtained IQA values, DRUNet-MCMS shows an improvement of up to 4.68 dB (0.107) and 2.04 dB (0.0457) than that of DRUNet and DRUNet-MC in averaged PSNR (SSIM), respectively.

Figure 2 shows the fifth reconstructed images of Sunflower, Man, Boat, Castle, Monarch, and Parrot scenarios reconstructed by the algorithms displayed in Table 1, respectively, together with the corresponding ground truths for comparison. As can be seen from the reconstructed images and the magnified image blocks in the red boxes, none of the SCSS methods is favorable for recovering image details, and their reconstructions all produce various degrees of image blurring and artifacts. Even with DRUNet-MC, the reconstructed results still show obviously degraded structures in Sunflower and Monarch scenarios. In contrast, DRUNet-MCMS produces sharper borders and better image details, especially for shadow structures with insignificant contrast (e.g., in Sunflower) and irregular structures (e.g., in Parrot). The main reason for the SOTA results is that the information of the image in each ADMM iteration can be fully extracted in MS denoising in DRUNet-MCMS. More importantly, the adaptive solver of denoising weights on the basis of MC-CUP can ensure that the most suitable parameters are selected at each iteration to optimize the image toward the optimal solution.
In video reconstruction, all the parameters in DRUNet-MCMS are set as the same as that in moving image scenarios. Similarly, Table 2 lists the averaged PSNR and SSIM values for reconstruction results of all videos, and the best results are also labeled in bold font. One can obtain the same conclusion as indicated in moving image scenarios, which is that DRUNet exceeds all the involved methods in the SCSS framework, and the reconstruction quality is further greatly improved in MC-CUP. It is worth noting that the proposed adaptively weighted MS denoising method, DRUNet-MCMS, can elevate 1.11 dB (0.0062) in averaged PSNR (SSIM) on the basis of DRUNet-MC. In addition, Fig. 3 displays 3 out of 10 frames (i.e., frames 3, 6 and 9) in the videos of Welding (A) and Detonator (B) reconstruction results by DRUNet, DRUNet-MC, DRUNet-MCMS, respectively, together with the ground truths. The values under each subfigure represent the PSNR (in dB) and SSIM corresponding to that frame. From the demonstrated results, it can be concluded that the DRUNet-MCMS method has the best performance in terms of recovery of background and details for CUP. Please refer to Supplemental 1 for the variation of multi-scale denoising weights with the increase of iteration for the Welding video scenario.

![Fig. 3. Selected video reconstruction frames of (A) Welding, (B) Detonator by DRUNet, DRUNet-MC, DRUNet-MCMS, together with the ground truths for comparison. A reconstruction movie of Welding and Detonator corresponding to the ground truths is provided in Visualization 1 (see Supporting Information).](image)

In the DRUNet and DRUNet-MC methods, when the collected data contains noise with noise level of $\sigma_0$, the denoising parameters in Eq. (10) need to be set larger than $\sigma_0$ in each iteration, otherwise the recovery images will be contaminated by noise leading to a worse result. On the other hand, the oversized denoising parameter in the ADMM iteration will cause the recovered results become over smooth and blurred, as shown in the results obtained by DRUNet and DRUNet-MC algorithms in Fig. 3. In contrast, the range of denoising parameters $S$ still does not need to be changed in the proposed DRUNet-MCMS algorithm, which confirms that the MCMS-PnP algorithm has stronger robustness to noise. It is worth noting that this is more favorable in experimental data reconstruction.

As shown above, it has been proven from numerical simulations that the MCMS-PnP algorithm brings higher image quality and possesses stronger noise immunity in the image reconstruction for CUP than the conventional algorithms. Next, we apply the proposed DRUNet-MCMS algorithm to recover transient scenes obtained by our home-built MC-CUP system for experimental validation. As indicated in Fig. 4(A), an ultrafast optical scene as the object is collected by an objective lens (OL) and imaged onto a DMD (Texas Instruments, DLP Light Crafter 3000), on which a static pseudo-random binary pattern is loaded. Then the ultrafast scene is split into two beams by the DMD along different directions of $\pm 12$ degrees, and both of the replicas are encoded by a pair of complementary masks, thus producing a dual-channel encoding. After that, the encoded
Table 2. The averaged results of PSNR (dB) and SSIM by different methods on video scenarios.

<table>
<thead>
<tr>
<th>Data</th>
<th>Methods</th>
<th>GAP-TV</th>
<th>TV-BM3D</th>
<th>DeSCI</th>
<th>FFDNet</th>
<th>DRUNet</th>
<th>DRUNet-MC</th>
<th>DRUNet-MCMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welding</td>
<td>PSNR</td>
<td>26.4</td>
<td>26.76</td>
<td>28.74</td>
<td>27.71</td>
<td>28.22</td>
<td>30.63</td>
<td>31.31</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.6592</td>
<td>0.7056</td>
<td>0.7623</td>
<td>0.7593</td>
<td>0.7708</td>
<td>0.8212</td>
<td>0.8283</td>
</tr>
<tr>
<td>Energy</td>
<td>PSNR</td>
<td>28.97</td>
<td>31.51</td>
<td>34.59</td>
<td>33.82</td>
<td>34.29</td>
<td>36.19</td>
<td>37.73</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.5024</td>
<td>0.7108</td>
<td>0.8925</td>
<td>0.968</td>
<td>0.9665</td>
<td>0.9635</td>
<td>0.9769</td>
</tr>
<tr>
<td>Filament</td>
<td>PSNR</td>
<td>25.16</td>
<td>25.93</td>
<td>28.85</td>
<td>29.19</td>
<td>30.64</td>
<td>34.63</td>
<td>36.11</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.5896</td>
<td>0.7645</td>
<td>0.8172</td>
<td>0.9554</td>
<td>0.964</td>
<td>0.9633</td>
<td>0.9764</td>
</tr>
<tr>
<td>Detonator</td>
<td>PSNR</td>
<td>28.1</td>
<td>29.35</td>
<td>32.31</td>
<td>29.75</td>
<td>30.42</td>
<td>32.92</td>
<td>33.62</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.697</td>
<td>0.8053</td>
<td>0.9169</td>
<td>0.9074</td>
<td>0.9098</td>
<td>0.934</td>
<td>0.9246</td>
</tr>
<tr>
<td>Blowtorch</td>
<td>PSNR</td>
<td>26.49</td>
<td>28.47</td>
<td>32.18</td>
<td>31.12</td>
<td>32.61</td>
<td>35.5</td>
<td>36.23</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.4651</td>
<td>0.661</td>
<td>0.7156</td>
<td>0.9431</td>
<td>0.9554</td>
<td>0.9593</td>
<td>0.9694</td>
</tr>
<tr>
<td>Match</td>
<td>PSNR</td>
<td>27.94</td>
<td>28.99</td>
<td>32.07</td>
<td>33.1</td>
<td>34.19</td>
<td>37.99</td>
<td>38.55</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.7159</td>
<td>0.8351</td>
<td>0.9091</td>
<td>0.891</td>
<td>0.9223</td>
<td>0.9488</td>
<td>0.9524</td>
</tr>
<tr>
<td>Water</td>
<td>PSNR</td>
<td>27.27</td>
<td>28.51</td>
<td>30.7</td>
<td>30.35</td>
<td>31.64</td>
<td>34.54</td>
<td>36.63</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.6546</td>
<td>0.777</td>
<td>0.8717</td>
<td>0.9343</td>
<td>0.9426</td>
<td>0.9591</td>
<td>0.9645</td>
</tr>
<tr>
<td>Average</td>
<td>PSNR</td>
<td>27.19</td>
<td>28.50</td>
<td>31.35</td>
<td>30.72</td>
<td>31.72</td>
<td>34.63</td>
<td>35.74</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.6120</td>
<td>0.7513</td>
<td>0.8408</td>
<td>0.9084</td>
<td>0.9188</td>
<td>0.9356</td>
<td>0.9418</td>
</tr>
</tbody>
</table>

An ultrafast scene of each channel is transferred through a 4f imaging system composed of L1 and L2 or L3 and L4, and reflected by a reflecting prism. Finally, both replicas are sent into a streak camera (Hamamatsu, C7700) with the slit fully opened, which is a scientific instrument that is able to capture ultrafast scenes based on the photo-electric effect and temporal shearing technique [33,34]. The streak camera and transient scenes are precisely synchronized in time by a digital delay generator (Stanford Research Systems, DG645) to temporally shear and spatiotemporally integrate them. Combining our proposed MCMS-PnP algorithm with the home-built MC-CUP system, we measured the spatiotemporal evolution of a spatially modulated E-shaped picosecond laser pulse and achieved volumetric (i.e., x-y-z) imaging of a 3D ladder structure by the time of flight (ToF) detection experimentally.

In the first scene, a single laser pulse from a mode-locked Ti:Sapphire laser amplifier (Spectra-Physics, 50 fs, 0.8 mJ) was stretched up to 200 ps by a pulse stretcher. The stretched pulse was spatially expanded to illuminate a hollow letter ‘E’ fabricated in a black nylon plate. Photons inside the shape could pass through the plate, while those on the outside were blocked. The resulting E-shaped laser pulse was projected onto a thin sheet of white paper for scattering, and further observed by the MC-CUP system. The three DRUNet-based algorithms are used to reconstruct the spatiotemporal evolution of the E-shaped laser pulse, and corresponding normalized results are shown in Fig. 4(B). 10 representative frames are selected from the reconstructed scenes with a time interval of 30 ps, where the reconstructed results for DRUNet, DRUNet-MC and DRUNet-MCMS are displayed for comparison. It needs to declare that the snapshot in the first row is generated by combining the static 2D image and dynamic 1D intensity evolution obtained by the streak camera as reference, and the DRUNet results in the second row are from one of the channels in MC-CUP. From Fig. 4(B), one can see that the DRUNet results show obvious incompleteness in shape, and the results of DRUNet-MC and DRUNet-MCMS are closer to the ground truth in terms of spatial structure. Moreover, the MS strategy has an absolute advantage in improving the dynamic range of reconstructed images as shown in frames at 90 and
Fig. 4. (A) Experimental system configuration of MC-CUP. OL, objective lens; DMD, digital micromirror device; M1-M4, mirror; L1-L4, lens; RP, reflecting prism; SC, streak camera. (B) Static image of 2D measurements overlay the intensity evolution of 1D measurements by streak camera as reference snapshot and reconstructed spatiotemporal images by DRUNet, DRUNet-MC, DRUNet-MCMS of E-shaped laser pulse; (C) PSNR and SSIM distribution calculated by DRUNet, DRUNet-MC, DRUNet-MCMS with ground truth, respectively; (D) the normalized intensities for ground truth (black line), DRUNet (green line), DRUNet-MC (yellow line) and DRUNet-MCMS (green line) along the white dotted line in the reconstructed images with the time of 60 ps in (B). A spatiotemporal evolution movie of E-shaped laser pulse corresponding to the reference snapshot is provided in Visualization 2 (see Supporting Information).

−90 ps. To quantitatively compare the reconstruction performance, we calculated the PSNR and SSIM values for the middle 8 frames of Fig. 4(B), and the results are shown in Fig. 4(C). It can be easily recognized from Fig. 4(C) that DRUNet-MCMS outperforms the other two methods in both PSNR and SSIM, especially under the condition of low signal to noise ratio (SNR). To further illustrate the improvement in spatial details, we plotted the normalized intensities along the white dotted lines for the frames at 60 ps in Fig. 4(B), and the results are given in Fig. 4(D). As expected, the spatial intensity distribution of DRUNet-MCMS still holds the highest SNR for the three peaks. Besides, the root-mean-square error (RMSE) of DRUNet, DRUNet-MC and DRUNet-MCMS calculated from Fig. 4(D) are 0.1930, 0.1822 and 0.1215, respectively, indicating the closest fidelity to the ground truth for DRUNet-MCMS.

In the second scene, we measured the surface profile of a volumetric object by capturing the backscattered photons from the surface via time-of-flight (ToF) detection. As shown in Fig. 5(A), a 3D ladder structure with alternating lengths is fabricated using white nylon and each step has a width and height of 3 mm, respectively. A 50-fs laser pulse with spatial expansion was employed to illuminate the 3D ladder structure. The MC-CUP system illustrated in Fig. 4(A) was placed perpendicular to the x-y plane of the ladder structure and collected the backscattered photons from the surface. The depth information of the object can be further obtained by measuring the round-trip ToF signal of the fs laser pulse, and it is given by $z = c \times \frac{t_{ToF}}{2}$, where $c$ and $t_{ToF}$
denote the speed of light and propagation time, respectively. Similarly, the reconstructed images via the three DRUNet-based algorithms are shown in Fig. 5(B). For comparison, 6 representative images out of 60 frames at the time delays of 0, 20, 40, 60, 80 and 100 ps are given, and the white dotted boxes identify the accurate sizes and locations of the steps. It can be visually observed that the DRUNet-MCMS algorithm results in more uniform intensity distribution. Moreover, the 3D morphology of the object based on the DRUNet-MCMS result with a temporal resolution of 2 ps is reconstructed as shown in Fig. 5(C), and the obtained 3D morphology shows a perfect match with the actual model. In order to quantitatively analyze the volumetric detection error, we compared the overlap degree between the corresponding images of different steps obtained from the three algorithms and the ground truth in Fig. 5(B) by calculating the Intersection over Union (IoU) [39] indices, and the results are displayed in Fig. 5(D). It is clearly seen that the IoU values of DRUNet-MCMS are much higher than those of the other two methods, and the mean IoU values of DRUNet, DRUNet-MC, DRUNet-MCMS is 0.46, 0.51, 0.69, respectively. In other words, the IoU value of DRUNet-MCMS have more than 0.18 improvement over the other two methods. Obviously, DRUNet-MCMS can provide an effective method to further improve the accuracy of volumetric detection.
4. Conclusion

In summary, we have developed a MS denoising PnP algorithm, shorten as MCMS-PnP, for MC-CUP to reconstruct ultrafast scenes, which is capable of realizing adaptive parameter tuning without manual settings. To test the reconstruction capability of this PnP algorithm, we first performed two numerical simulations, and the results show that the proposed method has the best performance compared to previous algorithms used in CUP, and the method can provide both higher image reconstruction quality and stronger noise immunity. Furthermore, two kinds of scenes are measured with our home-built MC-CUP system. The proposed method performs very well on the image details and shadow textures of the experimental data reconstruction results, which can effectively improve the spatial details. Moreover, more advanced denoisers or algorithms can be easily combined in this PnP framework to further improve the image reconstruction performance in the future. It can be prospected that the proposed MCMS-PnP algorithm would enable CUP to visualize ultrafast scenes with complicated spatial details, thus extending the applications in various areas, including shockwave diagnosis in inertial confined fusion [40] and implosions detection in Z-pinches [41].

Funding. National Natural Science Foundation of China (11727810, 12034008, 12074121, 62105101, 62175066, 91850202, 92150301); Science and Technology Commission of Shanghai Municipality (20ZR1417100, 21JIM0010700, 21XD1400900).

Disclosures. The authors declare no conflict of interest.

Data availability. Data and source code underlying the results presented in this paper are available in Ref. [42].

Supplemental document. See Supplement 1 for supporting content.

References
