2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA) | 979-8-3503-4503-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/DSAA60987.2023.10302598

Adaptive Compressed Sensing for Real-Time Video Compression, Transmission, and Reconstruction

Yaping Zhao[1,2,†]
zhaoyp@connect.hku.hk

Qunsong Zeng[1]
qszeng@connect.hku.hk

Edmund Y. Lam[1,2,*]
elam@eee.hku.hk

[*]*Corresponding Author*
[1]*Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong*
[2]*ACCESS — AI Chip Center for Emerging Smart Systems, Hong Kong SAR*

https://github.com/IndigoPurple/ART_Video

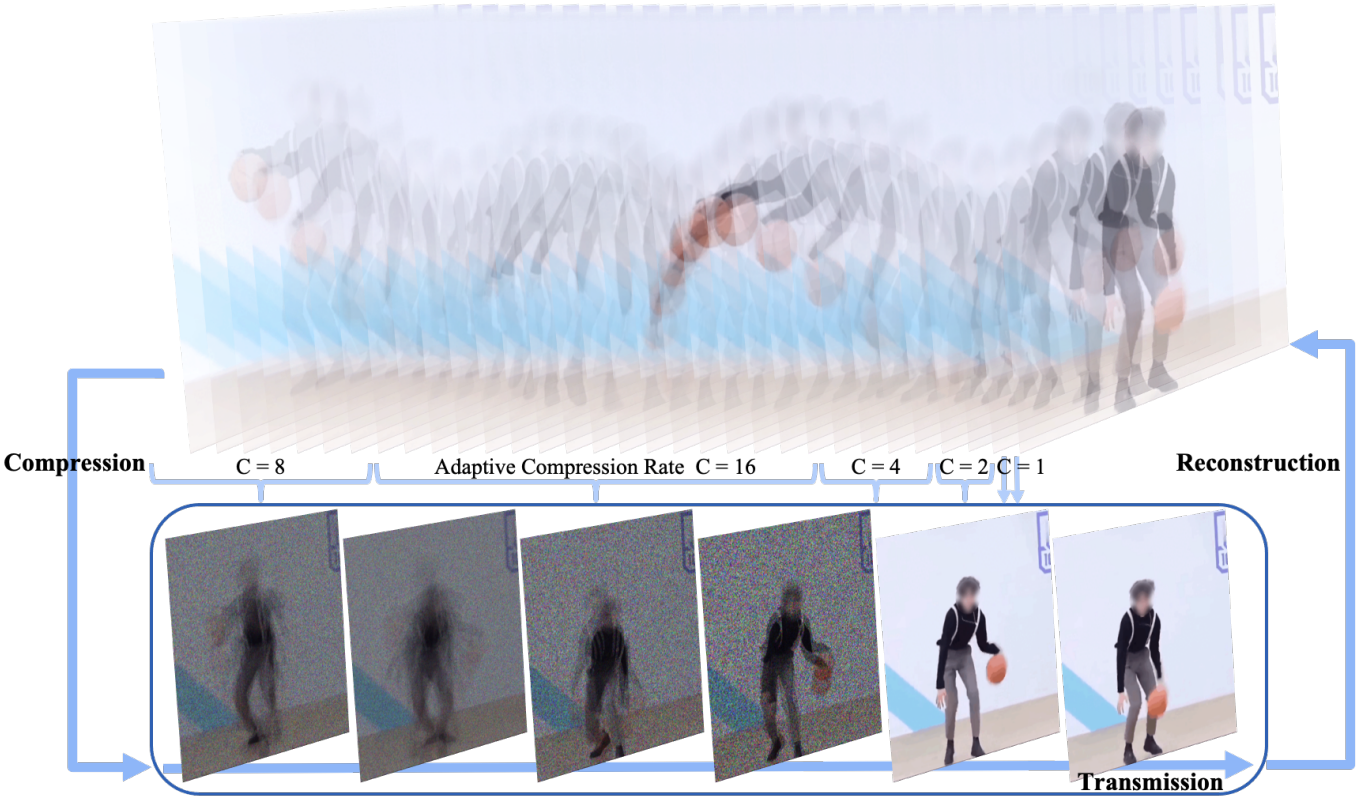


Fig. 1: We propose a real-time video compression and transmission framework that adaptively adjusts the compression rate based on the channel state to reduce communcation overhead while warranting the reconstruction quality. The original video is compressed into a low-dimensional sparse measurement, which is then transmitted to the receiver for reconstruction, ultimately recovering the high-dimensional dense video frames.

***Abstract*—The real-time transmission of videos with both high resolution and high frame rate is challenging, due to the limited storage space and significant communication overhead. To meet the real-time requirement, these issues are usually tackled by video quality reduction, which compromises the user experience. While previous methods, such as video compressed sensing, have attempted to address these issues, they often employ a fixed compression rate without considering the varying channel gain and do not adequately address the real-time transmission requirements. To mitigate these shortcomings, we propose an adaptive compressed sensing framework that optimizes the compression rate based on the channel state. This approach equivalently optimizes the video quality while ensuring real-time transmission by reducing communication overhead and thus latency. The feasibility and performance of our method are validated and discussed through extensive experiments on both classic and custom datasets.**

***Index Terms*—Compressed Sensing, Video Compression, Video Transmission, Video Reconstruction, Video Processing**

I. Introduction

Vision, as a vital human sensory modality for perceiving the surrounding environment, is predominantly documented and preserved in the medium of videos. The primary objective is to accomplish exceptional levels of resolution and frame rate in video capture. On one hand, higher resolution contributes to a more distinct spatial representation, enabling the visualization of intricate details, textures, and objects of smaller scale. On the other hand, elevated frame rates enhance the temporal depiction, facilitating smoother portrayal of object motion. Consequently, these videos possess the capability to unveil the trajectories of high-speed objects, enabling precise recording and comprehensive analysis.

Despite the advancements in imaging technology, which have enabled the capture of high-resolution and high-frame-rate videos, several underlying challenges persist. Firstly, there exists an inherent trade-off between spatial and temporal resolution in imaging, necessitating careful consideration during the capture process. Secondly, the simultaneous maintenance of high resolution and high frame rates in videos typically requires substantial storage capacity. Thirdly, the significant size of these high-resolution videos poses inevitable obstacles in terms of transmission. Specifically, real-time streaming of such videos presents considerable difficulties due to the substantial latency introduced by the transmission of large files, particularly in the presence of poor communication channel conditions. Consequently, viewers at the receiving end may encounter visual stuttering caused by the sluggish transmission.

In recent years, there has been a growing interest in the application of compressed sensing in imaging, specifically video compressed sensing, which enables the capture of high-speed videos at low frame rates through temporal compression [1]–[4]. However, there is still a lack of research focused on video compressed sensing systems that dynamically adjust the compression rate to accommodate real-time transmission requirements. Current imaging methodologies commonly employ a fixed compression rate and do not consider the fluctuating channel gain. While video codecs like H.264/MPEG4 [5] can be employed to reduce video storage size and facilitate real-time network transmission by generating smaller file sizes, these methods encounter challenges when the channel state is poor. In such cases, additional video compression often involves reducing the spatial resolution or lowering the frame rate, compromising the quality of the video and degrading the viewing experience.

To address these challenges, in this paper, we propose an adaptive compressed sensing framework for real-time video compression, transmission, and reconstruction, as Fig. 1 shows. On the sender's end, video signals are captured and then adaptively compressed based on channel state, transforming the original video into a low-dimensional sparse signal. This signal is then transmitted to the receiver, where it is reconstructed and restored into a high-dimensional dense video signal. Our method exhibits adaptability, dynamically responding to the variability of channel conditions and intelligently adjusting the compression rate for the best possible video quality. By effectively balancing the trade-offs between compression rate, transmission latency, and video quality, it promises a further step forward in the efficient transmission and reconstruction of high-dimensional video signals.

It is worth noting that our compression approach can be implemented using hardware-based sampling and encoding of video signals [4]. Furthermore, it can be synergistically integrated with existing software algorithms such as H.264/MPEG, thereby further reducing video file sizes. Our method is designed to complement, rather than entirely replace, current video compression algorithms.

In a nutshell, our main contributions are as follows:

- The framework incorporates time-varying channel states and introduces an adaptive compression rate during video compression. This dynamic approach enables us to minimize latency, ensuring real-time performance, while simultaneously optimizing video quality.
- Our proposed approach incorporates a synergistic integration of conventional software compression algorithms with hardware-based compression techniques. By combining these two methods, we can achieve an optimized layer of redundancy reduction, resulting in a more compact and efficient video representation.
- We extensively validated the feasibility and effectiveness of our proposed approach through a wide range of experimental settings, encompassing both classic and custom datasets.

II. Related Works

A. Video Compressed Sensing

Recent years have witnessed the emergence of a novel application of compressed sensing known as video compressed sensing, designed for capturing high-speed videos at low frame rates [1]–[4]. These video compressed sensing systems utilize per-pixel modulation during each integration time period, effectively circumventing the trade-off problem concerning spatial-temporal resolution in video capture:

$$\mathbf{Y} = \sum_{i=1}^{C} \mathbf{M}_i \odot \mathbf{X}_i + \mathbf{E}, , \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$ is a video sub-sequence, $\mathbf{M} \in \mathbb{R}^{H\times W\times C}$ is the coding mask, $\mathbf{E}$ is the measurement noise; $\mathbf{X}_i \in \mathbb{R}^{H\times W}, \forall i = 1,\dots,C$, is the i-th video frame and similarly $\mathbf{A}_i$ the corresponding i-th mask; H and W are the height and width of the video, respectively, and C is the compression rate. For convenience, we define $\boldsymbol{y} = \mathrm{vec}(\mathbf{Y})$, $\boldsymbol{x}_i = \mathrm{vec}(\mathbf{X}_i), \boldsymbol{x} = [\boldsymbol{x}_1^\top, \dots, \boldsymbol{x}_C^\top]^\top$, $\mathbf{A}_i = \mathrm{diag}(\mathrm{vec}(\mathbf{M}_i))$, $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_C]$, $\boldsymbol{e} = \mathrm{vec}(\mathbf{E})$, and use the following forward model [4], [6]:

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{e}, \tag{2}$$

Upon capture, high-speed videos are reconstructed using a variety of compressed sensing inversion algorithms [6]–[19].

Specifically, these algorithms obtain the estimated value $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$ by solving the following problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2 + R(\boldsymbol{x}), \tag{3}$$

where $\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2$ is the fidelity term and $R(\boldsymbol{x})$ is the regularization term.

By introducing an auxiliary parameter $\boldsymbol{v}$, the unconstrained optimization in Eq. (3) can be converted into:

$$(\boldsymbol{x}, \boldsymbol{v}) = \arg\min_{\boldsymbol{x},\boldsymbol{v}} \frac{1}{2}\|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + R(\boldsymbol{v}),\ s.t.\ \boldsymbol{x} = \boldsymbol{v}. \tag{4}$$

Using the generalized alternating projection (GAP) [20], Eq. (4) could be divided into the following two steps:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{v}^{(k)} + \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{v}^{(k)}), \tag{5}$$

$$\boldsymbol{v}^{(k+1)} = \mathcal{D}^{(k+1)}(\boldsymbol{x}^{(k+1)}). \tag{6}$$

where the superscript (k) denotes the iteration number, and $\mathcal{D}$ is a denoiser, *e.g.*, [21]. Eq. (5) can be solved efficiently due to the special structure of $\mathbf{A}$ in snapshot compressed sensing [22], which is utilized in GAP-TV [23].

B. Adaptive Measurements

These video compressed sensing systems mentioned in Sec. II-A were initially designed for a fixed time compression rate. Researchers have introduced the concept of adaptive temporal compressive sensing to address the variation in temporal correlation between video frames [24], [25]. This approach enables the development of a compressed sensing video system that captures the associated motion's velocity within the scene, allowing for an optimized representation of the dynamic content. Recently, saliency-aware adaptive compressive sensing is also investigated [26]. However, exploration of adaptive compression rate C considering the channel state during real-time video transmission is lacking.

C. Video Compression and Transmission

While advanced imaging systems permit the acquisition of high-quality videos [27]–[32], they are still facing the challenge of real-time communication with stringent latency requirements and limited memory occupation. To enhance video storage and transmission, the adoption of video codecs like H.264/MPEG4 [5] is beneficial. These codecs compress video files, resulting in smaller sizes, which in turn facilitates real-time transmission. By reducing file sizes, these codecs optimize bandwidth utilization and contribute to achieving efficient real-time video streaming over the link. However, when the channel state is poor, to further compress the video, conventional approaches are either to lower the spatial resolution or reduce the frame rate [33].

In Sec. II-A, we introduce video compressed sensing systems as an alternate avenue for video compression. Here, we delve deeper into their data compression mechanism. In accordance with [4], we consider a compact set $\mathcal{Q} \subset \mathbb{R}^{HWC}$, with $\boldsymbol{x} \in \mathcal{Q}$ comprising C frames $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_C$ in $\mathbb{R}^{HW}$. A lossy compression code of rate $l \in \mathbb{N}^+$ for $\mathcal{Q}$ is identified by its encoding mapping f, where

$$f : \mathcal{Q} \to \mathcal{T}, \tag{7}$$

with the set $\mathcal{T}$ defined as

$$\mathcal{T} = \left\{1, 2, \ldots, 2^{HWCl}\right\}, \tag{8}$$

and its *decoding mapping* g, where

$$g : \mathcal{T} \to \mathbb{R}^{HWC}. \tag{9}$$

The *average distortion* between $\boldsymbol{x}$ and its reconstruction $\hat{\boldsymbol{x}}$ is defined as [4]:

$$a(\boldsymbol{x}, \hat{\boldsymbol{x}}) \triangleq \frac{1}{HWC}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2\,. \tag{10}$$

Following [4], let $\tilde{\boldsymbol{x}} = g(f(\boldsymbol{x}))$. The distortion associated with the code (f, g) is symbolized by d, which is the maximum of all possible distortions, as per the subsequent equation. The codebook $\mathcal{C}$ of this code is formulated as:

$$d \triangleq \sup_{\boldsymbol{x}\in\mathcal{Q}} a(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \sup_{\boldsymbol{x}\in\mathcal{Q}} \frac{1}{HWC}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2^2. \tag{11}$$

Let $\mathcal{C}$ denote the *codebook* of this code defined as

$$\mathcal{C} = \{g(f(\boldsymbol{x})) :\ \boldsymbol{x} \in \mathcal{Q}\}. \tag{12}$$

Since the code's rate is l, $|\mathcal{C}|$ cannot exceed 2^{HWCl}. Contemplating a collection of compression codes $(f_l, g_l)_l$ for the set $\mathcal{Q}$, indexed by their rate l, we define the deterministic distortion-rate function for this code collection as:

$$d(l) = \sup_{\boldsymbol{x}\in\mathcal{Q}} \frac{1}{HWC}\|\boldsymbol{x} - g_l(f_l(\boldsymbol{x}))\|_2^2. \tag{13}$$

Consequently, the deterministic rate-distortion function for this code collection is defined as:

$$l(d) = \inf\{l : d(l) \le d\}. \tag{14}$$

The theoretical derivation presupposes that the compression code $g(f(\boldsymbol{x}))$ yields a codeword in $\mathcal{C}$ that is closest to $\boldsymbol{x}$:

$$g(f(\boldsymbol{x})) = \arg\min_{\boldsymbol{c}\in\mathcal{C}} \|\boldsymbol{x} - \boldsymbol{c}\|_2^2. \tag{15}$$

III. Method

A. Theoretical Analysis

Jalali et al. [22] proposed a Compressible Signal Pursuit (CSP)-type optimization for snapshot compressed sensing recovery. Building upon Yuan et al.'s work [4], we delve deeper into this optimization. Given a compact set $\mathcal{Q} \subset \mathbb{R}^{HWC}$ with a rate-l compression code defined by mappings (f, g) per Eq.(7)-(9), we reconstruct a signal $\boldsymbol{x} \in \mathcal{Q}$, resulting in $\hat{\boldsymbol{x}}$, through the following optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{c}\in\mathcal{C}} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{c}\|_2^2, \tag{16}$$

where $\mathcal{C}$ is determined in accordance with Eq.(12) and $\mathbf{A}$ embodies the sparse acquisition matrix.

This optimization targets the most compressible signal within the codebook, approximating the observed measurement per $\mathbf{A}$ and measurement vector $\boldsymbol{y}$. Notably, the computational feasibility of Eq.(16), similar to Eq.(15), is uncertain.

Building on Yuan et al. [4], we discuss a theorem clarifying the efficacy of video compressed sensing recovery using CSP-type optimization. The theorem relates the parameters of the code, its rate l, distortion d, and frame count C. The frame count C notably affects the compressive sampling rate and overall reconstruction quality.

Theorem 1: [4], [22] Assume that $\forall \boldsymbol{x} \in \mathcal{Q}$, $\|\boldsymbol{x}\|_\infty \leq \frac{\rho}{2}$. *Further, assume the rate-l code achieves distortion d on $\mathcal{Q}$. Moreover, assuming each element in* $\mathbf{M}$ *is drawn from the standard Gaussian distribution,* i.e., $M_{i,j,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \forall i = 1,\ldots,H; j = 1,\ldots,W; k = 1,\ldots,C$. *Let $\hat{\boldsymbol{x}}$ denote the solution of compressible signal pursuit optimization in Eq.* (16). *Assume that $\epsilon > 0$ is a free parameter, such that $\epsilon \leq \frac{16}{3}$. Then,*

$$\Pr\left(\frac{1}{HWC}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 \leq d + \rho^2\epsilon\right) > 1 - 2^{HWCl+1}\mathrm{e}^{-HW(\frac{3\epsilon}{32})^2}. \tag{17}$$

Our framework presumes the signal is constrained, with $||\boldsymbol{x}||_\infty \leq \frac{\rho}{2}$. This premise, while key for theorem proofs, reflects real-world camera-captured intensities bound by the sensor's dynamic range.

Theorem 1 states that for a compressible signal characterized by its compression rate l and distortion d, the reconstruction error, or the discrepancy between actual and reconstructed signals, is apt to remain within the distortion threshold in a video compressed sensing system.

In light of this theorem, we draw the following observations that align with prior analysis [4]:

- With fixed spatial dimensions H and W, Eq.(17) indicates that a larger C leads to increased reconstruction error, consistent with the understanding that greater compression rates compromise reconstruction quality.
- Theorem1 is rooted in the quantized space of the compressed signal. Adapting these results to a continuous space requires added investigation.
- The proof approach of Theorem 1, per [22], consists of defining an error limit within the encoded space using the concentration of measure principle and then applying a union bound. A larger ϵ encompasses a wider tail distribution, boosting both the success rate and error limit.
- The system design should consider the direct relationship between the compression rate C and reconstruction error, as informed by the signal's rate-distortion function or the (l, d) pair. A corollary detailing this connection is provided in [22].

  Corollary 1: [4], [22] Consider the same setup as in Theorem 1. Given $\eta > 0$, assume that

$$C < \frac{1}{\eta}\left(\frac{\log\frac{1}{d}}{2l}\right). \tag{18}$$

Then,

$$\Pr\left(\frac{1}{HWC}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d + 8\rho^2\sqrt{\frac{\log\frac{1}{d}}{\eta}}\right) \leq 2e^{-\frac{\log\frac{1}{d}}{5\eta}HW}. \tag{19}$$

By substituting $\epsilon = 8\sqrt{\frac{\log\frac{1}{d}}{\eta}}$ into Theorem 1, we arrive at a corollary that decouples C from the probability limit in Eq.(19). When C aligns with Eq.(18), both the error and probability are bounded by the distortion d, given the parameters H, W, ρ, and η.

This finding emphasizes a pivotal system trade-off: reducing encoding distortion d allows larger C, meaning more frames can be compressed into a single measurement. This leads to significantly reduced reconstruction error with high certainty. Thus, refining the encoding to lessen distortion can greatly enhance our compressive sensing-based video transmission system.

The findings from Theorem 1 are pertinent to video compressed sensing systems using the forward model in Eq. (2). A key assumption is the high compressibility of the high-dimensional data, a trait common in video compressed sensing systems. Notably, the reconstruction error directly correlates with signal compressibility, setting the maximum error limit within our system.

B. Video Compression, Transmission, and Reconstruction

Our proposed method accounts for the fluctuating nature of channel gain and correspondingly modulates the compression rate C, as denoted in Eq. (1).

It is assumed that the codebook $\mathcal{C}$ with all the sizes has been known at the receiver for spatial-temporal domain decoding. The transmission operates in the epoch manner. In each epoch, the device transmits the measure $\boldsymbol{y} \in \mathbb{R}^{HW}$ with bitwidth l to the receiver, where the measure $\boldsymbol{y}$ is a superposition of C frames. In this case, the overhead of transmitting $\boldsymbol{y}$ is fixed as HWl, while it carries the information regarding C frames. For tractability, we assume the channel state remains invariant during each epoch. Although the duration of one epoch can be longer than the channel coherence time when the channel is poor, it can be realized by resource allocation and power control to maintain the transmission rate in each epoch. In the i-th epoch, the achievable transmission rate is given by Shannon's equation as follows:

$$r_i = B\log_2\left(1+\gamma|h_i|^2\right), \tag{20}$$

where r_i denotes the achievable transmission rate [in bits/sec.], B is the bandwidth [in Hz], γ is the transmit signal-to-noise ratio (SNR), and $|h_i|^2$ stands for the channel power gain. Then, the latency of transmission of one measure $\boldsymbol{y}$ is calculated by

$$\Delta t_i = \frac{HWl}{r_i}. \tag{21}$$

We assume that the remote end receives, decompresses, and plays the video stream in a real-time manner, and plays N_p

frames per second at a constant speed. Thus, it requires up to $N_p\Delta t_i$ frames in this epoch. On the other hand, the number of reconstructed frames from $\boldsymbol{y}$ is C_i. To achieve real-time play, the following constraint should be satisfied:

$$C_i \geq N_p \Delta t_i. \tag{22}$$

According to Theorem 1, larger C gives higher distortion, we just set the value of $C_i = N_p\Delta t_i$, which ensures the latency requirement while enabling the high quality as possible. In this case, given compression rate C_i, the probability of distortion exceeding a certain range in epoch-i is quantified as

$$\begin{aligned}&\Pr\left(\frac{1}{HWC_i}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d+\frac{16}{3}\rho^2\Big|C_i\right)\\&\leq 2^{HWC_il+1}\mathrm{e}^{-\frac{HW}{4}},\end{aligned} \tag{23}$$

where we have set the free parameter ϵ as $\frac{16}{3}$ to minimize the upper bound. Note that the one-epoch latency is determined by the channel state in this epoch, characterized by h_i. Hence, by substituting the results in Eq.(20), Eq.(21) and Eq.(22) into Eq.(23), the upper bound of the probability is

$$\begin{aligned}&\Pr\left(\frac{1}{HWC_i}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d+\frac{16}{3}\rho^2\Big|h_i\right)\\&\leq 2\mathrm{e}^{-\frac{HW}{4}}\times 2^{\frac{H^2W^2l^2N_p}{B\log_2(1+\gamma|h_i|^2)}}.\end{aligned} \tag{24}$$

Furthermore, we consider the Rayleigh fading channel, such that the channel follows the zero-mean complex Gaussian distribution, i.e., $h_i \sim \mathcal{CN}(0,1)$. Accordingly, the channel power gain follows the exponential distribution with the probability density function:

$$f_{|h_i|^2}(z) = \mathrm{e}^{-z}. \tag{25}$$

Finally, we characterize the performance of the proposed real-time adaptive compressed video transmission scheme by quantifying the upper bound of the probability of exceeding a certain distortion:

$$\begin{aligned}&\Pr\left(\frac{1}{HWC_i}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d+\frac{16}{3}\rho^2\right)\\&=\int_0^\infty\left(\frac{1}{HWC_i}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d+\frac{16}{3}\rho^2\Big||h_i|^2\right)f_{|h_i|^2}(z)dz\\&\leq 2\mathrm{e}^{-\frac{HW}{4}}\int_0^\infty 2^{\frac{H^2W^2l^2N_p}{B\log_2(1+\gamma z)}}\mathrm{e}^{-z}dz.\end{aligned} \tag{26}$$

Unfortunately, the integral in Eq.(26) cannot converge on the range $(0,\infty)$. It indicates that the performance cannot be guaranteed without further handling the deep fading of the channel. To tackle this issue, we consider the time-out mechanism with time-threshold t_{th}. When the transmission time is longer than t_{th}, the time-out event occurs, i.e., $\Delta t_i > t_{th}$. In this case, channel gain is below a certain threshold, given that

$$|h_i|^2 < \frac{1}{\gamma}\left(2^{\frac{HWl}{Bt_{th}}}-1\right). \tag{27}$$

Then, the probability of time-out is given by

$$\begin{aligned}\Pr(t_i > t_{th}) &= \int_0^{\frac{1}{\gamma}\left(2^{\frac{HWl}{Bt_{th}}}-1\right)} f_{|h_i|^2}(z)dz\\&= 1-\exp\left(-\frac{1}{\gamma}\left(2^{\frac{HWl}{Bt_{th}}}-1\right)\right).\end{aligned} \tag{28}$$

We emphasize that the time-out probability is usually small by setting an appropriate threshold. If the event occurs, the video viewer has to wait for t_{th} period which deteriorates the quality of the user experience. Finally, we characterize the quality of the reconstructed video frames under the condition of no time-out events as follows:

$$\begin{aligned}&\Pr\left(\frac{1}{HWC_i}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2 > d+\frac{16}{3}\rho^2\Big|t_i \leq t_{th}\right)\\&\leq 2\mathrm{e}^{-\frac{HW}{4}}\int_{\frac{1}{\gamma}\left(2^{\frac{HWl}{Bt_{th}}}-1\right)}^\infty 2^{\frac{H^2W^2l^2N_p}{B\log_2(1+\gamma z)}}\mathrm{e}^{-z}dz.\end{aligned} \tag{29}$$

Though the closed-form expression is hard to derive, the integral in Eq.(29) is integrable, so that the performance of our scheme is guaranteed.

Upon obtaining a compressed measurement, we can employ a myriad of compressed sensing inversion algorithms [6]–[18] to reconstructthe video. An important point to note here is that the sensing matrix $\mathbf{A}$, as expressed in Eq.(2), can be pre-generated and stored at both the transmitter and receiver ends. This crucial step mitigates the necessity of transmitting anything other than the compressed measurements, thereby considerably reducing latency. It further lays the groundwork for the facilitation of real-time applications, a noteworthy advantage in the age of live streaming and immediate data processing.

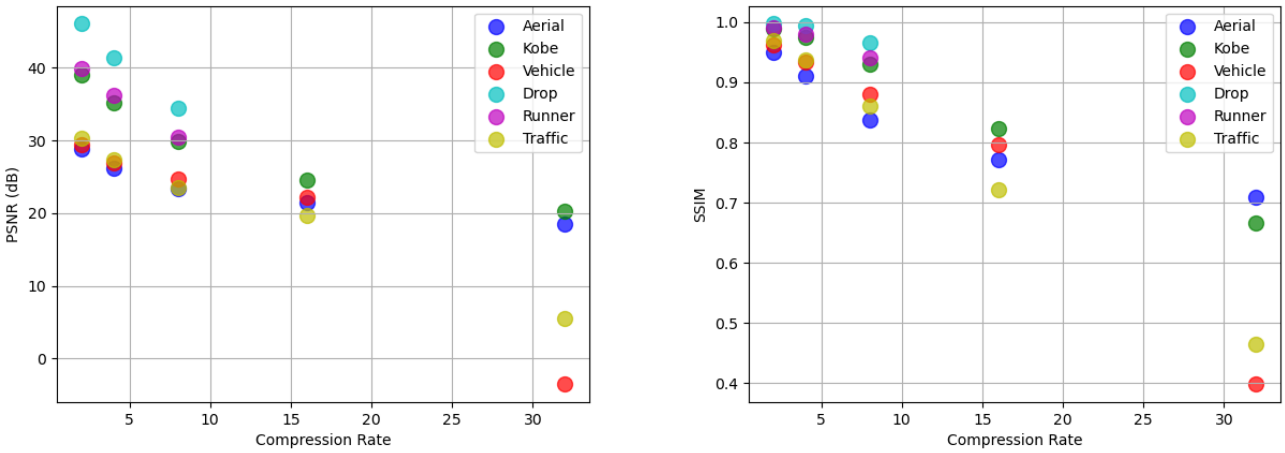
Our method exhibits adaptability, dynamically responding to the variability of channel conditions and intelligently adjusting the compression rate for the best possible video quality. By effectively balancing the trade-offs between compression rate, transmission latency, and video quality, it promises a further step forward in the efficient transmission and reconstruction of high-dimensional video signals. In the subsequent sections, we will further illustrate the practical implications and performance benefits of our approach through comprehensive experimental results and discussions.

IV. Experiment

A. Experimental Settings

In our evaluation, we used six traditional snapshot compressed sensing video datasets: `Aerial`, `Kobe`, `Vehicle`, `Drop`, `Runner`, `Traffic`. We also curated five custom videos to address the existing datasets' grayscale limitation and their short sequence lengths. These changes made our evaluation more relevant to real-world scenarios and tested our framework's adaptability for dynamic compression rate adjustments.

Initially, we examined the relationship between compression rate and reconstruction quality using the traditional datasets. Using a basketball dribbling video, we demonstrated



(a) (b)

Fig. 2: Quantitative evaluations of reconstruction quality under different compression rates on classical datasets, in terms of (a) PSNR and (b) SSIM. Displaying results from six classical datasets, it portrays a near-linear correlation between the increase in compression rate and the decrease in reconstruction quality. This trend indicates that the trade-off between compression rate and video quality is fairly predictable, enabling the anticipation of reconstruction quality for a particular compression rate.

our method's capability, as outlined in Section III-B, under variable channel capacities. Additionally, we used four high-resolution, publicly available videos `Golf`[1], `Leaf`, `Flower`, `Walk`[2] to showcase our system's efficiency in storage management.

For reconstruction, we employed the GAP-TV [23] algorithm. We tested compression rates from $C \in \{1, 2, 4, 8, 16, 32\}$ for initial validation, intending to explore more settings in future research. Our evaluation metrics included Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [34], and we tracked memory usage through file size in MB.

B. Results on Classical Datasets

A series of experiments were performed on six classical datasets to investigate the effects of varying the compression rate C on the quality of video reconstruction. Our exploration generated both qualitative and quantitative outcomes, illustrated in Fig.2 and Fig.3, respectively. These figures present a clear visual documentation of our observations and deductions, offering valuable insights into the performance of our method under varying compression rates.

Firstly, a prevalent trend was identified across the datasets: an incremental increase in the compression rate C typically leads to a corresponding decline in the quality of reconstructed videos. This observation is rather intuitive, as higher compression rates tend to discard more information, inevitably impacting the final reconstruction quality.

Secondly, Fig. 2 provides a quantitative perspective on this decrement in reconstruction quality. Interestingly, the deterioration in quality did not appear to be random or abrupt; rather, it exhibited an almost linear correlation with the increase in compression rate. This consistent behavior suggests that the compromise between the compression rate and the video quality is predictable, thereby enabling us to anticipate the potential reconstruction quality for a given compression rate.

Additionally, Fig. 3 offered a qualitative glimpse into this phenomenon. We discovered that under extremely low compression rates, such as 2 or 4, the reconstructed videos are virtually indistinguishable from the ground truth videos, exhibiting high fidelity. However, as we pushed the compression rate up to 32, the reconstructed videos bore significant distortion, thus underscoring the cost of extreme compression.

On the basis of these findings, we suggest that a judicious selection of the compression rate is instrumental in achieving a balance between data compression and reconstruction fidelity. A compression rate of 8 emerges as a viable compromise in our tests, as it seemed to uphold an acceptable level of reconstruction quality while still ensuring an appreciable degree of compression. This empirical evidence underlines the criticality of prudently determining the compression rate, bearing in mind the nuanced trade-off between compression efficiency and reconstruction quality.

C. Performance Evaluation on Custom Datasets

In order to ensure the practicality and real-world relevance of our proposed method, it is imperative to test its effectiveness on color videos, as the vast majority of videos encountered in daily life are not as grayscale as the classical datasets in Sec. IV-B. As a representative test case, we selected a vibrant color video featuring a man dribbling a basketball.

[1]https://www.quinticsports.com/sample-videos/
[2]:https://www.pexels.com/search/videos/120fps/

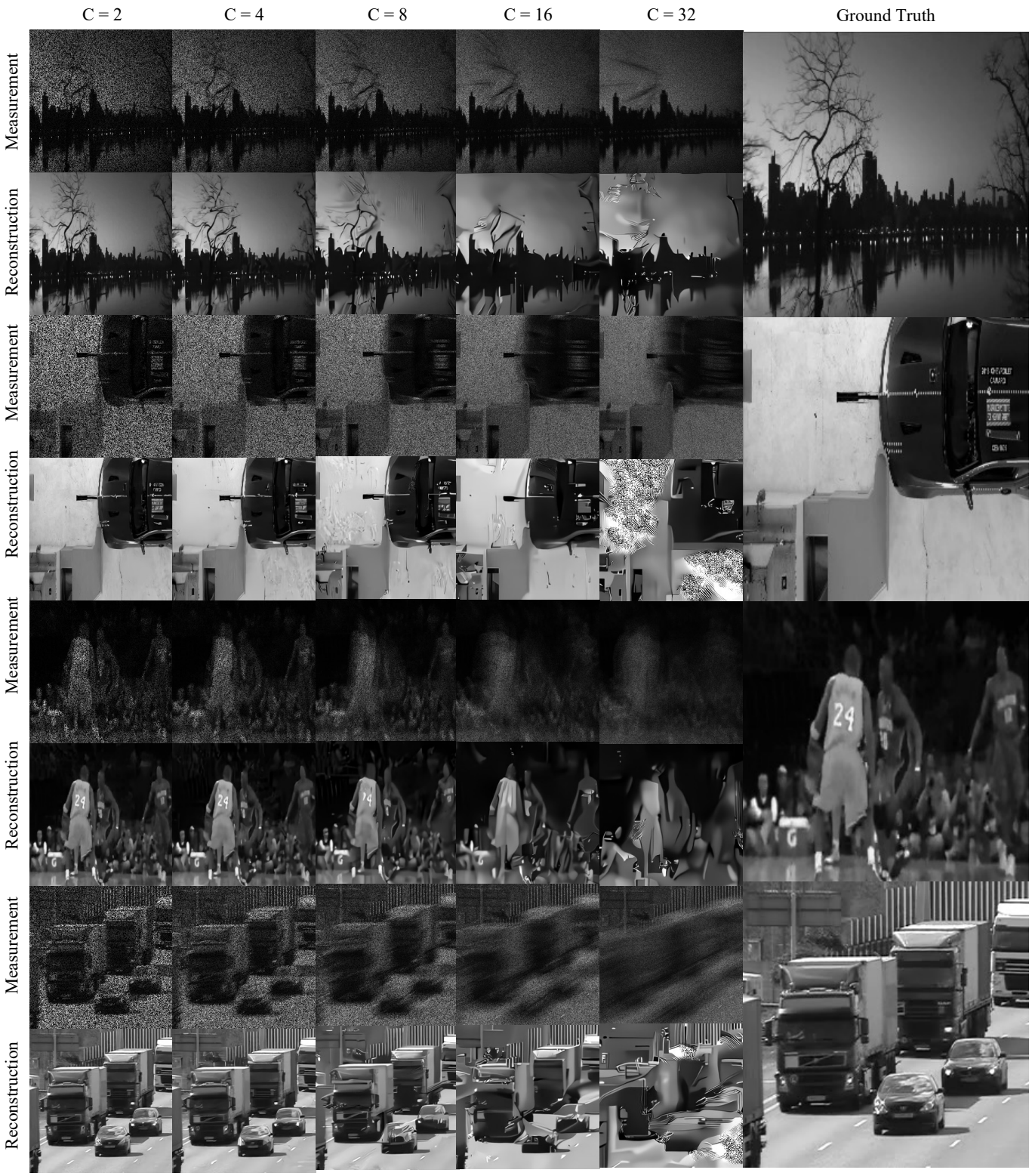


Fig. 3: Qualitative evaluations of reconstruction quality under different compression rates on classical datasets `Aerial`, `Vehicle`, `Kobe`, `Traffic`. This figure provides a qualitative analysis of the impact of varying compression rates on the quality of reconstructed videos. Using six classical datasets, it visually represents the degradation in quality as the compression rate increases. Particularly, it showcases that at very low compression rates such as 2 or 4, the reconstructed videos closely resemble the original ground truth, while at a compression rate of 32, significant distortion becomes apparent, demonstrating the compromise necessary with higher compression.

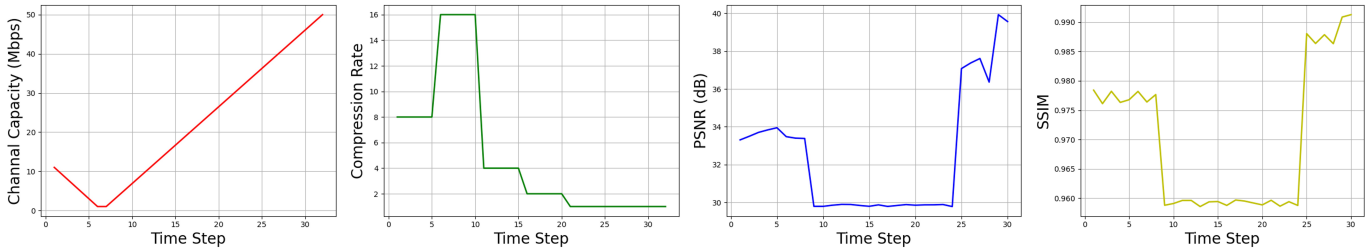


Fig. 4: The dynamics of adaptive video compression process under varying channel capacity. The left panel shows the assumed fluctuations in channel capacity over time, while the right panel illustrates the corresponding changes in reconstruction quality at different compression rates. As the channel capacity initially declines, an increase in the compression rate to 16 leads to a notable drop in reconstruction quality. However, as channel capacity rises in subsequent periods, the reduction in compression rate contributes to an improvement in the quality of the reconstructed video.

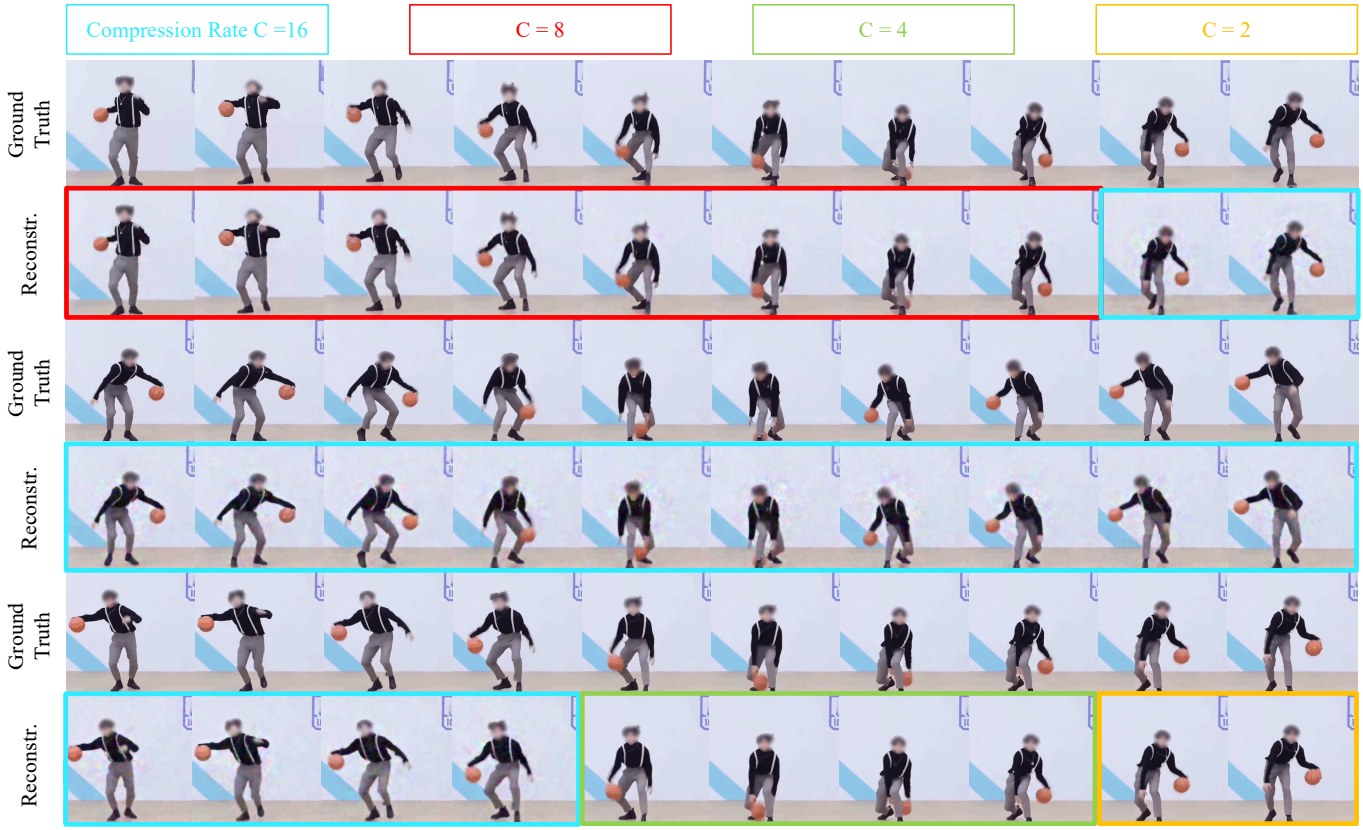


Fig. 5: The adaptive video compressed sensing process using different compression rates. Notably, at a compression rate of 16, the quality of the reconstructed video visibly deteriorates, manifesting as noise around the basketball player. In contrast, when the compression rate is set at or below 8, the reconstructed frames appear almost identical to the original ground truth, thereby affirming the efficacy of a compression rate of 8 for achieving a balanced compromise between data compression and reconstruction quality.

Data	Spatial Resolution (Width×Height)	Frame Rate (fps)	Frame Number	Size w/o Ours	Size w/ Ours	PSNR (dB)	SSIM
Golf	1024×576	2000	505	2.5 MB	1.2 MB	36.14	0.9696
Leaf	1920×1080	120	360	5.4 MB	3.1 MB	41.86	0.9918
Flower	1080×1920	120	706	64.3 MB	24 MB	28.79	0.9267
Walk	1920×1080	120	717	112.8 MB	44.5 MB	23.00	0.8105

TABLE I: The properties of the four high-resolution and high-frame-rate videos that were selected for further evaluation of our proposed framework. The videos differ in terms of spatial resolution, frame rate, and total frame count. The table also documents the remarkable compression achieved by our method, demonstrating an average file size reduction of 55%.

Initially, we set the compression rate C to 8, treating it as a reasonable starting point given our previous observations on classical datasets in Sec. IV-B. Considering the variability of real-world conditions, we assumed a dynamic channel capacity as depicted in the left panel of Fig.4. Leveraging the power of dynamic programming and taking into account the constraints provided by Sec. II-A, we successfully adapted the compression rate for each successive time slot.

During the initial phase, as the channel capacity decreased, the compression rate had to be incrementally raised as a necessary response. However, in the subsequent periods, the channel capacity demonstrated an increasing trend, allowing us to reduce the compression rate and thereby enhancing the quality of video reconstruction. Throughout this dynamic process, we modulated the compression rate as follows: $C = 8 \rightarrow 16 \rightarrow 4 \rightarrow 2 \rightarrow 1$.

From the right panel of Fig. 4, it is evident that the reconstruction quality experiences a downturn when the compression rate is escalated to 16. However, following this high compression period, the quality recovers gradually as increased channel capacity and buffered video frames permit a reduction in compression rate.

Fig. 5 provides a visual illustration of this adaptive video compression process. When the compression rate is raised to 16, noticeable noise appears around the human body in the reconstructed video, signifying a drop in performance due to increased compression. Nevertheless, at compression rates not exceeding 8, the reconstructed frames closely resemble the ground truth, thus affirming the effectiveness of our recommended compression rate $C = 8$ for balancing data compression and reconstruction quality.

D. Analysis of Memory Utilization

To further assess the potency of our proposed framework, we expanded our test base to include four additional videos. These videos, which were freely available online and had considerably more frames than the classical datasets, allowed us to explore the scalability of our method. The properties of these high-resolution, high-frame-rate videos are outlined in Table I. Given that our compression technique can be implemented through hardware-based sampling and encoding of video signals, it is not to entirely replace current video compression algorithms. Rather, by coupling it with established software algorithms such as H.264/MPEG [5], we can achieve further reductions in video file sizes. Consequently, our experiments aim to ascertain the extent to which file sizes can be further compressed beyond the base provided by H.264/MPEG, as illustrated in Table 1.

In our testing procedure, we applied our recommended compression rate $C = 8$ to these videos and stored them in the MPEG4 format. Our results, tabulated in Table I, suggest that our method can reduce the file size by an average of 55%. Taking the video titled `Walk` as an example, it has a spatial resolution of 1920×1080, a frame rate of 120, and contains 717 frames in total. Under our proposed framework, this video only requires 44.5 MB for transmission, illustrating the remarkable compression achieved.

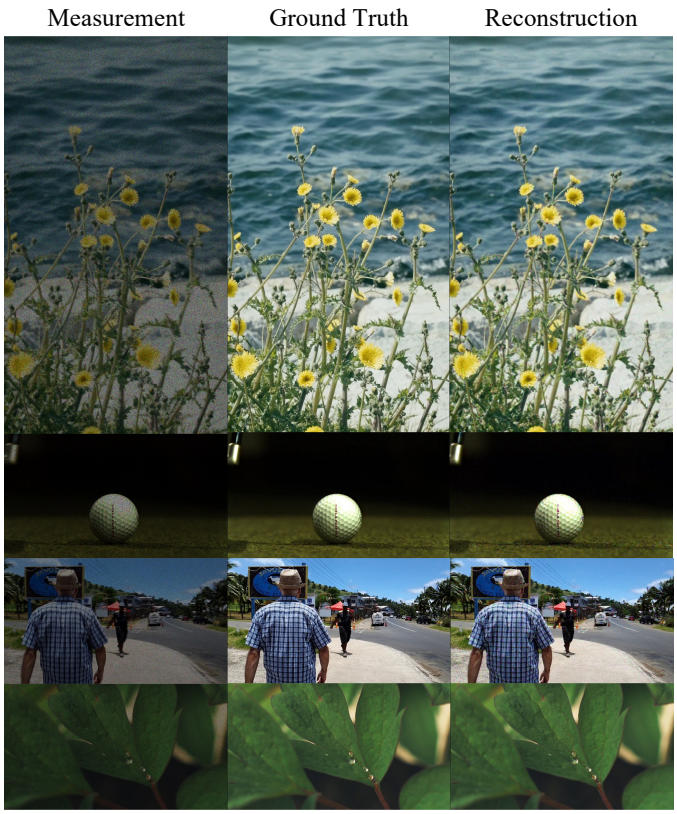


Fig. 6: The video frames reconstructed from the measurements. Despite the substantial compression, the quality of the reconstructed frames remains strikingly similar to the original ground truth frames. The visual evidence corroborates the efficiency of our proposed framework, showcasing its ability to support real-time transmission of smaller video files with lower latency, while maintaining high standards of reconstruction quality. Zoom in to see details.

A visual representation of the reconstructed video frames is provided in Fig. 6, where the quality of the reconstruction is compellingly similar to the ground truth. This not only attests to the robustness of our proposed method, but also establishes its capacity to enable real-time transmission of smaller video files with lower latency, while simultaneously preserving high-quality reconstruction.

V. Discussion

In this paper, we primarily aim to introduce an adaptive compressed sensing framework tailored for real-time video compression, transmission, and reconstruction. Notably, our compression encoding can be realized through hardware, allowing for its integration with existing software compression algorithms. As for the time cost associated with compression and reconstruction, it largely hinges on the efficiency of the current algorithms, with the bottleneck not rooted in our framework itself. To further enhance its practical applicability,

future endeavors will focus on devising more efficient algorithms that consume minimal resources, such as considering the recent advances in DEQSCI [35].

VI. Conclusion

This paper presents an adaptive compressed sensing framework for real-time video compression, transmission, and reconstruction, effectively addressing the inherent challenges associated with high-resolution, high-frame rate video processing. The proposed method dynamically adjusts the video compression rate based on varying channel quality, thereby optimizing video quality while reducing latency for real-time performance. We offer a comprehensive theoretical analysis of the influence of varying compression rates on video distortion. The framework's effectiveness and performance have been substantiated through rigorous experimentation under different settings on both classic and custom datasets, making it a viable solution for real-time capture, transmission, and playback of high-resolution, high-frame rate videos.

Acknowledgement

This work is supported in part by the Research Grants Council (GRF 17201620), by the Research Postgraduate Student Innovation Award (The University of Hong Kong), and by ACCESS — AI Chip Center for Emerging Smart Systems, Hong Kong SAR.

References

[1] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *IEEE International Conference on Computer Vision (ICCV)*, Nov 2011, pp. 287–294.

[2] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2c2: Programmable pixel compressive camera for high speed imaging," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2011, pp. 329–336.

[3] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. Brady, "Coded aperture compressive temporal imaging," *Optics Express*, vol. 21, no. 9, pp. 10 526–10 545.

[4] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 65–88, 2021.

[5] I. E. Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. John Wiley & Sons, 2004.

[6] Y. Zhao, "Mathematical cookbook for snapshot compressive imaging," *arXiv preprint arXiv:2202.07437*, 2022.

[7] Q. Yang and Y. Zhao, "Revisit dictionary learning for video compressive sensing under the plug-and-play framework," in *Seventh Asia Pacific Conference on Optics Manufacture and 2021 International Forum of Young Scientists on Advanced Optical Manufacturing (APCOM and YSAOM 2021)*, vol. 12166. SPIE, 2022, pp. 2018–2025.

[8] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for video snapshot compressive imaging," *arXiv preprint arXiv:2201.06931*, 2022.

[9] M. Qiao, Z. Meng, J. Ma, and X. Yuan, "Deep learning for video compressive sensing," *APL Photonics*, vol. 5, no. 3, p. 030801, 2020.

[10] S. Zheng, C. Wang, X. Yuan, and H. L. Xin, "Super-compression of large electron microscopy time series by deep compressive sensing learning," *Patterns*, vol. 2, no. 7, p. 100292, 2021.

[11] L. Wang, M. Cao, Y. Zhong, and X. Yuan, "Spatial-temporal transformer for video snapshot compressive imaging," *arXiv preprint arXiv:2209.01578*, 2022.

[12] Z. Cheng, B. Chen, R. Lu, Z. Wang, H. Zhang, Z. Meng, and X. Yuan, "Recurrent neural networks for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[13] Z. Meng and X. Yuan, "Perception inspired deep neural networks for spectral snapshot compressive imaging," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 2813–2817.

[14] Z. Meng, S. Jalali, and X. Yuan, "Gap-net for snapshot compressive imaging," *arXiv preprint arXiv:2012.08364*, 2020.

[15] Z. Wu, J. Zhang, and C. Mou, "Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging," *arXiv preprint arXiv:2109.06548*, 2021.

[16] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[17] X. Yuan, Y. Liu, J. Suo, F. Durand, and Q. Dai, "Plug-and-play algorithms for video snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[18] Z. Wu, C. Yang, X. Su, and X. Yuan, "Adaptive deep pnp algorithm for video snapshot compressive imaging," *arXiv preprint arXiv:2201.05483*, 2022.

[19] Z. Y. Chong, Y. Zhao, Z. Wang, and E. Y. Lam, "Solving inverse problems in compressive imaging with score-based generative models," in *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2023.

[20] X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted-$\ell_{2,1}$ minimization with applications to model-based compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 797–823, 2014.

[21] Y. Zhao, H. Zheng, Z. Wang, J. Luo, and E. Y. Lam, "MANet: improving video denoising with a multi-alignment network," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 2036–2040.

[22] S. Jalali and X. Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8005–8024, 2019.

[23] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," 2015.

[24] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using gaussian mixture models," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4863–4878, 2014.

[25] X. Yuan, J. Yang, P. Llull, X. Liao, G. Sapiro, D. J. Brady, and L. Carin, "Adaptive temporal compressive sensing for video," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 14–18.

[26] Y. Zhao and E. Y. Lam, "SASA: saliency-aware self-adaptive snapshot compressive imaging," *in preparation*, 2023.

[27] G. Li, Y. Zhao, M. Ji, X. Yuan, and L. Fang, "Zoom in to the details of human-centric videos," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

[28] Y. Zhao, G. Li, and E. Y. Lam, "Cross-camera human motion transfer by time series analysis," *arXiv preprint arXiv:2109.14174*, 2021.

[29] Y. Zhao, M. Ji, R. Huang, B. Wang, and S. Wang, "EFENet: Reference-based video super-resolution with enhanced flow estimation," in *Artificial Intelligence: First CAAI International Conference, CICAI 2021, Hangzhou, China, June 5–6, 2021, Proceedings, Part I 1*. Springer, 2021, pp. 371–383.

[30] Y. Zhao, H. Zheng, M. Ji, and R. Huang, "Cross-camera deep colorization," in *CAAI International Conference on Artificial Intelligence*. Springer, 2022, pp. 3–17.

[31] Y. Zhao, H. Zheng, J. Luo, and E. Y. Lam, "Improving video colorization by test-time tuning," in *IEEE International Conference on Image Processing (ICIP)*, 2023.

[32] Y. Zhao and E. Y. Lam, "Video colorization with test-time training," *in preparation*, 2023.

[33] S. Lu, X. Yuan, and W. Shi, "Edge compression: An integrated framework for compressive imaging processing on cavs," in *2020 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 2020, pp. 125–138.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[35] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for snapshot compressive imaging," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 3642–3650.