Adaptive Compressive Video Coding for Embedded Camera Sensors: Compressed Domain Motion and Measurements Estimation

Amit Satish Unde, *Student Member, IEEE* and Deepthi P. Pattathil, *Member, IEEE*

Abstract—This paper presents a new framework for wireless video streaming services using the resource-constrained embedded camera sensors based on block compressive sensing. We propose adaptive encoding scheme that exploits high redundancy between successive video frames to reduce transmission cost while effectively handling occlusion effects and, at the same time, maintains the simple and energy conserving encoder design. The proposed methodology adapts the compression ratio for different blocks of the non-key frame depending on temporal correlation. We also propose compressed domain motion and measurements estimation techniques to exploit the high correlation between successive frames at the decoder. The proposed motion estimation technique makes use of restricted isometry property of the sensing matrix to seek the best matching measurement vector for motion estimation as opposed to block matching in conventional video coding. In the proposed measurement estimation technique, efficient utilization of bandwidth is achieved by skipping some measurements at the transmitter side. The skipped measurements are estimated at the receiver by exploiting the correlation between CS measurements of the non-key frame and corresponding motion predicted frame using multiple regression model. Extensive simulation results on a set of diverse video sequences are presented to demonstrate the effectiveness of the proposed video coding technique.

Index Terms—Compressive sensing, adaptive video coding, compressed domain motion estimation, measurements estimation

1 Introduction

Advances in low-power and low-cost camera sensors and mobile devices have been significantly fostering wireless video streaming applications [1] such as video surveillance [2], [3], multimedia sensor network (WMSN) [4], Internet of Things (IoT) [5]. These camera sensors are limited in memory, processing capability and energy supply. In such a stringent system setting, acquisition, storage, and transmission of a huge amount of multimedia data with available bandwidth and computing power impose several critical challenges in the development of video coding technology [6]. Therefore, there is a great demand to develop low-complexity and energy-conserving encoding schemes for use in camera sensors with extremely tight resource budget, possibly at the cost of increased decoding complexity.

In surveillance applications, successive frames in video sequences are highly correlated since the camera is still (not moving), and the background is not changing. By exploiting this inherent correlation, it is possible to substantially reduce the amount of transmitted data and thereby transmission cost, allowing efficient use of the available bandwidth [6]. The traditional video coding techniques (e.g., MPEG, H.264) [7] exploits this correlation by adopting computationally complex motion estimation techniques at the encoder. As a result, the computational complexity of the encoder is significantly higher than the corresponding decoder, prohibiting its use in resource-constrained applications. This critique of the high encoding and low decoding complexity has generated great enthusiasm in the distributed video coding (DVC) techniques [8].

The DVC is able to reduce the encoder complexity by modeling the correlation between successive frames only at the decoder. This shifting of the complexity from the encoder to the decoder enables an extension of DVC to resource-deprived camera sensors. While promising, DVC necessitate feedback from the decoder for rate control which increases both transmission overhead and latency. In addition, DVC is unsuited for video acquisition and storage applications in which video is stored in compressed form and decoded at a later time. Over the past few years, compressive sensing (CS) [9] based image acquisition framework has been emerging as a promising solution to address the aforementioned problems.

The CS theory asserts that the exact recovery of the $K-$sparse signal $x \in \Re^N$ is possible from only $M = \mathcal{O}(K\ logN)$ linear and non-adaptive random measurements $y = \Phi x$, where Φ is the $M \times N$ sensing matrix with $M \ll N$. By exploiting the sparsity of the signal, CS unifies sampling and compression. The strength of CS-based image acquisition system is that the encoder can be made inexpensive and signal-independent, at the expense of high decoding complexity. This is because the decoder solves a large scale optimization problem to recover the signal. This asymmetric design is

• *The authors are with the Department of Electronics and Communication Engineering, National Institute of Technology, Calicut, Kerala 673601, India. E-mail: amitsunde@gmail.com, deepthi@nitc.ac.in.*

Manuscript received 19 Oct. 2017; revised 29 Apr. 2019; accepted 25 June 2019. Date of publication 5 July 2019; date of current version 31 Aug. 2020. (Corresponding author: Amit Satish Unde.)
Digital Object Identifier no. 10.1109/TMC.2019.2926271

highly beneficial for image acquisition system whenever the encoder is severely resource constrained [10], [11].

1.1 Previous Work

There has been significant progress in video CS with a focus on leveraging the correlation between successive frames at the decoder. Pudlewski et al. [10] designed CS based low complexity video coding scheme together with transmission and channel coding rate control to maximize quality of the received video. Aswin et al. [12] proposed a methodology for video CS by modeling the scene evaluation as a linear dynamical system (LDS) and estimated the LDS parameters from CS measurements. The framework for video CS based on Gaussian mixture model (GMM) with online learning to exploit spatio-temporal correlation was proposed in [13]. The video CS algorithms based on total variation [14] assume sparsity in gradients of temporally correlated video frames and minimizes gradients to recover the video frames.

The dictionary learning based video CS reconstruction methods have been reported in [15]. These methods aim to learn the dictionary either offline from training videos or adaptively using recovered key frames and seek the sparsest representation of a video patch. Mun and Fowler proposed block compressive sensing based recovery algorithm for video CS [16]. Motion estimation (ME) and motion compensation (MC) techniques were incorporated to take advantage of temporal correlation. Chen Zhao et al. [17] used multihypothesis technique for motion prediction and proposed reweighed residual-based recovery process to improve quality of motion predicted frames. Distributed video coding architecture proposed in [18] combines DVC and CS theory wherein the correlation was exploited through the interframe sparsity model.

1.2 Relation to Prior Work

Studies have shown that the recovery quality of non-key frames in video CS depends heavily on generation and utilization of side information at the decoder. In other words, the accuracy in predicting temporal correlation reflects in the recovery performance. While dictionary learning and GMM based algorithms require access to an appropriate training data, maintaining low latency is extremely challenging due to computationally expensive online learning. The algorithms that employ ME/MC techniques require initial recovery of non-key frames prior to correlation estimation. However, since non-key frames are sampled at much lower bit-rate, initial non-key frame recovery experiences high-frequency oscillatory artifacts which result in poor recovered video quality.

After exhaustive study, we found that the prediction of the accurate side information from limited available measurements of non-key frames while maintaining low latency is a highly challenging task. In addition, since the correlation between successive video frames is exploited only at the decoder, existing algorithms suffer heavy losses in the recovery performance due to occlusion effects. More precisely, temporal correlation prediction techniques employed in current CS recovery algorithms are not robust to occlusion problems which cause significant performance degradation and it is substantially unexplored in the existing literature.

1.3 Our Contribution

In this paper, we propose a novel architecture for video CS that can substantially improve the recovery performance, while maintaining the low complexity of the encoder. The proposed solution is based on block compressive sensing (BCS) paradigm in which the video frame is divided into non-overlapping blocks and each block is sampled independently. We propose a simple and inexpensive adaptive encoding scheme that enables the encoder to exploit high temporal correlation between successive video frames, which is not much explored in the existing literature. The proposed encoding strategy rectifies the occlusion problems and, at the same time, benefits from a significant reduction in the transmission cost. We also propose motion estimation technique in the compressed domain at the decoder so as to obtain high-quality motion predicted frames. We propose measurement estimation technique that can substantially improve the recovery quality of non-key frames through the efficient use of motion predicted frames as a side information. The major contributions of our work are detailed as follows:

Adaptive Compressive Video Encoding. We propose a novel video encoder based on BCS to reduce transmission cost and overcome occlusion problems. Our proposed method is motivated from background subtraction techniques used for detecting and tracking of moving objects in surveillance applications. Specifically, when the camera is fixed, only moving objects in successive frames occupy different pixel positions due to the unchanging background. This fact opens the door for adaptively allocating the bit-rate for each block independently. For this reason, we encode blocks of the non-key frame in *SKIP*, *LOW* or *HIGH* mode depending on the algebraic difference between each block of the non-key frame and its co-located block in the nearest key frame. Even though the proposed encoding mechanism slightly increases the encoder complexity for computing this difference value, it leads to significant reduction in transmission cost. This could be acceptable in the light that energy consumption for transmission is much more than that for computations in the encoder. Also, our experimental analysis shows that the proposed method is robust to occlusion effects, able to handle large object motions effectively and achieves greater recovery accuracy.

Motion Estimation in Compressed Domain. We develop a methodology to exploit the temporal correlation between successive frames directly from the available CS measurements of key and non-key frames at the decoder. This new technique is motivated from the restricted isometry property (RIP) of the sensing matrices which guarantee to approximately preserve the distance between any two K-sparse signals. The proposed method seeks the best matching measurement vector for each block of the non-key frame from measurement vectors corresponding to blocks of reference key frame within the specified search window.

Measurement Estimation in Compressed Domain. We look forward to effectively utilize the side information obtained using compressed domain motion estimation technique. Instead of transmitting all measurements corresponding to blocks of non-key frames, we propose to transmit only fewer

number of measurements and intentionally skip the remaining measurements. At the decoder, from the given sequence of received CS measurements, a prediction of the non-key frame is obtained using the proposed compressed domain motion estimation technique. Then, by using measurements corresponding to motion predicted frames as a side information, we estimate the skipped measurements using multiple regression model and perform reconstruction using state-of-the-art recovery algorithms for the BCS framework.

The remainder of this paper is organized as follows. We review in Section 2 the basic principles of compressive sensing. In Section 3, we present the proposed adaptive video CS system that allocate different compression ratio for blocks of the non-key frame at the encoder and significantly improve the recovery performance at the decoder through motion estimation in compressed domain followed by measurements estimation. The effectiveness of the proposed system through extensive simulation studies is demonstrated in Section 4 and conclusions are drawn in Section 5.

2 CS Background [9], [19]

In this section, we present mathematical preliminaries of compressive sensing.

CS theory hinges on a key concept of the signal sparsity. The signal $x \in \Re^{N\times 1}$ is said to be K-sparse in transform domain Ψ, if at most K of its transform coefficients, $\theta = \Psi x$ are non-zero. In practice, natural signals such as images are typically known to be sparse in discrete cosine transform (DCT) or discrete wavelet transform (DWT). In CS theory, the sparse signal is sampled and compressed at the same time through a random projection on non-adaptive and signal-independent sensing matrix Φ. More specifically, CS theory demonstrates that the recovery of the signal x is possible from only $M = \mathcal{O}(KlogN)$ linear measurements $y = \Phi x$ if signal exhibits sparsity in some transform domain. The term M/N is referred as compression ratio.

The sensing matrix needs to satisfy restricted isometry property (RIP) and incoherence property for faithful recovery of the signal. In practice, random Gaussian matrix, random Bernoulli matrix and structurally random matrix are known to satisfy above-mentioned properties with overwhelming probability. However, an infinite number of solutions exist for the system of linear equations since $M \ll N$. Hence, the unique solution can be obtained by imposing additional constraints. The solution with the sparsest representation is certainly an appealing representation which can be solved as,

$$\min_{\theta} \| \theta \|_0 \quad s.t. \quad y = \Phi x = \Phi\Psi^{-1}\theta, \tag{P0}$$

where $\| . \|_0$ is the $\| l \|_0$ norm.

However, exact computation of sparsest representation is known to be an NP-hard problem. Thus, many recovery algorithms have been recently proposed that approximates the solution instead by minimizing the l_1 optimization problem given as,

$$\min_{\theta} \| \theta \|_1 \quad s.t. \quad y = \Phi x = \Phi\Psi^{-1}\theta, \tag{P1}$$

where $\| . \|_1$ is the $\| l \|_1$ norm.

This equivalence problem is easier to solve. Various methods such as orthogonal matching pursuit (OMP), focal underdetermined system solver (FOCUSS), model guided adaptive recovery (MARX), gradient projection sparse reconstruction (GPSR) have been proved to reconstruct the signal exactly and efficiently with high probability.

3 Proposed Adaptive Compressed Video System

In this section, we present the overall architecture of the compressive sensing based video acquisition-reconstruction system. We describe in detail the motivation and design of the proposed adaptive compressive video CS encoder. We also discuss the novel methodology for recovery of CS videos including motion and measurements estimation in compressive domain.

3.1 System Architecture

In our model, video acquisition is performed at the encoder through the BCS framework [20] wherein each frame is divided into number of non-overlapping blocks and projected independently on respective sensing matrices. Let $\mathbf{x}_i$ be the vector representation of ith $B \times B$ block of the image X after raster scanning with $N_B = B^2$ pixels. Then, the block can be sampled using a structurally random matrix (SRM) [21] as $y_i = \Phi_B \mathbf{x}_i$, where y_i represents a measurement vector and Φ_B is the $M_B \times N_B$ $(M_B \ll N_B)$ sensing matrix.

The proposed adaptive CS encoder discussed in the following can be used for video acquisition in CCD/CMOS cameras as well as single-pixel camera [22]. The video sequence consists of a group of picture (GOP) with $IPPP...$ $PIPPP...$ frame pattern. Intra-frames (I frames) are encoded at a higher bit-rate, independent of its neighboring frames and used as a reference frame during decoding. Predictive frames (P frames) are encoded using the proposed adaptive strategy which takes advantage of high redundancy between successive frames for effective compression. While I frames are recovered at the decoder using conventional BCS recovery algorithms, P frames are recovered using proposed compressed domain motion and measurements estimation techniques, as described in more detail in Sections 3.3 and 3.4 respectively. For convenience, we refer I and P frames as key and non-key frames respectively in the rest of the discussion.

3.2 Proposed Adaptive Compressed Video Encoding

There has been a number of efforts investigated in recent literature [10], [13], [16] aimed at reducing the computational complexity of the encoder and improving the recovery accuracy at the decoder. The reduction in the encoder complexity is achieved by using structured matrices rather than complete random matrices for sensing. Besides, CS encoder samples each frame independently to avoid any additional hardware cost or complexity without looking at the high correlation between successive frames. Thus, since such high inter-frame correlation is exploited only at the decoder, it suffers heavy losses in the recovery performance in the presence of occlusion effects. This independent frame encoding also leads to increase in transmission cost for a

given reconstruction quality. Hence, there is still great scope to reduce transmission cost at the encoder and at the same time, overcome occlusion effects at the decoder which is overlooked in the existing literature.

While existing video CS encoder can sense the signal more parsimoniously, they do not exploit the high inter-frame redundancy. But, bridging the gap between encoder hardware complexity and the transmission cost can efficiently address the problems associated with efficient reconstruction. In the following, we consider the design of smart and adaptive video CS encoder. Our design is motivated by the success of frame redundancy exploitation at the encoder of conventional video coding standard. The realization of the proposed encoder can lead to significant reduction in transmission cost, offers a greater flexibility in GOP structure and simultaneously get rid of occlusion effects with only a little increase in hardware complexity compared to a conventional video CS encoder.

The reduction in transmission cost can be attained by transmitting only the minimum number of measurements for each block that guarantees faithful recovery. One way to accomplish this challenge is to allocate different compression ratio for various blocks of the non-key frame depending on inter-block redundancy. For this reason, we compute mean absolute difference (MAD) d between blocks of the non-key frame and its co-located block in the nearest key frame. Then, each block is encoded in any one of *SKIP*, *LOW* or *HIGH* modes as described in the following.

- *SKIP mode:* If d is less than predefined lower threshold, it indicates that corresponding block of the non-key frame is highly correlated to its counterpart in the nearest key frame. In this case, the encoder does not process the block and transmits a bit indicating SKIP mode.
- *LOW mode:* If d is in between a pre-specified lower threshold and an upper threshold, it is an indication of a medium level of correlation between respective blocks. Therefore, the encoder transmits only M_T measurements among the total M_B measurements.
- *HIGH mode:* If d is greater than upper threshold due to large object motion or occlusion effect, the encoder transmits all M_B measurements.

The proposed adaptive encoder is particularly beneficial for the still camera in which the difference between successive frames is only due to moving objects within field of view (FOV). Because of this unchanging background, the total number of blocks in *SKIP* mode is large and very few blocks need to be processed in *HIGH* mode. In this way, the proposed encoding design achieves a significant reduction in transmission cost.

The recovery of the non-key frame is performed at the decoder through following steps:

- *SKIP mode:* Copy measurements corresponding to co-located blocks in the nearest key frame
- *LOW mode:* Apply compressed domain motion and measurements estimation techniques as discussed in Sections 3.3 and 3.4.
- *HIGH mode:* Use received measurements without any processing

After gathering measurements for all blocks of the non-key frame, the final recovery is achieved using state-of-the-art BCS recovery algorithms. It is worth noting that measurements corresponding to blocks in *HIGH* mode are utilized without any further modifications. These blocks are responsible for occlusion effect with high probability. Since these blocks are processed in *HIGH* mode, the recovered non-key frame can be guaranteed to be occlusion-free and can exhibit significant improvement in the visual quality.

3.3 Proposed Compressed Domain Motion Estimation and Prediction at the Decoder

In conventional video coding, it has been witnessed that the signal prediction plays a key role in exploiting temporal redundancy between successive frames [23], [24], thereby improving signal compressibility and reducing the transmission cost. The successive frames in a video sequence are very similar. So, instead of encoding each frame independently, many video coding standards employ ME and MC techniques at the encoder to compute motion predicted frame. The significant compression is achieved by encoding the only compressible difference between the reference frame and motion predicted frame. In essence, quality of frame prediction reflects into signal compressibility and constitutes basic building block of current video coding standards. But ME and MC at the encoder make the structural complexity of the encoder very high.

ME and MC techniques are widely adopted for video CS framework [16], [17] to exploit high degree of temporal correlation among successive video frames at the decoder. However, the existing approaches require independent CS recovery prior to the prediction of the non-key frame. But, since non-key frames are sampled at much lower bit-rate, the initial recovery of non-key frame suffers from high-frequency oscillatory artifacts. This frame after initial recovery can be considered as a very crude approximation of the original non-key frame which causes degradation in the quality of subsequent ME and MC stages dramatically. To overcome aforementioned problems, we propose to obtain motion predicted frame directly from available CS measurements of non-key frame, instead of recovering each non-key frame independently.

This new compressed domain motion estimation technique stems from the following observation. It is well known in CS theory that the sensing matrix Φ obeys the *restricted isometry property* which states that there exists a smallest number δ_K such that [9], [19],

$$(1-\delta_K) \parallel x \parallel_2^2 \;\leq\; \parallel \Phi x \parallel_2^2 \;\leq\; (1+\delta_K) \parallel x \parallel_2^2, \tag{1}$$

holds for any K-sparse signal x. The $\delta_K \in (0,1)$ is restricted isometry constant of the matrix Φ. Various results demonstrate that if $\delta_K < \sqrt{2}-1$, the recovery of the signal is exact.

The RIP of random matrix illustrates that the distance between two $K-$ sparse vectors will change only very little with high probability after random projection. It is worth noting that the distance between any two $K-$sparse vectors can be at most $2K-$sparse. In order to recover $K-$sparse signal x from corresponding CS measurements y uniquely, any difference between two $K-$sparse signals, $x_k - x_j$ is approximately preserved in the measurements y_k and y_j as,

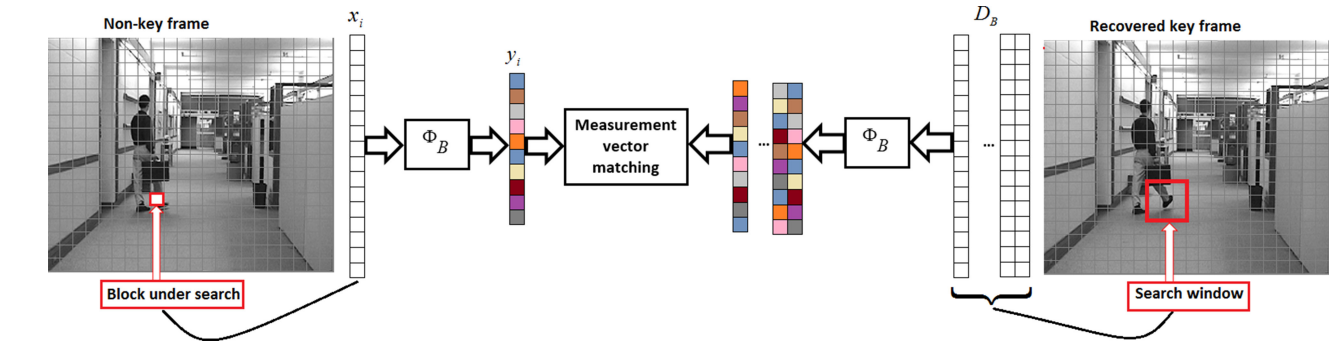


Fig. 1. Motion estimation and prediction in compressed domain.

$$(1-\delta_{2K}) \parallel x_k - x_j \parallel_2^2 \leq \parallel y_k - y_j \parallel_2^2 = \parallel \Phi(x_k - x_j) \parallel_2^2 \leq (1+\delta_{2K}) \parallel x_k - x_j \parallel_2^2, \quad (2)$$

when Φ is RIP of order $2K$.

The RIP implies that all pairwise distance between $K-$sparse signals must be well preserved in measurement space and guarantees that a transformation varies little with respect to distance. In this context, we extend the advantage of distance preserving property of the sensing matrix for accurate frame prediction from the only far fewer number of CS measurements. To be more specific, we model motion estimation as measurement vector matching problem as opposed to block matching in the spatial domain.

The compressed domain motion estimation model depicted in Fig. 1 assumes that measurement vector in CS retains nearly all the useful information of the image block. The motion estimation technique in conventional video coding compares each block of the non-key frame with a co-located block and its adjacent neighboring blocks decided by search window in a nearby key frame. Different from this, we compare measurement vector of each block of the non-key frame with measurement vectors corresponding to blocks inside search window of the nearest (preceding or following) recovered key frame at the decoder. It is assumed that the key frame is recovered at the decoder prior to the prediction of the non-key frame.

The proposed comparison of measurement vectors is similar to exhaustive block matching search in conventional ME techniques. It is worth noting that exhaustive search strategy gives accurate prediction in comparison with other search strategies, at the expense of high computational cost. Since the proposed measurements comparison is performed at the decoder and it is aimed exclusively to obtain accurate prediction, the exhaustive search strategy is employed in the proposed method. Our proposed compressed domain motion estimation and prediction method consists of two stages, namely *initial* stage and *prediction* stage.

In the *initial* stage, coordinates (u_i, v_i), of the center of current block of the non-key frame is obtained from index i, of the received CS measurements y_{NK}^i. The subscript NK denotes the non-key frame. Next, all overlapping $B \times B$ blocks inside the search area of diameter $2s$ in the nearest recovered key frame $\hat{I}(u_i - s : u_i + s, v_i - s : v_i + s)$, are stacked as columns of matrix D_B after raster scanning. These column vectors of D_B are projected onto the sensing matrix Φ_B to generate candidate measurement vectors, which will be further used for motion estimation.

In the *prediction* stage, we seek the measurement vector giving the best matching to y_{NK}^i, among candidate measurement vectors corresponding to blocks in a search area of the recovered key frame. Then, motion predicted block is taken as the one in the recovered key frame that provides the best matching measurement vector. The schematic of the proposed compressed domain motion estimation and prediction method is shown in Fig. 2.

We illustrate in Algorithm 1 the process of the proposed compressed domain motion estimation and prediction technique and refer it as C-MEMP algorithm. The proposed method is optimal in the sense that it mimics state-of-the-art motion estimation techniques for the compressed domain while keeping the complexity of the encoder at an optimal low level. Furthermore, the proposed method is more powerful than the existing spatial domain block matching algorithms in video CS, as it enables block prediction from corresponding CS measurements rather than low-resolution frame approximation. In essence, the proposed design augments the conventional ME technique with the CS theory, aiming to retain the best features of both theories.

Algorithm 1. Compressed Domain Motion Estimation and Prediction (C-MEMP)

Input: measurement vector y_{NK}^i of ith block, the corresponding sensing matrix Φ_B, coordinates (u_i, v_i) of center of the current block, search diameter $2s$ and recovered nearest key frame $\hat{I}$
Output: $\mathbf{x}_p^i$ (prediction of the block $\mathbf{x}_{NK}^i$)
Procedure:
1 **Candidate correlated blocks:** Rasterize all overlapping $B \times B$ blocks of $\hat{I}(u_i - s : u_i + s, v_i - s : v_i + s)$ and stack resultant column vectors in a matrix D_B
2 **Random projection:** $y_{ref} = \Phi_B D_B$
3 **Motion estimation:** Compare each vector of y_{ref} with y_{NK}^i
4 **Motion prediction:** Select the column vector of D_B corresponding to best matching measurement vector of y_{ref} with y_{NK}^i in mean absolute difference sense
5 **Unrasterize** the selected vector to obtain $\mathbf{x}_p^i$

It is worth noting that the existing video CS algorithms utilize frame prediction only as side information (SI) to further refine the quality of non-key frames. Motivated by this fact, we propose in the following a novel measurement estimation technique that reduces, to a certain extent, the transmission cost and, at the same time, further improves the quality of motion predicted frames.

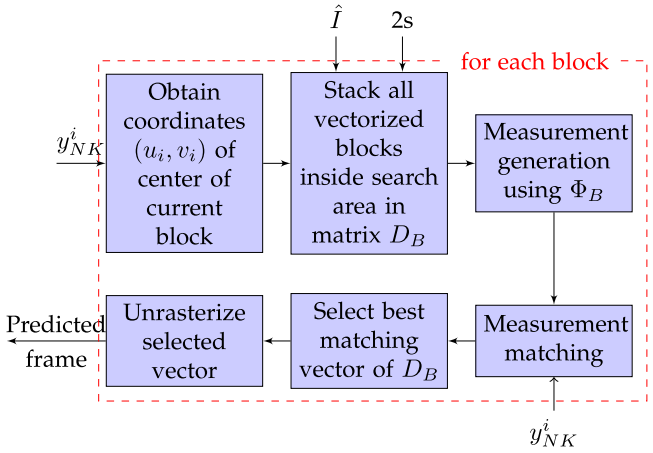


Fig. 2. Motion estimation and prediction in compressed domain at the decoder from measurements of non-key frame y^i_{NK}, nearest recovered key frame $\hat{I}$ and search diameter $2s$.

3.4 Proposed Measurements Estimation Algorithm

Distributed compressive video coding (DCVC) is an extension of distributed source coding (DSC) [25] for video compression targeting a low encoding complexity and error resilience. In DSC, Slepian and Wolf proved that the source X can be coded at a rate close to conditional entropy $H(X/Y)$, if the correlated source Y is transmitted at full rate $H(Y)$ and used as a side information (SI) to decode X. This, in principle, is the success behind DCVC where motion predicted frames constructed only at the decoder are used as SI for recovery of the non-key frame. In fact, the quality of SI allows controlling the rate of non-key frame in a more accurate manner.

Variety of video CS decoding approaches are investigated in recent literature that perform SI driven recovery of non-key frame [16], [17]. Many of the prior works incorporate SI either to model joint optimization problem or to drive residual based CS recovery resulting from frame difference sparsity. Unfortunately, these approaches hold promising when sufficient number of measurements are available at the decoder. However, through experimental analysis, we noticed that our proposed prediction method is able to construct high-quality SI from a far fewer number of CS measurements. But, existing video CS recovery algorithms coupled with SI may not be directly adopted due to the availability of only fewer number of CS measurements.

It has been known that motion predicted frame is highly correlated to the respective non-key frame. At this point, a natural question arise about the correlation between measurements resulting from block-based sampling of motion predicted frame using Φ_B with that of non-key frame. After extensive research, we noticed that CS measurements inherit high correlation between respective frames, thanks to the RIP property of the sensing matrix. Based on this hidden correlation that exists between CS measurements of the non-key frame and corresponding motion predicted frame (SI), we propose measurements estimation technique for improvement in the recovery of non-key frame.

The core idea of the proposed method is to reduce the number of CS measurements transmitted to the decoder as much as possible, thereby improving energy efficiency of the encoder. The motivation behind the proposed method is

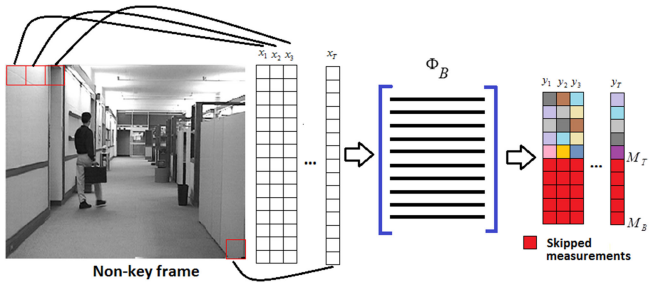


Fig. 3. Measurements skipping process at the encoder.

twofold: (1) missing data estimation techniques in wireless sensor networks (WSNs) and (2) matrix completion theory. These two methods are discussed in the following.

- In WSNs, the missing of sensor data (such as temperature or humidity) is a common and inevitable problem due to inherent characteristics of wireless channel [26]. Various works reported in the literature to solve this problem are based on the principle that data captured by sensors which are spatially correlated in a WSNs scenario exhibit high correlation. Multiple regression (MR) model or graph theory based approaches are adopted to exploit spatial correlation between data from neighboring sensor nodes to make accurate estimation of the missed or lost data [26], [27].
- In recent years, matrix completion approaches that recover missing entries of the partially filled matrix, are successfully applied to many practical problems [28]. The matrix completion problems are based on the assumption that the matrix to be recovered is low-rank or approximately low-rank, i.e., a maximum number of linearly independent row or column vectors are small. As a motivating example, image/video inpainting techniques exploit spatio-temporal correlation using matrix completion to recover missing parts of images or video frames [29].

The proposed method is a special case of missing data estimation wherein we exploit high correlation among CS measurements for the estimation of skipped measurements. Besides, the proposed method can be viewed as a low-rank matrix completion through a concatenation of CS measurements of non-key frames and the corresponding motion predicted frame in a matrix. The resulting matrix of CS measurements is low-rank due to high correlation among them, validating applicability of the proposed measurement estimation techniques.

In the proposed approach, we sample non-key frame at a higher bit-rate similar to key frame at the encoder. However, instead of transmitting all measurements corresponding to blocks of non-key frames, we intentionally ignore some of the measurements and transmit the remaining measurements as shown in Fig. 3. Specifically, if $\{y^i_{NK}\}_{1:M_B} = \{y^{i,1}_{NK}, y^{i,2}_{NK}, \ldots, y^{i,M_B}_{NK}\}$ are measurements of a block, $\mathbf{x}^i_{NK}$, then we transmit only $\{y^i_{NK}\}_{1:M_T}$ with $M_T \ll M_B$ measurements to the decoder and intentionally skip remaining $M_B - M_T$ measurements. This significantly reduces the number of measurements transmitted to the decoder and as a consequence, reduces transmission cost.

At the decoder, we propose a three stage recovery process, namely *motion prediction*, *measurement estimation* and *refinement* stage. In the first *motion prediction* stage, given the sequence of received CS measurements $\{y^i_{NK}\}_{1:M_T}$, a

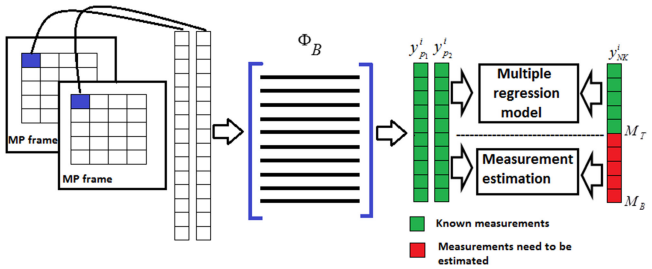


Fig. 4. Measurements estimation process at the decoder.

prediction of non-key frame from preceding and following key frames is created using our proposed C-MEMP algorithm; i.e.,

$$X_p = \text{C-MEMP}(\hat{I},\ \{Y_{NK}\}_{1:M_T}), \tag{3}$$

where $\{Y_{NK}\}_{1:M_T} = (\{y^1_{NK}\}_{1:M_T}, \ldots, \{y^T_{NK}\}_{1:M_T})$ is the matrix comprising of measurement vectors corresponding to all blocks of the non-key frame and T denotes the total number of blocks. We implicitly assume that the recovered key frame $\hat{I}$ is available prior to prediction of the non-key frame.

In the *measurement estimation* stage, we make use of high correlation between received measurements $\{y^i_{NK}\}_{1:M_T}$ and that $\{y^i_p\}$, resulting from sampling of co-located blocks of motion predicted frames using Φ_B (see Fig. 4). We integrate the most popular multiple regression algorithm to estimate the missing measurements as detailed in section 3.5. The MR algorithm first computes regression coefficients $\hat{\beta}$ from received M_T measurements of the non-key frame and those of the corresponding motion predicted frame as,

$$\hat{\beta} = \text{MR}(\{y^i_p\}_{1:M_T},\ \ \{y^i_{NK}\}_{1:M_T}). \tag{4}$$

Then, using computed regression coefficients and known $M_B - M_T$ measurements of motion predicted frame, the skipped measurement of non-key frame are estimated as,

$$\{y^i_{NK}\}_{M_T+1:M_B} = \text{MR}(\hat{\beta},\ \{y^i_p\}_{M_T+1:M_B}). \tag{5}$$

In the *refinement* stage, the final recovery of the non-key frame is performed using both received M_T and estimated $M_B - M_T$ measurements with state-of-the-art BCS recovery algorithm. The proposed three-stage motion and measurements estimation model is schematically described in Fig. 5.

In spirit and functionality, measurements resulting from motion predicted frames is similar to side information in the distributed video coding. As in the DVC, the proposed method exploits the correlation between measurements only at the decoder and retain all the basic advantages of CS video coding: reduced compression ratio and transmission cost, low encoding complexity, while greatly improving the recovery performance of the non-key frame. In the following, we encourage the use of multiple regression model for estimation of missing measurements. This is inspired by the fact that the estimated measurements are the best possible estimation in the sense that it is unbiased (equal to the original measurements on average).

3.5 Multiple Regression Model [26]

The application of MR algorithm in WSNs for missing data estimation has been thoroughly investigated in the literature.

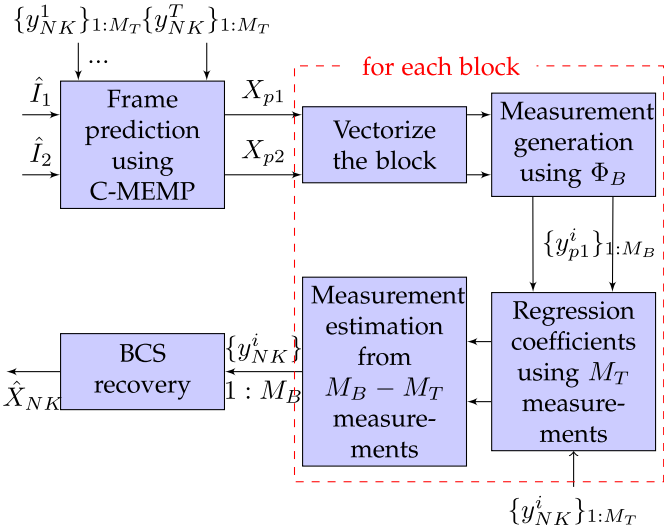


Fig. 5. Video CS recovery system using motion and measurements estimation in compressed domain.

The major part of the research is concentrated on missing data estimation (such as temperature or humidity) of sensor nodes by utilizing the spatial correlation between neighboring sensor nodes. Different from this approach, we insist the use of MR algorithm for missing measurements estimation rather than missing data estimation.

Assume that we need to estimate the missing measurements of ith block of non-key frame, $\mathbf{x}^i_{NK}$, at an instance k from available measurements of its co-located blocks in corresponding motion predicted frames $\mathbf{x}^i_{p1}$ and $\mathbf{x}^i_{p2}$. Let $y^{i,k}_{NK}$ denotes measurement corresponding to ith block of non-key frame at an instance k. Using MR model, the task is to estimate missing measurement, $y^{i,k}_{NK}$, as a linear combination of known measurements $y^{i,k}_{p1}$ and $y^{i,k}_{p2}$ from motion predicted frames. It is given as,

$$y^{i,k}_{NK} = \beta_0 + \beta_1 y^{i,k}_{p1} + \beta_2 y^{i,k}_{p2} + \epsilon_k, \tag{6}$$

where, ϵ_k is the noise variable at an instance k with mean zero and standard deviation σ. The unknown coefficients β_0, β_1 and β_2 are the regression coefficients that need to be estimated from available measurements.

In the terminology of MR model, $y^{i,k}_{NK}$ is referred as response variable (or dependent variable) and $y^{i,k}_{p1}$, $y^{i,k}_{p2}$ are referred as predictor variables. The available measurements consist of M_T rows of observations given as $y^{i,j}_{NK}$, $y^{i,j}_{p1}$ and $y^{i,j}_{p2}$; $j = 1, 2, \ldots, M_T$. The estimation of the β coefficients is computed so as to minimize the sum of squared differences between observed measurements and predicted measurements. The sum of squared differences is given as,

$$\sum_{j=1}^{M_T} (y^{i,j}_{NK} - \beta_0 - \beta_1 y^{i,j}_{p1} - \beta_2 y^{i,j}_{p2})^2. \tag{7}$$

The coefficients that minimize Eq. (6) is denoted as $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$. After estimating these coefficients, they are plugged into the linear regression model. The predicted measurement is given as,

$$\hat{y}^{i,k}_{NK} = \hat{\beta}_0 + \hat{\beta}_1 y^{i,k}_{p1} + \hat{\beta}_2 y^{i,k}_{p2}, \quad k = M_T + 1, \ldots, M_B. \tag{8}$$

If the known measurements corresponding to non-key frame is denoted as $\mathbf{y}_{NK}^{i,known} = (y_{NK}^{i,1}, \ldots, y_{NK}^{i,M_T})^T$ and the measurements corresponding to motion predicted frames is denoted as,

$$\mathbf{\Gamma} = \begin{bmatrix} 1 & y_{p1}^{i,1} & y_{p2}^{i,1} \\ 1 & y_{p1}^{i,2} & y_{p2}^{i,2} \\ \vdots & \vdots & \vdots \\ 1 & y_{p1}^{i,M_T} & y_{p2}^{i,M_T} \end{bmatrix},$$

then,

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T = (\mathbf{\Gamma}^T\mathbf{\Gamma})^{-1}(\mathbf{\Gamma}^T\mathbf{y}_{NK}^{i,known}). \quad (9)$$

As reported in the literature of MR model, the estimated coefficients, $\hat{\beta}$, are minimal variance, the unbiased estimated value of β coefficients. We detail in Algorithm 2 the recovery of the non-key frame at the decoder using proposed three-stage motion prediction and measurements estimation technique. We term this algorithm as the C-MEst-BCS algorithm.

Algorithm 2. Recovery of Non-Key Frame using Motion and Measurements Estimation (C-MEst-BCS)

Input: measurement vector $\{y_{NK}^i\}_{1:M_T}$, the corresponding sensing matrix Φ_B, recovered preceding and following key frames $\hat{I}_1$ and $\hat{I}_2$ respectively

Output: $\hat{X}_{NK}$ (recovered non-key frame)

Procedure:

1 **Motion predicted frames:** Obtain X_{p1} and X_{p2} from preceding and following key frames using C-MEMP
2 **Correlated measurements:** Sample $\mathbf{x}_{p1}^i$ and $\mathbf{x}_{p2}^i$ using Φ_B to obtain $\{y_{p1}^i\}_{1:M_B}$ and $\{y_{p2}^i\}_{1:M_B}$.
3 **Regression coefficients:** Obtain $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ using Eq. (8)
4 **Estimate** missing measurements using Eq. (7)
5 **Repeat** steps $2-4$ for each block of the non-key frame
6 **Recover** non-key frame using both received and estimated measurements

4 Experimental Results

We evaluate the performance of the proposed video CS system on a variety of video sequences and compare them with current leading video CS methods. The MATLAB code of the proposed techniques reproducing the experimental results is available at https://amitsunde.github.io/.

4.1 Experimental Settings

We test the performance and validity of the proposed solutions on a set of diverse video sequences with grayscale CIF resolutions (288×352 pixels/frame) at a frame rate of 25 frames/second. The GOP of size 4 frames is considered for all sequences. In our simulations, each video frame is divided into non-overlapping blocks of size 16×16 for BCS sampling. The key frames are encoded at the rate of 4 bpp and used as reference frames in the recovery of non-key frames. We adopt structurally random matrix proposed in [21] due to its low structural and computational complexity.

The CS measurements are then quantized using uniform scalar quantization (SQ) followed by fixed length coding. The CS simultaneously performs sampling and compression; the bit-rate of CS coder [30] is expressed as,

$$\text{bit-rate} = \frac{M}{N}R \quad \text{bpp}, \quad (10)$$

where R represents the rate of quantizer.

It is worth noting from Eq. (10) that the compression ratio and R cannot be minimized simultaneously. Hence, the maximum reconstruction quality can be attained through an optimal combination of the number of CS measurements and quantization levels. Due to the absence of closed-form solution to optimize the bit-rate, a good estimate on quantization precision can be obtained through statistical distribution of the CS measurements. The SRM has got the property that it retains Laplacian distribution of the measurements irrespective of the distribution of original input data [30]. So motivated from rate-distortion analysis in [31], we observed empirically that quantization precision with $Q = 4$ bits per CS measurements obtained using SRM works well for a given bit budget over a wide range of video sequences. Hence, we quantize CS measurements using SQ together with fixed length coding with Q = 4 bits per CS measurements. The quantization step size, Δ, is given as,

$$\Delta = \frac{y_{max} - y_{min}}{2^Q}, \quad (11)$$

where y_{max} and y_{min} represent a maximum and minimum value of CS measurements of each frame.

The frame recovery is performed at the decoder through BCS-FOCUSS [20] algorithm using discrete cosine transform (DCT) as a sparsifying transform. The choice of BCS-FOCUSS algorithm is motivated by the fact that it gives the best performance in the BCS framework. The quality of the recovered frame is evaluated in terms of peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [32].

4.2 Comparison with Other Methods

The proposed video CS system is compared with the following prominent video CS methods.

- Motion compensated block compressive sensing (MC-BCS)[16].[1] The MC-BCS is block-based video CS coder wherein the independent CS recovery of the non-key frame is performed initially from corresponding CS measurements. Then, ME is performed on the initial recovered non-key frame and respective key frame at the decoder to estimate a motion field. Using these motion vectors, a motion compensated (MC) frame that forms a prediction of the non-key frame is obtained. The motion predicted frames in MC-BCS are referred to as MEMC in the simulation studies. Furthermore, the residual-based recovery process is performed to refine the quality of motion predicted frames.
- Multihypothesis reweighted residual sparsity (MH-RRS) [17]. The MH-RRS is block-based CS video coder that uses more sophisticated multihypothesis (MH) technique for motion estimation. In this method, the prediction of

1. The source of MC-BCS can be downloaded from http://my.ece.msstate.edu/faculty/fowler/BCSSPL/

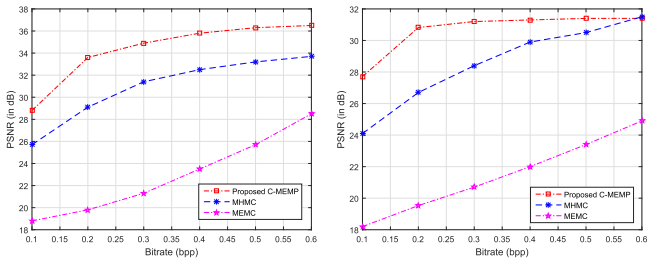


Fig. 6. Performance analysis of the proposed C-MEMP technique in comparison with MEMC and MHMC methods for (a) *Hall-monitor* and (b) *Container* video frames.

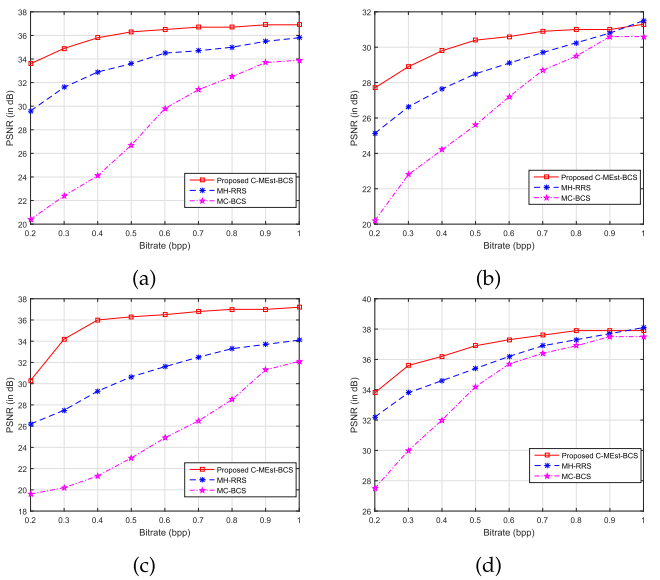


Fig. 7. Performance analysis (in terms of PSNR) of the proposed C-MEst-BCS method in comparison with MC-BCS and MH-RRS methods for various video frames (a) *Hall-monitor*, (b) *Traffic*, (c) *Container*, (d) *Crew*.

each block of the non-key frame is obtained through an effective linear combination of all blocks inside the specified search window in the recovered key frames. The motion predicted frames in MH-RRS are referred to as MHMC in the simulation studies. Furthermore, the quality of motion predicted frames is enhanced through re-weighted residual-based recovery strategy. The different residual coefficients are weighted according to their probability of being zero.

4.3 Performance Evaluation of the Proposed Compressed Domain Motion Estimation Technique

The CS video coder usually aims to obtain accurate predictions of the non-key frame at the decoder. More specifically, better the frame prediction, superior will be the reconstruction performance. Therefore, we compare in Fig. 6 the performance of the proposed motion estimation and prediction (C-MEMP) method with MEMC and MHMC methods to get a closer comparison of baseline motion prediction approaches. The proposed C-MEMP, MEMC, and MHMC are interim stages of the proposed adaptive video coder, MC-BCS and MH-RRS schemes respectively.

It can be clearly seen in Fig. 6 that the proposed C-MEMP method performs consistently better than both MEMC and

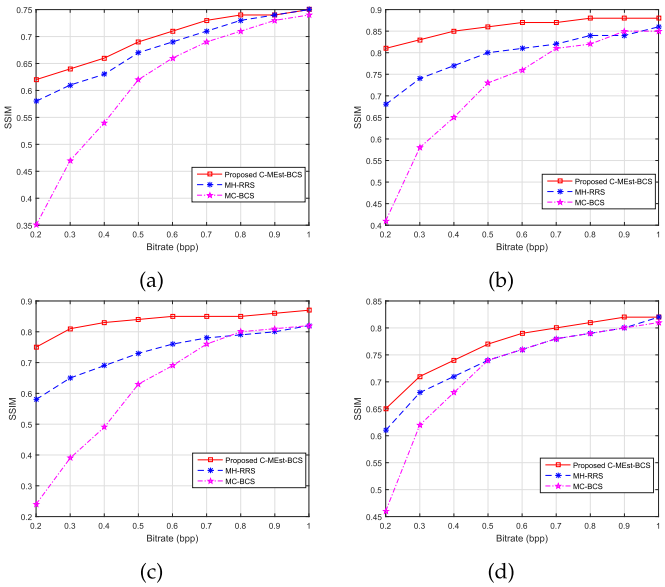


Fig. 8. Performance analysis (in terms of SSIM) of the proposed C-MEst-BCS method in comparison with MC-BCS and MH-RRS methods for various video frames (a) *Hall-monitor*, (b) *Traffic*, (c) *Container*, (d) *Crew*.

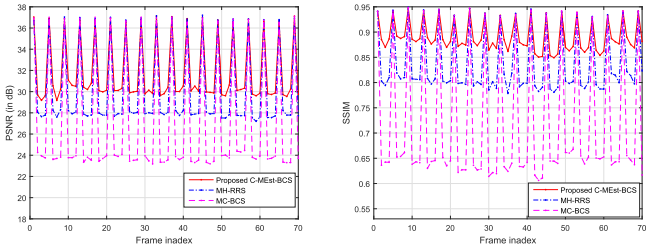
MHMC techniques. More importantly, the proposed method is able to achieve the better frame prediction at extremely lower bit-rates. This is due to the fact that the proposed C-MEMP enables motion estimation directly from lossless CS measurements. Nevertheless, it is noticed through extensive simulations studies that $M_T = 25$ CS measurements of blocks of the non-key frame are sufficient to find the best matching measurement vector. However, an increase in the number of CS measurements above specific M_T does not much further improve the quality of frame prediction. Therefore, the proposed method exhibits much superior performance, particularly at lower bit-rates. This vividly emphasizes the importance of the proposed method since non-key frames are sampled at a much lower bit-rate in CS video coding.

4.4 Performance Evaluation of the Proposed Measurement Estimation Methods

We evaluate the effectiveness of the proposed measurement estimation (C-MEst-BCS) technique on four video sequences[2] namely, *Hall-monitor*, *Traffic*, *Container* and *Crew*. All videos contain one or multiple moving objects with slow, moderate or fast motion.

We plot in Fig. 7 the quality of reconstructed frames in terms of PSNR for different video sequences. It can be noticed from Fig. 7 that the proposed C-MEst-BCS method provides a significant increase in the recovery quality as opposed to both the MC-BCS and MH-RRS methods, especially at a lower bit-rate. This clearly indicates that the superior performance of the proposed C-MEMP technique inherits into the C-MEst-BCS method. The much better recovery performance at lower bit-rates signifies the importance of the proposed method in reducing the transmission cost. Fig. 8 presents the quality of recovery accuracy in

2. The videos used in experiments can be downloaded from https://media.xiph.org/video/derf/, http://dyntex.univ-lr.fr/index.html, https://www.microsoft.com/en-us/download/details.aspx?id=52358



(a) (b)

Fig. 9. (a) PSNR and (b) SSIM comparison of various algorithms for the first 70 frames of *Traffic* dataset. The bit-rate of non-key frames is set to 0.4 bpp.

terms of SSIM. The results shown in this figure follow an identical pattern to that of PSNR results shown in Fig. 7, and confirms the effectiveness of the proposed C-MEst-BCS method.

We show in Fig. 9 the quality in PSNR and SSIM for first 70 frames of *Traffic* video. From the figure, the proposed C-MEst-BCS method completely outperforms other approaches by the significant margin. We also noticed similar kind of observations for the other video sequences, but those results are not shown here due to the space limitation. We display in Figs. 10 and 11 the reconstructed frames of *Traffic* and *Container* videos for visual comparison. It can be witnessed that the proposed C-MEst-BCS method preserve the richness or details in the foreground as well as background regions and it is much superior to other methods.

4.5 Performance Evaluation of the Proposed Adaptive Video CS Encoder

In this section, we evaluate the rate-distortion performance of the proposed adaptive video encoder and demonstrate its efficiency in overcoming the occlusion problems. In addition, the computational cost and transmission overhead analysis using the proposed adaptive scheme is presented. The simulation studies are performed on the *Break-dancer* and *Ballet* videos with two challenging tasks: high compression efficiency and complicated object motion. The encoder transmits $M_T = 23$ and $M_B = 200$ measurements for blocks in *LOW* mode and *HIGH* mode respectively. The lower and upper threshold values are tuned to balance the required bit-rate and recovery quality of non-key frames.

4.5.1 Rate-Distortion Analysis

We illustrate in Fig. 12 the performance of the proposed adaptive video CS system in terms of PSNR as a function of the bit-rate. We can clearly see that the proposed scheme outperforms non-adaptive approaches for all values of the bit-rate. The proposed method offers as much as $8-10$ dB and $2-6$ dB improvement in PSNR over the MC-BCS and MH-RRS schemes respectively. The reason is that the proposed method incorporates the powerful C-MEst algorithm for blocks in *LOW* mode and independently decodes low temporally correlated blocks in the *HIGH* mode which makes the performance more stronger.

We show in Fig. 13 SSIM curves for an in-depth comparison of recovery accuracy of various methods at different values of the bit-rate. It is apparent from the figure that the proposed method results in higher SSIM than non-adaptive methods for each of the bit-rate. As seen in Fig. 13, the proposed method demonstrates much better performance at a sufficiently low bit-rate. This is quite remarkable since this permits the encoder to effectively utilize the available power and bandwidth in severely resource-constrained conditions.

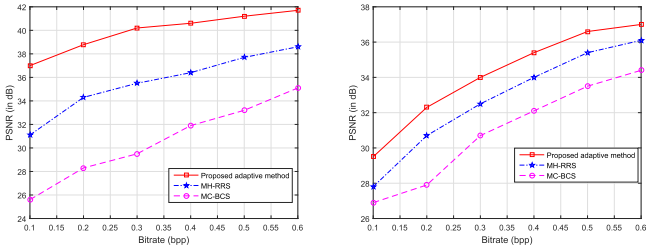
For a more comprehensive comparison, we present in Fig. 14 PSNR and SSIM values for first 80 frames of *Ballet* video. Clearly, the proposed method consistently achieves a significant coding gain over MC-BCS and MH-RRS techniques for a given bit-rate. Encouragingly, the proposed video CS system suggests a promising direction to remove redundancy between successive frames for substantially improving coding performance.

(a) (b) (c) (d)

Fig. 10. Comparison of various methods on *Traffic* frame at 0.4 bpp (a) Ground truth, (b) Proposed C-MEst-BCS (PSNR = 28.4 dB, SSIM = 0.84), (c) MH-RRS (PSNR = 26.9 dB, SSIM = 0.79), (d) MC-BCS (PSNR = 23.6 dB, SSIM = 0.64).

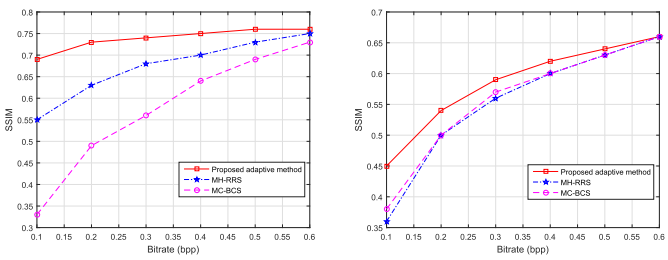
(a) (b) (c) (d)

Fig. 11. Comparison of various methods on *Container* frame at 0.4 bpp (a) Ground truth, (b) Proposed C-MEst-BCS (PSNR = 36.7 dB, SSIM = 0.84), (c) MH-RRS (PSNR = 30.1 dB, SSIM = 0.69), (d) MC-BCS (PSNR = 22.1 dB, SSIM = 0.60).



(a) (b)

Fig. 12. Performance analysis (PSNR) of the proposed adaptive method in comparison with MC-BCS and MH-RRS schemes for various video frames (a) *Ballet*, (b) *Break-dancer*.



(a) (b)

Fig. 13. Performance analysis (SSIM) of the proposed adaptive method in comparison with MC-BCS and MH-RRS schemes for various video frames (a) *Ballet*, (b) *Break-dancer*.

Finally, Fig. 15 demonstrates a noticeable improvement in visual quality by the proposed method. The frames recovered with the proposed method appears to be sharp and retains the highest fidelity, particularly in areas of moving objects.

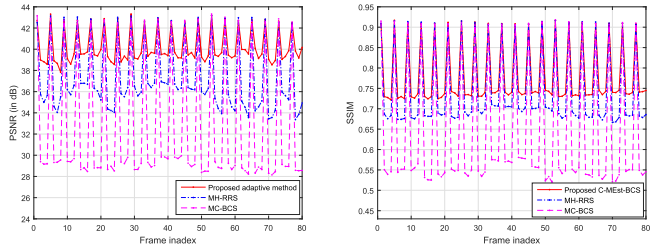
4.5.2 Overcoming Occlusion Problems

We demonstrate the important strength of the proposed adaptive encoding scheme in overcoming the occlusion effects. Given two consecutive frames of a video sequence, occlusion occurs if some objects (parts of the scene) in the non-key frame have not appeared in the corresponding key (reference) frame due to fast object motion or sudden rapid movement. We show in Fig. 16 the original key frames and corresponding non-key frames of *Traffic* and *Ballet* videos wherein occluded regions are marked in a rectangular box.

The performance of the various video CS methods is illustrated in Fig. 16. It can be clearly noticed that the proposed method shows a great edge in overcoming occlusion problems. This is because the proposed method encodes blocks in *SKIP*, *LOW* or *HIGH* mode depending on mean absolute difference d between blocks of the non-key frame and its co-located block in the nearest key frame. The difference d for blocks of the non-key frame corresponding to occlusion effect is usually very high and therefore such blocks are encoded in *HIGH* mode (at a higher bit-rate). Since these blocks are decoded independently (without the help of respective key frames), the proposed method is able to handle occlusion problem in a very efficient manner.

4.5.3 Computational Complexity

We evaluate the computational complexity of the proposed algorithm in terms of the number of additions/subtractions



(a) (b)

Fig. 14. (a) PSNR and (b) SSIM comparison of various algorithms for the first 80 frames of *Ballet* dataset. The bit-rate of non-key frames is set to 0.25 bpp.

and absolute operations required. For this analysis, video frames of size $\sqrt{N} \times \sqrt{N}$ pixels are divided into T non-overlapping $B \times B$ blocks with $N_B = B^2$ pixels for acquisition using BCS.

In the proposed method, the computation of mean absolute difference to decide whether blocks of the non-key frame to be encoded in *SKIP*, *LOW* or *HIGH* mode requires total $2N$ additions/subtractions and N absolute operations. Furthermore, blocks which are encoded in *LOW* or *HIGH* modes are sampled using SRM, necessitating $N_B log_2 N_B$ additions/subtractions per block. For the sake of convenience, T_{LH} denotes the total number of blocks encoded in *LOW* or *HIGH* mode using the proposed method. On the contrary, non-adaptive schemes require $N_B log_2 N_B$ additions/subtractions per block for BCS sampling using SRM. We report in Table 1 the computational complexity required in the case of the proposed adaptive encoding strategy and the conventional non-adaptive method.

We quantify in Table 2 the computational complexity with a particular case for more detailed analysis. For this comparative analysis, we divide video frames of size 288×352 pixels into blocks of size 16×16 pixels. Since the proposed method adaptively encodes blocks of the non-key frame, the corresponding computational cost varies for each frame depending on values of the lower threshold. For this reason, we compare the complexity in terms of the total number of additions/subtractions required per 100 frames. For the fairness of evaluations, the lower threshold in the proposed method is set to 600 which allows to capture only high correlations between successive frames and works well for all video sequences. The reduction in the complexity is measured using computational complexity reduction ratio (CCRR) and is given as,

$$\text{CCRR} = \left(1 - \frac{\text{no. of computations for proposed scheme}}{\text{no. of computations for existing scheme}}\right) \times 100. \tag{12}$$

Table 2 vividly illustrates the benefits of the proposed method from the computational point of view. It is worth noting that the proposed method requires comparatively more computations for *Traffic* video sequence which is more dominated by fast-moving objects. This is consistent with the fact that usually large frame-to-frame differences will occur for videos with fast motion activity. Therefore, the proposed adaptive method encodes a large number of blocks in *Low* or *HIGH* mode, thereby causing an increase in the associated computational cost. But it should be noted that even in

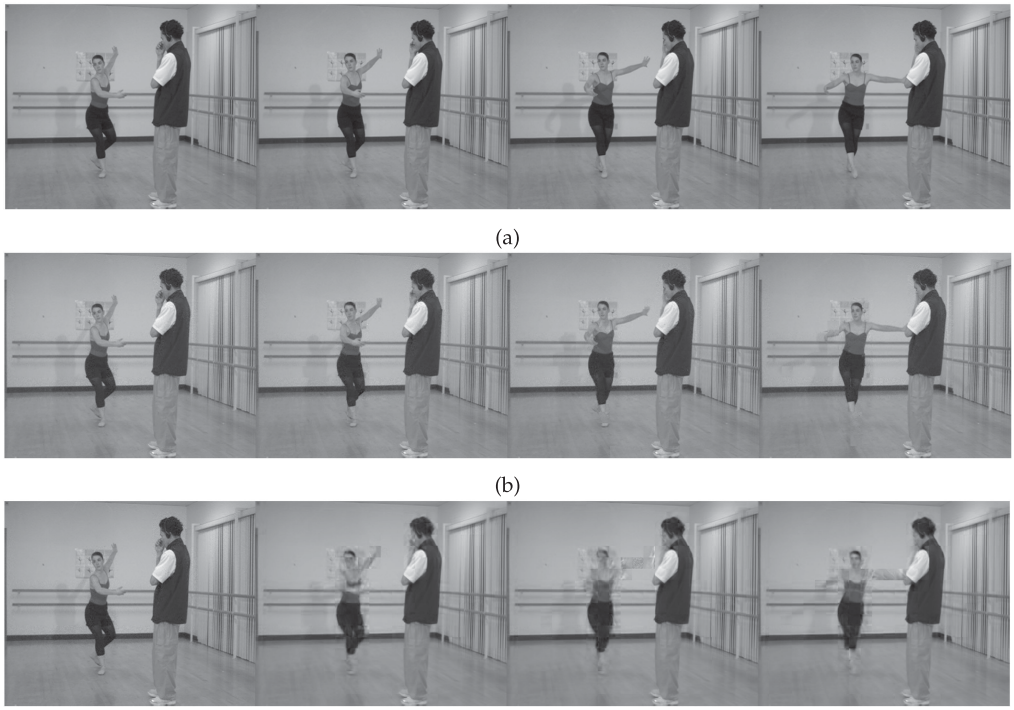


Fig. 15. (a) Ground truth (b) Proposed adaptive video coding (Average bit-rate of non-key frames = 0.27 bpp, PSNR = 39.5 dB, SSIM = 0.73) (c) MH-RRS (Average bit-rate of non-key frames = 0.27 bpp, PSNR = 34.92 dB, SSIM = 0.67).

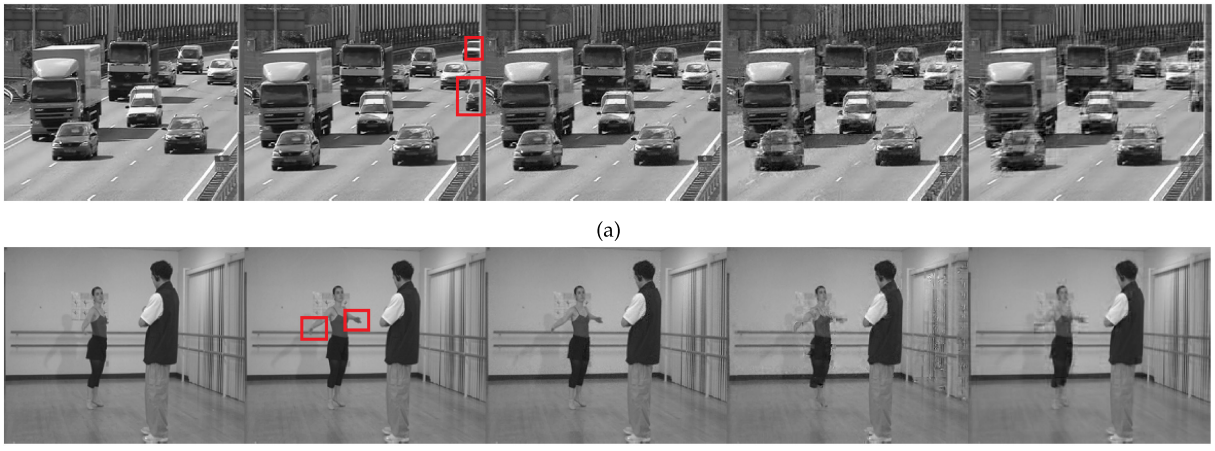


Fig. 16. Recovery comparison of various methods on (a) *Traffic* and (b) *Ballet* video frames (left to right: Original key frame, Original non-key frame, Proposed adaptive method, MH-RRS scheme, and MC-BCS scheme1).

such a scenario, the number of computations is lesser than the corresponding non-adaptive scheme. The proposed method is particularly designed for video surveillance applications wherein the videos include mostly very low to moderate objects motion. In this scenario, the high correlation between successive video frames reduces the average number of blocks to be processed in *LOW* or *HIGH* mode, thereby diminishing the computational burden.

TABLE 1
Computational Complexity of the Proposed Adaptive Encoding Scheme in Comparison with Non-Adaptive Scheme

Algorithm	No. of additions/ subtractions	No. of absolute operations
Proposed adaptive encoding scheme	$2N + T_{LH} \times N_B log_2 N_B$	N
Non-adaptive encoding scheme	$T \times N_B log_2 N_B$	

4.5.4 Transmission Overhead

The transmission overhead in the proposed method is due to adaptive encoding strategy and quantization process. The proposed adaptive encoding scheme requires an indexing mechanism to support synchronization between the encoder and decoder. For this reason, the encoder transmits bits 0, 10, and 11 indicating that the corresponding block of the non-key frame is encoded in *SKIP*, *LOW* and *HIGH*

TABLE 2
Computational Complexity Analysis in Terms of Average no. of Additions/Subtractions using the Proposed Adaptive Method and Non-Adaptive Encoding Scheme

Video Sequence	Proposed method	Non-adaptive scheme	CCRR
Crew	53913600 ($2^{25.74}$)	81100800 ($2^{26.27}$)	33.5
Break-dancer	72038400 ($2^{26.10}$)	81100800 ($2^{26.27}$)	11.17
Ballet	53760000 ($2^{25.68}$)	81100800 ($2^{26.27}$)	33.71
Traffic	80898048 ($2^{26.3}$)	81100800 ($2^{26.27}$)	0.75

mode respectively. We quantify the overhead in bits in the proposed adaptive scheme for the particular case. Consider the video frame of size 288×352 pixels is sampled through BCS with 16×16 blocks i.e., the video frame is divided into 396 non-overlapping blocks. With the assumption that all blocks are processed in *LOW* or *HIGH* mode (i.e., considering maximum indexing bits), the upper bound on the overhead in terms of bits due to the proposed indexing scheme can be given as,

$$\text{Transmission overhead in bits} \leq 792. \tag{13}$$

In the case of the quantization process, the quantization step size and minimum or maximum values of CS measurements need to be transmitted to the decoder. In the proposed approach, 8 bits each are allocated to represent the quantization step size and a minimum value of CS measurements. In essence, the upper bound on transmission overhead in the proposed method can be given as,

$$\text{Transmission overhead} \leq \frac{808}{288 \times 352} = 0.0080 \ \text{bpp}. \tag{14}$$

It is worth noting that due to the high correlation between successive video frames very less number of blocks will be processed in *LOW* or *HIGH* mode in a practical case, thereby diminishing the transmission overhead.

5 Conclusions

In this paper, we present a novel adaptive video coding system based on block compressive sensing. The proposed system makes use of the correlation between CS measurements resulting from the correlated blocks of successive video frames for efficient video coding. The system adaptively allocates different compression ratios for various blocks of the non-key frame depending on the algebraic difference between measurement vectors corresponding to each block of the non-key frame and its co-located block in the nearest key frame. The proposed technique helps to remove the redundancy between frames at the encoder while solving occlusion issues. This brings in reduced transmission cost without adversely affecting the complexity and resource requirement of the CS encoder.

We proposed the compressed domain motion estimation and prediction technique at the decoder that takes advantage of RIP property of the sensing matrix to provide high-quality motion predicted frame from the far fewer number of available CS measurements. We modeled motion estimation as measurement vector matching problem as opposed to spatial domain block matching technique in conventional video coding. Further improvement in the quality of motion predicted frame is achieved through the proposed measurement estimation technique which exploits the hidden correlation between CS measurements of the non-key frame and corresponding motion predicted frames. In addition to high coding efficiency, the proposed system also exhibits much better error resilience performance compared conventional video coding standard. The performance of the proposed system is analyzed by applying it to a wide range of video sequences. Simulation results demonstrate significant improvement in coding efficiency over existing video CS techniques.

References

[1] X. Tian, C. Zhao, H. Liu, and J. Xu, "Video on-demand service via wireless broadcasting," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2970–2982, Oct. 2017.

[2] D. Ho, G. S. Park, and H. Song, "Game-theoretic scalable offloading for video streaming services over LTE and WiFi networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1090–1104, May 2018.

[3] Y. Shen, W. Hu, M. Yang, J. Liu, B. Wei, S. Lucey, and C. T. Chou, "Real-time and robust compressive background subtraction for embedded camera networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 2, pp. 406–418, Feb. 2016.

[4] T. Ma, M. Hempel, D. Peng, and H. Sharif, "A survey of energy efficient compression and communication techniques for multimedia in resource constrained systems," *IEEE Commun. Surveys Tutorials*, vol. 15, no. 3, pp. 963–972, Jul.-Sep. 2013.

[5] J. Xu, Y. Andrepoulos, Y. Xiao, and M. van Der Schaar, "Non-stationary resource allocation policies for delay-constrained video streaming: Application to video over internet-of-things-enabled networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 782–794, Apr. 2014.

[6] W. Ji, X. Ji, and Y. Chen, "Feedback-free binning design for mobile wyner-ziv video coding: An operational duality between source distortion and channel capacity," *IEEE Trans. Mobile Comput.*, vol. 16, no. 6, pp. 1615–1629, Jun. 2017.

[7] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[8] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[9] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] S. Pudlewski, A. Prasanna, and T. Melodia, "Compressed-sensing-enabled video streaming for wireless multimedia sensor networks," *IEEE Trans. Mobile Comput.*, vol. 11, no. 6, pp. 1060–1072, Jun. 2012.

[11] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1182–1195, Jun. 2017.

[12] A. C. Sankaranarayanan, P. K. Turaga, R. G. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *Proc. 11th Eur. Conf. Comput. Vis.*, Sep. 2010, pp. 129–142.

[13] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, Nov. 2014.

[14] J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe, "Flutter shutter video camera for compressive sensing of videos," in *Proc. IEEE Int. Conf. Comput. Photography*, 2012, pp. 1–9.

[15] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[16] J. E. Fowler, S. Mun, E. W. Tramel, et al., "Block-based compressed sensing of images and video," *Foundations Trends® Signal Process.*, vol. 4, no. 4, pp. 297–416, 2012.

[17] C. Zhao, S. Ma, J. Zhang, R. Xiong, and W. Gao, "Video compressive sensing reconstruction via reweighted residual sparsity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1182–1195, Jun. 2017.
[18] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in*Proc. 16th IEEE Int. Conf. Image Process.*, 2009, pp. 1393–1396.
[19] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[20] A. S. Unde and P. Deepthi, "Block compressive sensing: Individual and joint reconstruction of correlated images," *J. Visual Commun. Image Representation*, vol. 44, pp. 187–197, 2017.
[21] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 139–154, Jan. 2012.
[22] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[23] H. R. Tohidypour, H. Bashashati, M. T. Pourazad, and P. Nasiopoulos, "Online-learning-based mode prediction method for quality scalable extension of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2204–2215, Oct. 2017.
[24] T. K. Lee, Y. L. Chan, and W. C. Siu, "Adaptive search range for HEVC motion estimation based on depth information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2216–2230, Oct. 2017.
[25] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 80–94, Sep. 2004.
[26] L. Pan, H. Gao, H. Gao, and Y. Liu, "A spatial correlation based adaptive missing data estimation algorithm in wireless sensor networks," *Int. J. Wireless Inf. Netw.*, vol. 21, no. 4, pp. 280–289, 2014.
[27] H. Zhang, J. M. Moura, and B. Krogh, "Estimation in sensor networks: A graph approach," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, 2005, pp. 203–209.
[28] Y. Chi, "Low-rank matrix completion," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 178–181, Sep. 2018.
[29] M. Le Pendu, X. Jiang, and C. Guillemot, "Light field inpainting propagation via low rank matrix completion," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1981–1993, Apr. 2018.
[30] A. S. Unde and P. Deepthi, "Rate–distortion analysis of structured sensing matrices for block compressive sensing of images," *Signal Process.: Image Commun.*, vol. 65, pp. 115–127, 2018.
[31] L. Wang, X. Wu, and G. Shi, "Binned progressive quantization for compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 2980–2990, Jun. 2012.
[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Amit Satish Unde received the BE degree in electronics engineering from the Dr. J. J. Magdum College of Engineering, Jaysingpur (Shivaji University, Kolhapur), in 2009, and the MTech in signal processing from National Institute of Technology, Calicut, India, in 2012. He is currently doing his research with the Department of Electronics and Communication Engineering of the National Institute of Technology, Calicut. His research interests include compressive sensing, multimedia surveillance, security and privacy. He is a student member of the IEEE.

Deepthi P. Pattathil received the BTech degree in electronics and communication engineering from the N.S.S. College of Engineering, Palakkad (Calicut University), in 1991, and the MTech degree in instrumentation from the Indian Institute of Science, Bangalore, in 1997, and the PhD degree from the National Institute of Technology, Calicut in the field of Secure Communication, in 2009. She has been working as Faculty in institutions under IHRD, Thiruvanthapuram from 1992 to 2001 and in the Department of Electronics and Communication Engineering, National Institute of Technology, Calicut from 2001 onwards. Her current interests include Signal Processing with Security Applications, Cryptographic system implementations, Information Theory and Coding Theory. She is a member of the IEEE.