

Article

# Block Compressive Sensing Single-View Video Reconstruction Using Joint Decoding Framework for Low Power Real Time Applications

## Mansoor Ebrahim<sup>1,\*</sup>, Syed Hasan Adil<sup>1</sup>, Kamran Raza<sup>1</sup> and Syed Saad Azhar Ali<sup>2</sup>

- <sup>1</sup> Faculty of Engineering, Science & Technology, Iqra University, Karachi 75500, Pakistan; hasan.adil@iqra.edu.pk (S.H.A.); drkamran@iqra.edu.pk (K.R.)
- <sup>2</sup> Center for Intelligent Signal and Imaging Research (CISIR), Electrical and Electronics Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Malaysia; saad.azhar@utp.edu.my
- \* Correspondence: mebrahim@iqra.edu.pk

Received: 29 September 2020; Accepted: 29 October 2020; Published: 10 November 2020



Abstract: Several real-time visual monitoring applications such as surveillance, mental state monitoring, driver drowsiness and patient care, require equipping high-quality cameras with wireless sensors to form visual sensors and this creates an enormous amount of data that has to be managed and transmitted at the sensor node. Moreover, as the sensor nodes are battery-operated, power utilization is one of the key concerns that must be considered. One solution to this issue is to reduce the amount of data that has to be transmitted using specific compression techniques. The conventional compression standards are based on complex encoders (which require high processing power) and simple decoders and thus are not pertinent for battery-operated applications, i.e., VSN (primitive hardware). In contrast, compressive sensing (CS) a distributive source coding mechanism, has ransformed the standard coding mechanism and is based on the idea of a simple encoder (i.e., transmitting fewer data-low processing requirements) and a complex decoder and is considered a better option for VSN applications. In this paper, a CS-based joint decoding (JD) framework using frame prediction (using keyframes) and residual reconstruction for single-view video is proposed. The idea is to exploit the redundancies present in the key and non-key frames to produce side information to refine the non-key frames' quality. The proposed method consists of two main steps: frame prediction and residual reconstruction. The final reconstruction is performed by adding a residual frame with the predicted frame. The proposed scheme was validated on various arrangements. The association among correlated frames and compression performance is also analyzed. Various arrangements of the frames have been studied to select the one that produces better results. The comprehensive experimental analysis proves that the proposed JD method performs notably better than the independent block compressive sensing scheme at different subrates for various video sequences with low, moderate and high motion contents. Also, the proposed scheme outperforms the conventional CS video reconstruction schemes at lower subrates. Further, the proposed scheme was quantized and compared with conventional video codecs (DISCOVER, H-263, H264) at various bitrates to evaluate its efficiency (rate-distortion, encoding, decoding).

**Keywords:** compressive sensing; video compression; image processing; joint reconstruction; redundancies; residual based coding

## 1. Introduction

The transformation of wireless sensor networks (WSNs) into visual sensor networks (VSNs) has set new horizons in the IT ecological space, making it broadly adapted in various applications. Equally,



the coupling of high-quality cameras with sensors has greatly increased the size of the data that has to be managed and transmitted (increasing the computational burden). Furthermore, as the sensor nodes are battery-operated, power utilization is a key concern that must be considered. One potential solution to the aforementioned issues is to reduce the amount of data transmitted using specific image/video compression techniques. In a conventional video capturing arrangement, a compression algorithm is usually applied to the video data once the complete frame set is obtained to exploit the inter-view and intra-view correlations between the frames. Various approaches have been proposed to exploit these correlations for standard video compression [1,2]. The standard complex-encoder simple-decoder method is mostly used by traditional video compression schemes i.e., at encoder heavy processing compared to the decoder. The encoder usually uses the motion estimation and compensation (ME/MC) method to utilize the correlations between the frames that require high computation. Thus, the conventional schemes are not apt for low power real-time applications, i.e., VSN (battery-powered nodes with primitive hardware). Most of the conventional schemes entail the use of extra computational resources for data processing leading to additional power requirements, with few trade-offs between the reconstruction quality and computational constraints [1].

The emergence of compressive sensing (CS) [3] has widened the approach for 3D reconstruction, multi-camera imaging, armed personnel tracing and surveillance, magnetic resonance imaging (MRI) and seismic identification for applications such as surveillance, mental state monitoring, driver drowsiness and patient care, etc. CS is specifically useful for the transmission of correlated data for low power real-time applications, i.e., visual sensors networks and wireless sensors networks. CS has transformed the idea of conventional data compression schemes as it is based on the principle of exemplifying a signal using a much smaller sampling rate than the standard Nyquist rate. Hence, using CS is one of the best solutions. Unlike the conventional schemes it is based on a simple encoder (i.e., transmitting fewer data) and a complex decoder. It is also known that the primary motivation for considering CS is that it offers the potential to reduce sensor costs and computational complexity significantly. For example, it is proposed that the famous single-pixel camera, with only one photo-sensor, could provide an imaging sensor that is much cheaper to build than one with several million photosensors. Moreover, CS is a form of dimensionality reduction only, and it is not specifically a type of compression [4].

CS makes sense only when the dimensionality reduction that it provides takes place directly within the sensing device's hardware. In other words, the image never exists anywhere in the sensor in its full dimensionality, unlike most of the conventional video coding schemes in which the acquired image is used in its full dimensionality, increasing the load on the sensor node. However, CS can be used for compression, provided that the dimensionality reduction it offers is coupled with quantization and entropy coding to create a bit of data from the CS measurements.

In this paper, a joint decoding (JD) framework is proposed. The proposed scheme aims to decrease the redundancies among video frames at the decoder using side information. The proposed scheme first decodes the encoded keyframes and non-key frames of a video sequence received at the host workstation. After independently decoding the frames, a side information process is initiated. The side information produced is utilized to enhance the reconstruction quality of the non-key frames. The side information process consists of registration and fusion steps used to exploit the inter/intra view correlation between the frames to generate a frame prediction. Next, the difference among the predicted frame measurements and the acquired measurements is calculated to minimize the prediction errors. The difference at the level of the measurement, known as the residual measurements, is recovered to produce the residual. The final frame reconstruction is performed by adding the residual and the predicted frame to compensate for the difference.

The rest of the paper is organized as follows: Section 2 presents an overview of the various existing CS-based video reconstruction coding schemes. A detailed explanation of the proposed JD scheme for single view video reconstruction using block CS is provided in Section 3. In Section 4 the experimental results are presented and the conclusions are drawn in Section 5.

#### 2. Literature Review

Compressive sensing (CS) has emerged as one of the most substantial mechanisms for compressing visual data (video) in recent times. The principle is to sample each video frame independently using CS (encoder), and then at a decoder the correlations within the video frames are exploited using joint reconstruction schemes for image reconstruction. In literature various video reconstruction coding schemes have been proposed (dictionary-based coding, 3D-transformation coding, residual-based coding, etc), each having their benefits and challenges. In the following, only the state-of-the-art residual coding schemes are discussed as the proposed model relies on residual-based coding and will be compared with these schemes. However, complete explanations of the other coding schemes can be found in [5–11].

In [12], the modified-CS-residual scheme is presented. The scheme is based on a residual reconstruction approach in which side information helps manage the reconstruction issue related to sparse signals (minutest quantity of linear projections). The least mean squares or Kalman filter estimation methods are used to produce side information. However, this side information can be exposed to errors. The proposed scheme aims to counter the convex relaxation problem correlated to data restraints and sparsity outside the side information.

A reconstruction scheme known as k-t FOCUSS is presented in [13,14]. It is also built on a similar residual reconstruction idea. The scheme first generates side information assuming n keyframes using disparity (compensation/estimation) predictions and a residual encoding approach. This leads to an ideal sample allocation among the estimated and residual steps. The residuals among the bidirectional (DC/DE) estimation of each keyframes and non-key frame is used to attain the reconstruction.

A DC/DE-based reconstruction scheme with residuals is introduced in [15]. The scheme generates the side information by integrating DC/DE-based prediction and is referred to as DC-BCS-SPL. The final reconstruction is aided by calculating the residuals among the side information and the original view.

Also, in [16], a DC/DE-based joint reconstruction scheme is proposed. The proximal gradient method is adopted by the proposed scheme to resolve the optimization problem.

The motion (compensation/estimation)-based scheme is introduced in [17]. The proposed MC/ME scheme is incorporated into the block compressive sensing (BCS-SPL) video restoration method and called MC-BCS-SPL. In this approach, each frame of the video sequence is initially sampled using random block-based CS measurements (at the encoder) and transmitted to the decoder. The received encoded measurements are then decoded and the proposed residual dependent ME/MC approach is applied to generate the final view. The proposed scheme generates video frames alternately along with the motion fields associated with the frames i.e., using one to improve the other iteratively.

In [18], a joint reconstruction algorithm using MC/ME and fusion is presented. The proposed scheme uses the down-sample approach (lowering sampling rate of the views) and then MC/ME and a fusion approach are applied on the down-sampled views to produce a view prediction, which helps generate the final view.

A view estimation and residual reconstruction-based joint reconstruction method is introduced in [19]. The method integrates a MC/ME approach into the reconstruction to generate the side information. The side information then helps in the reconstruction of the final output.

The residual coding schemes discussed above help improve the reconstruction quality of the video frames, however, they are exposed to a few issues such as imprecise estimation or computational intricacy. For example, Ref. [12] claims that the proposed scheme is more suitable for video applications and uses least mean squares (LMS) or Kalman filter prediction methods. However, the scheme is exposed to prediction errors. The schemes proposed in [13–19] are mostly based on DC/MC and DE/ME prediction methods. In such a scheme, precise predictions are hard to achieve if basic transformation (translation/affine) models are used. This is because video captured from different view angles might display some distortions that are hard to handle by using fundamental transformations. Thus, the proposed scheme usually uses a more complex transformation model to resolved prediction issues.

However, such a solution might lead to an additional computational burden. A summary of CS-based residual coding schemes for single view video reconstruction is presented in Table 1.

Work	Scheme	Description	Issues
[12]	Kalman-Filter (KF) based Prediction	Make use of side information to handle the sparse signals reconstruction issue (minimum number of linear projections). It also resolves the convex relaxation issue linked with data restraints and sparsity, outer the side information.	It is usually viable for video applications, as it assumes that the sparsity pattern develops progressively from frame to frame. Besides, Prediction make use of state dynamics and measurement models. The prediction might fail due to incorrect initialization of the model and filter. A tree-structured KF algorithm can be used to cater the issue at the cost of higher computational resources.
[13]	Disparity Estimation & Compensation (DE/DC) based	Incorporates disparity estimation and compensation prediction methods into the reconstruction process of BCS-SPL that generate side information aiding final reconstruction.	Introduces discontinuities at the block borders (blocking artefacts). Accurate predictions are hard to achieve with fundamental
[14]	[14]	Solve the optimization problem by implementing the proximal-gradient method, and side information is generated by DE/DC prediction approach.	estimation, and compensation algorithms as images/frames captured from different view angles may exhibit some deformations.
[15,16]	_	Uses motion-based prediction and residual encoding to optimize the sample allocation among the estimated and residual encoding steps.	May result in producing false edges and ringing effects. Requires more complex estimation and compensation algorithms that result in additional computational burden.
[17]	Motion Estimation & Compensation	The MC/ME approach is incorporated with the BCS-SPL reconstruction process for the video. The video sequence frames are generated alternatively i.e., one helps to improve the quality of the other iteratively.	Significant improvements at higher subrate as compared to lower subrate i.e., the scheme does not accurately predict the motions due to the smaller number of measurements at lower subrates that leads to low-quality initial
[18]	Predication	The scheme uses low sample rate, ME/MC and fusion method to produce a view prediction, which helps in the generation of an ultimate view for the multi-view video.	reconstructions
[19]	-	The MC/ME approach is incorporated with the BCS-SPL reconstruction process for multi-view video.	

Table 1. Summary of various CS-based residual coding schemes for visual reconstruction.

## 3. Proposed Joint Decoding (JD) Framework for Video

In this section, the proposed joint decoding (JD) framework for single view video is discussed. The archetype of the complete system is shown in Figure 1. At the encoder side, we consider a visual node S, capturing a scene (video), and block-based compressive sensing (BCS) is applied on each frame in the video sequence for encoding. The encoded frames are then independently transmitted to the host workstation.



Figure 1. Archetype of the complete system for single-view video coding using the proposed JD (joint decoding).

At the host workstation (decoder), a sequence of J recurrent frames is received from the visual node S, referred to as a group of pictures (GoP). As the video is uninterrupted, it is established that the current GoP is tailed by another GoP. The GoP comprises a keyframe and J-1 non-key frames, represented as  $F_K$  (the primary) and  $F_{NK}$ . The encoding of  $F_K$  and  $F_{NK}$  is performed at the subrate of  $M_K$  and  $M_{NK}$ , respectively, with  $M_K > M_{NK}$ .

At the visual node (encoder), each frame  $F_x$  (where x is the frame number) in the video sequence is initially portioned into a small block of 16 × 16. Then, the sensing matrix  $\Phi_x$  as in Equation (1) is used to sample each block within a frame producing a set of measurements (Y<sub>x</sub>) as in Equation (2) [20–22]:

$$\Phi_{\mathbf{x}} = \begin{bmatrix} \Phi_{\mathbf{x}} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \Phi_{\mathbf{x}} \end{bmatrix}$$
(1)

The set of measurements  $(Y_x)$  produced at the encoder is sent to the host workstation (server) independently. At the server side (decoder), firstly, the received encoded frames  $(Y_x)$  are decoded independently (frame by frame) using the total variation (TV) minimization [23,24] problem till a complete GoP is obtained. The TV minimization makes use of piece-wise smooth features of the signals to provide a better solution within the possible space rather than finding the sparse solution within the transformation domain  $\Psi$ . The basic TV minimization function is given in Equation (4):

$$TV(F) = \sum_{i,j} |F_{i+1,j} - F_{i,j}| + |F_{i,j+1} - F_{i,j}|$$
(3)

$$\mathbf{F} = \operatorname{argmin}_{\mathbf{F}} \left| \mathbf{y} - \Theta \mathbf{F} \right|_{\ell_2} + \lambda \operatorname{TV}(\mathbf{F}) \tag{4}$$

However, the CS reconstruction using TV minimization as in Equation (4) is exposed to extra computational problems that limit its use for CS, i.e., it is harder to computationally access and explain certain properties of TV minimization (non-differentiable, non-linear) than  $\ell_1$  minimization. To counter such issues, a scheme [24] referred as TV-AL3 is presented that uses the conventional augmented Lagrangian (AL) with variable splitting and alternating direction method. The approach decreases the computational burden and provides the same output as standard TV.

After obtaining the complete GoP, the proposed joint decoding (JD) method is implemented to utilize the inter and intra-view correlations among the frames. As shown in Figure 1, the first frames of the current and next GoP's aid as the keyframes  $F_K$  (reference frames) for the JD to produce side information of frames (prediction, residual) that will help in enhancing the quality of (J-1) non-key frames  $F_{NK}$  of the current GoP.

#### 3.1. Correlation Estimation of the CS Measurements among Adjacent Frames

A correlation estimation of the CS measurements among the adjacent frames is presented in this section. As discussed earlier, adjacent frames with a video sequence have a high inter-frame level correlation with each other. Consequently, it can also be assumed that the CS measurements of such frames are also highly correlated. Although, acquiring CS measurements is entirely different from linear transformations, as CS measurements form a random Gaussian distribution. The correlation among the two random entities can be estimated by using Pearson's correlation coefficient. It is defined as "the correlation coefficient of two random variables is a measure of their linear dependence". Consider the CS measurements of two adjacent frames  $Y_n$  and  $Y_{n+1}$  of a video sequence, then their correlation coefficient is estimated as:

$$\rho(\mathbf{Y}_{n}, \mathbf{Y}_{n+1}) = \frac{1}{\varphi - 1} \sum_{n=1}^{\varphi} \left( \frac{\mathbf{Y}_{n} - \mu_{\mathbf{Y}_{n}}}{\sigma_{\mathbf{Y}_{n}}} \right) \cdot \left( \frac{\mathbf{Y}_{n+1} - \mu_{\mathbf{Y}_{n+1}}}{\sigma_{\mathbf{Y}_{n+1}}} \right)$$
(5)

where,  $\varphi$  is the number of measurements,  $\mu_{Y_n}$  and  $\mu_{Y_{n+1}}$  are the mean of  $Y_n$  and  $Y_{n+1}$ , respectively, while  $\sigma_{Y_n}$  and  $\sigma_{Y_{n+1}}$  are the standard deviation of  $Y_n$ , and  $Y_{n+1}$ , respectively.

The evaluation of CS measurement correlation among various frames is performed on various standard grayscale CIF video sequences (Foreman, Coastguard, Container, Hall Monitor, Mother-Daughter). Each video sequence contains 300 frames and is divided into GoPs of size 8. The CS measurements for each frame within a video sequence are obtained at a measurement rate of 0.5. The CS measurement correlation coefficient for each non-keyframe for various video sequences is estimated and shown in Figure 2. The results presented are an average of all the correlations for non-key frame w.r.t keyframes. It can be noticed that all the frames within the video sequence show a high correlation, i.e., above 0.92. Moreover, for moderate motion content video sequences, the correlation is also higher.



**Figure 2.** Correlation analysis of CS measurements among the adjacent frames of various CIF video sequences.

## 3.2. Joint Decoding Framework

This subsection discusses in detail the decoding process of the proposed JD scheme. The proposed JD consists of three significant steps, as presented in Figure 3.



Figure 3. Reconstruction of frames using the proposed joint decoding (JD) framework.

The descriptions of each step are provided in the following sections.

## 3.2.1. Frame Estimation

This step proposes a frame estimation technique using registration and fusion methods. The idea is to utilize the frames' correlations to estimate the J-1 non-key frame ( $F'_{NK}$ ) in the GoP from the keyframes ( $F'_{K}$ ). The proposed method assists in utilizing the inter and intra-view correlations between the frames and produce a set of predicted non-key frames.

## **Registration Approach**

The frame estimation step is initialized by performing registration (intensity) on the two independently generated keyframes  $F'_K$  in the GoP. The aim is to align the  $F'_K$  on the same plane as of  $F'_{NK}$  ( $F'_K$  is aligned to  $F'_{NK}$ , and the redundancies between them are exploited). Firstly, a phase correlation method (finding the gross alignment) is used on the  $F'_K$  and  $F'_{NK}$  frames to calculate the initial transformation matrix. Next, the translation transformation over affine is used for aligning  $F'_K$ 

w.r.t.  $F'_{NK}$  to produce transformed  $F'_{K}$  called  $F''_{KT}$ . The reason for using a translation transform is that affine transform makes sense when multiple images are not on the same plane and are to be rectified, whereas, in the case of a video sequence, each frame is usually on the same plane, so it would be sufficient to consider translation transform.

The similarity metric and optimization function is then applied on the transformed frame  $F''_{KT}$  to estimate the registration precision and generate the final registered image  $F''_{K}$ , as shown in Equation (6):

$$F'' K = \operatorname{argmin}_{N} \left\| \sum_{y \in L_{F'_{KT}}} \sum_{x \in L_{F'_{NK}}} P_{F'_{KT}F'_{NK}}(y, x; a) \log_{2} \left( \frac{P_{F'_{KT}F'_{NK}}(y, x; a)}{P_{F'_{KT}}(x)P_{F'_{NK}}(y; a)} \right) \right\|$$
(6)

The similarity metric makes use of the mutual information (MI) method where the optimizer uses one evolutionary (OE). The evolutionary optimizer is used to exploits the alike alignment that usually occurs within the frames of a video sequence that better helps the OE optimizer than the other optimizers such as gradient descent (GD). The optimizer is one of the vital parameters of the registration process that defines the method to exploit the attained similarity metric M and produce a final result. Once both the keyframes  $F''_K$  is registered, a wavelet-based fusion process is applied on them to produce the predicted frame  $F_P$ .

#### **Fusion Approach**

In the fusion process, the registered keyframes  $F''_K$  are down-sampled up to three levels using symlet 4-tap filter into their respective approximation (A) and detail (D) coefficient maps. Then, the point-to-point operations are performed to fuse the A and D coefficients in the two decomposition maps. The coefficients A and D are set as shown in Equations (7) and (8), respectively:

$$A_{Mean}(a,b) = (A_K(a,b) + A_{NK}(a,b))/2$$
(7)

$$D_{Mean}(a,b) = (D_K(a,b) + D_{NK}(a,b))/2$$
(8)

The average magnitudes for each A and D coefficient having a similar coordinate of the two decomposition maps is calculated. The average value obtained then serves as the output for the fused map (FM). Once the fused map is generated, an inverse transformation is applied to generate the  $F_P$ 's predicted frames. The proposed frame estimation method predicts the object motions and produces predicted frames  $F_P$ :

$$F_{p} = \text{Transform}^{-1}(\text{FM}) \tag{9}$$

The defined set of parameters for registration and fusion processes remained best when related to other sets of parameters for various video sequences that have been used.

## 3.2.2. Residual Reconstruction

In this step, the predicted frames  $F_P$  is projected onto the measurement basis  $Y_P = \Phi_x I_P$ . Once, the predicted measurements  $Y_P$  is generated, they are then differentiated from the given frame measurements  $Y_x$  to produce the residual measurement  $Y_r$  as in Equation (10):

$$Y_r = Y_x - Y_P \tag{10}$$

The obtained residual measurements  $Y_r$  are then decoded by using BCS-TV-AL3 method to generate the residual frames  $F_r$ .

#### 3.2.3. Final Frame Reconstruction

The final reconstructed frames  $F''_{NK}$  within the GoP are generated by adding the  $F_r$  and  $F_P$  frames. It is a standard point-to-point addition that is expressed in Equation (11). By doing so, uniformity in terms of frame measurements (Y) is achieved, i.e., the measurements computed for  $F''_{NK}$  is to some extent equal to the measurements  $Y_{NK}$ :

$$F''_{NK} = F_r + F_P \tag{11}$$

After the keyframes,  $F'_K$  ( $F_0$  and  $F_J$  from  $Y_0$  and  $Y_J$ ) are reconstructed using BCS-TV-AL3, they are used as the reference frames for reconstructing the non-key frames  $F'_{NK}$  between them.

The proposed scheme produces the non-key frame  $F''_1$  from  $Y_1$ ,  $F_0$ , and  $F_J$ ; similarly  $F''_2$  is generated from  $Y_2$ ,  $F_0$  and  $F_J$ . The method endures for all the remaining non-key frames. The reconstructed frames' quality is expected to drop as the distance of the key frames from the non-key frames increases. Thus, the quality of the reconstruction may decrease with the increase in the GoP size (J). A complete GoP (J = 8) reconstruction for News video sequence is shown in Figure 4. The highlighted red dotted circles clearly show that as the difference among the key frames and non-key frame increases the reconstruction quality decreases.



**Figure 4.** Visual quality analysis for a complete GoP (J = 8) of News video sequence at a subrate of 0.1 using proposed JD-TV framework.

#### 4. Experimental Results

## 4.1. Setup

In this subsection, the validation of the proposed JD scheme incorporated with TV-AL3 (i.e., JD-TV) in terms of performance and efficiency is presented. A standard set of grayscale CIF [25] and HD resolution [26] video sequences with sizes of  $352 \times 288$  and  $1024 \times 678$  are selected on which the

proposed scheme is applied for evaluation. The selected video sequences are based on different parameters such as # of frames, motion content type (slow, moderate, fast). The details of the video sequences (# frames, motion type) are presented in Table 2. The video sequence indicated as low, moderate, and high contents are based on spatial details plus camera and object movement.

The experimental setup comprises of the application of proposed JD-TV having various GoP (J) sizes (3, 5, and 8) to validate its efficiency at various levels. The assessment is performed by observing the peak signal to noise ratio (PSNR) at various sampling rates (subrate). Subrate or sampling rate is the number of recorded samples (pixel) per unit distance when converting from an analog signal to digital and is give as  $S = M^B/N$ . Where, M is the number of CS measurements,  $N = B \times B$  (B = block size).

	Video Sequence	No. Frames	Content-Type
	Hall-Monitor	300	Low
	Mother Daughter	300	Low
CIF	Coast-Guard	300	Moderate
(352 × 288)	Foreman	300	Moderate
	Mobile Calendar	300	High
	Stefan	300	High
UD	Love Birds	300	Low
HD (1024 × 678)	News	300	Moderate
	Book-Arrival	300	Moderate

Table 2. Various grayscale CIF and HD standard video sequences.

Also, the structural similarity index (SSIM) is also recorded as it is anticipated as more precise and consistent (human visual perception) than PSNR. All PSNR and SSIM values presented are the mean of five independent experiments. The values are averaged because the  $\Phi$  is of random nature, and thus, the image quality may vary. A smaller block size of  $16 \times 16$  is implemented compared to larger ones  $(32 \times 32, 64 \times 64)$ , as it is observed that larger block size will produce more samples and will require additional time and energy to transmit. Although, the larger block size provides better reconstruction quality but at the expense of complexity. The selection of block size is a tradeoff between reconstruction quality and computational complexity. In this regard, it is not feasible for a battery-powered device to always encode and transmit the captured images at larger block size of  $16 \times 16$  or  $32 \times 32$  as it provides better image quality with less computational complexity. The encoding of all keyframes are performed at a fixed subrate of 0.5 while the non-key frames within a GoP are encoded at various lower subrates (0.05, 0.1, 0.15, 0.2, 0.25, 0.3).

## 4.2. Effect of GoP Size on the Performance of Proposed JD-TV

In this subsection, the proposed JD-TV is assessed and equated with independent BCS-TV-AL3 using various GoP sizes. Table 3 shows the impact of three different GoP sizes (3, 5, 8) on the performance of the proposed JD at various subrates for several video sequences.

The results shown are the average values of the non-key frames obtained for each complete video sequence. The result shows that the proposed scheme performs well when compared with independent BCS-TV-AL3. For low, moderate, and high content types, the improvement over independent BCS-TV-AL3 is 3 dB-7 dB, 2.5 dB-5 dB and 2 dB-4 dB, respectively, on average for all GoP sizes. Also, it is also noticed that the gain for reconstructed low content video sequence is better than moderate and high content as the association between the frames is higher than in moderate and high content videos. The better correlation results in more precise frame estimation and residual reconstruction.

	Mother-Daughter							
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	24.88	30.36	32.35	33.73	35.55	37.15	
	JD-TV GoP3	36.35	40.10	41.03	41.41	41.94	42.41	
	JD-TV GoP5	34.16	35.55	36.73	38.43	39.88	40.72	
	JD-TV GoP8	32.00	34.26	35.81	37.65	38.63	39.61	
Low Motion Content Video Sequence			Hal	l Monitor				
I	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	21.20	23.71	25.26	26.68	28.18	29.71	
	JD-TV GoP3	32.31	32.70	33.28	33.93	34.45	35.08	
	JD-TV GoP5	29.52	31.56	32.10	32.98	33.64	34.26	
	JD-TV GoP8	28.98	30.65	31.77	32.81	33.83	34.85	
			Fe	oreman				
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	23.66	25.38	27.05	29.0	30.70	33.83	
	JD-TV GoP3	29.82	31.94	32.31	33.59	34.14	35.81	
	JD-TV GoP5	26.3	29.9	31.06	32.84	33.59	35.03	
Moderate Motion	JD-TV GoP8	26.29	27.86	30.05	31.42	32.96	34.61	
Content Video	Coast Guard							
Sequence	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	21.86	23.98	25.01	26.15	27.18	27.99	
	JD-TV GoP3	29.18	29.73	30.22	30.9	31.51	32.00	
	JD-TV GoP5	25.43	26.84	27.51	28.6	29.27	30.02	
	JD-TV GoP8	24.53	25.61	26.69	27.6	28.38	29.14	
			Mobi	le Calenda	r			
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	17.79	19.71	20.69	21.77	22.71	23.61	
	JD-TV GoP3	24.54	25.38	25.82	26.49	27.2	27.76	
	JD-TV GoP5	22.46	23.49	24.26	24.95	25.71	26.39	
High Motion	JD-TV GoP8	21.26	22.39	23.35	24.08	24.9	25.76	
Content Video			5	Stefan				
Sequence	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV	19.92	21.76	22.89	24.14	25.23	26.42	
	JD-TV GoP3	24.3	27.06	27.13	28.71	29.53	30.44	
	JD-TV GoP5	23.68	25.34	26.59	27.87	28.52	29.76	
	JD-TV GoP8	23.39	25.08	26.27	27.07	27.93	29.00	

**Table 3.** Average PNSR evaluation of proposed JD-TV scheme and the independent BCS-TV-AL3 for different low, moderate, and high motion content video sequences at various GoP.

Note: The bold values represent the highest PSNR (dB) achieved at specific subrate for video sequence.

It is also observed that the larger the GoP size, the lower the PSNR gain will be. This is because the principle of the proposed scheme is to reconstruct the non-key frames by utilizing the correlation of the keyframes. Thus, the non-key frames closer to the keyframe are more correlated than those further away. For GoP = 3 the average gain is 3dB to ~6dB, while for GoP = 8 ~2dB to 4dB is achieved.

Additionally, as the subrate increases, the gain decreases. In other words, as the keyframes,  $F_K$  is encoded at a subrate higher than that of non-key frames  $F_{NK}$ , it produces a larger measurement, covering the related smaller set of  $F_{NK}$  measurements. This, in result decreases the estimation errors of  $F_{NK}$  that arises due to smaller set measurements and generates a better form of  $F_{NK}$ .

The reconstructed frames' visual quality is also tested by using SSIM metric on the same video sequences used for PSNR. The SSIM graphs of various video sequences with GoP sizes of 3, 5, and 8 at different subrates are presented in Figure 5.



**Figure 5.** Average SSIM evaluation at GoP sizes of 3, 5, and 8 for different video sequences at different subrates for independent BCS-TV-AL3and proposed JD.

The SSIM shows a similar trend in terms of gain as of PSNR. Moreover, the result also indicates an enhancement in visual quality and a notable increase over the independent scheme. Also, the proposed scheme is tried at higher GoP sizes (16, 32). However, the frame reconstruction was not substantial enough to be imitated. In the case of VSN, the frequent transmission of the keyframes is not encouraged due to limited battery life as in the instance with GoP size 3 and 5; it will upsurge the computational load at the encoder (Low power node). Thus, GoP size 8 is measured as a more stable point between all the GoP sizes and will later opt-in trials.

## 4.3. Subrate

This subsection presents the impact of different  $F_K$  subrate used in the proposed JD-TV on the  $F_{NK}$  reconstruction. A GoP size of 8, as indicated from the former experimentation (balance point) instead of 16/32 is opted. The evaluation is carried on two arrangements.

- First setup, encoding of both  $F_K$  and  $F_{NK}$  at the similar subrate i.e.,  $M_K = M_{NK}$  (0.05, 0.1, 0.15, 0.2, 0.25, 0.3).
- Second setup, the encoding of  $F_K$  at a static subrate of MR = 0.5 and  $F_{NK}$  at various subrates i.e.,  $M_{NK}$  (0.05, 0.1, 0.15, 0.2, 0.25, 0.3) with 0.05 interval among the subrate.

The evaluation result shown in Table 4 indicates the importance of subrate on the reconstruction quality of the frames. In other words,  $F_K$  subrate has a substantial effect on the  $F_{NK}$  reconstruction quality. In the first setup, where subrate  $M_K = M_{NK}$  the reconstruction quality, when compared with traditional scheme, improves on average from 2 dB–3.5 dB. However, when the subrate is  $M_K > M_{NK}$ , it is noted that the gain increases significantly (3.5 dB–5 dB) when compared with JD-TV ( $M_K = M_{NK}$ ) and BCS-TV. The reason that the reconstruction quality at subrate  $M_K > M_{NK}$  is better than at subrate  $M_K = M_{NK}$  is because, when encoding of both  $F_K$  and  $F_{NK}$  is performed at the same subrate, the generated  $F_K$  does not comprehend enough material that could considerably assist the reconstruction of  $F_{NK}$ .

	Hall Monitor							
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	21.70	23.53	24.76	26.44	28.16	29.77	
	$JD-TV (M_R = M_{NR})$	22.88	25.54	27.37	29.34	30.84	32.62	
Low Motion	JD-TV ( $M_R = 0.5$ )	28.98	30.65	31.77	32.81	33.83	34.85	
Content			Mother Dau	ughter				
Video	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	24.88	30.36	32.19	33.73	35.55	37.15	
	$JD-TV (M_R = M_{NR})$	29.02	32.15	33.81	35.57	37.83	39.15	
	JD-TV ( $M_R = 0.5$ )	32.00	34.26	35.81	37.65	38.63	39.61	
			Coast gu	ard				
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	21.86	23.98	25.01	26.15	27.18	27.99	
	$JD-TV (M_R = M_{NR})$	22.32	24.90	25.37	26.41	27.67	28.58	
Moderate Motion	JD-TV ( $M_R = 0.5$ )	24.53	25.61	26.69	27.6	28.38	29.14	
Content	Foreman							
Video	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	23.66	25.38	27.05	29.0	30.70	33.83	
	$JD-TV (M_R = M_{NR})$	24.92	26.39	28.10	29.84	31.35	33.35	
	JD-TV ( $M_R = 0.5$ )	26.29	27.86	30.05	31.42	32.96	34.61	
			Mobile Cal	endar				
	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	16.68	18.58	19.58	20.52	21.58	22.50	
	JD-TV ( $M_R = M_{NR}$ )	17.99	19.43	20.65	21.80	22.88	24.03	
High Motion	JD-TV ( $M_R = 0.5$ )	20.15	21.28	22.24	22.97	23.79	24.65	
Content			Stefar	ı				
Video	Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
	BCS-TV-AL3	19.92	21.76	22.89	24.14	25.23	26.42	
	$JD-TV (M_R = M_{NR})$	20.23	22.65	23.85	25.09	26.12	27.50	
	$JD-TV (M_R = 0.5)$	23.39	25.08	26.27	27.07	27.93	29.14	

**Table 4.** Average PNSR evaluation of proposed JD-TV scheme and the independent BCS-TV-AL3 for different low, moderate and high motion content video sequences at GoP = 8.

Note: The bold values represent the highest PSNR (dB) achieved at specific subrate for video sequence.

## 4.4. Proposed JD Framework and HD Video Sequences

In this section, the proposed JD framework is tested with HD video sequences in order to evaluate its effectiveness at various video resolutions. The HD video sequence selected contains both low content and moderate content types. A block size of  $16 \times 16$  and GoP size of 8 is adopted.

The PSNR and SSIM results presented in Tables 5 and 6 respectively, clearly show that the proposed JD framework also performs well for HD video sequences compared with conventional BCS framework. A similar trend as that of CIF video sequence can be noticeable using JD reconstructions. For low motion video (Lovebird) the gain in term of PSNR and SSIM is more significant than moderate motion videos (Newspaper, Book arrival).

Book Arrival							
Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
BCS-TV-AL3	27.24	30.21	32.08	33.64	34.77	35.8	
JD-TV	32.99	33.8	35.04	36.19	36.92	37.76	
Gain	5.75	3.59	2.96	2.55	2.15	1.96	
		N	ewspaper				
Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
BCS-TV-AL3	26.00	29.58	31.71	33.30	34.64	40.13	
JD-TV	32.58	35.06	36.36	37.32	38.23	38.80	
Gain	6.58	5.48	4.65	4.02	3.59	3.33	
		L	ovebird				
Subrate	0.05	0.10	0.15	0.20	0.25	0.30	
BCS-TV-AL3	25.08	28.64	30.35	32.02	33.54	34.86	
JD-TV	34.50	36.52	37.95	38.94	40.08	40.12	
Gain	9.42	8.88	7.60	6.92	6.54	5.26	

**Table 5.** Average PNSR evaluation of proposed JD-TV scheme and the independent BCS-TV-AL3 for different HD video sequences at various subrates having GoP = 8.

Note: The bold values represent the PSNR (dB) gain difference achieved by JD-TV for various video sequence.

**Table 6.** Average SSIM evaluation of proposed JD-TV scheme and the independent BCS-TV-AL3 for different HD at various subrates having GoP = 8.

Book Arrival								
Subrate	0.05	0.10	0.15	0.20	0.25	0.30		
BCS-TV-AL3	0.79	0.85	0.88	0.9	0.9	0.92		
JD-TV	0.91	0.92	0.94	0.95	0.96	0.97		
Gain	0.12	0.07	0.06	0.05	0.06	0.05		
		Ne	ewspaper					
Subrate	0.05	0.10	0.15	0.20	0.25	0.30		
BCS-TV-AL3	0.77	0.84	0.88	0.91	0.92	0.93		
JD-TV	0.93	0.94	0.95	0.97	0.97	0.98		
Gain	0.16	0.10	0.07	0.06	0.05	0.05		
	Lovebird							
Subrate	0.05	0.10	0.15	0.20	0.25	0.30		
BCS-TV-AL3	0.75	0.83	0.87	0.9	0.91	0.93		
JD-TV	0.95	0.96	0.96	0.97	0.97	0.98		
Gain	0.20	0.13	0.09	0.07	0.06	0.05		

Note: The bold values represent the PSNR (dB) gain difference achieved by JD-TV for various video sequence.

#### 4.5. Visual Result Comparison

Visual quality analysis is one of the vital evaluation parameters for the reconstruction of compressed images. In this subsection, the visual quality of the reconstructed frames using proposed JD is analyzed. The proposed framework performs better in terms of PSNR and SSIM at lower sub-rate. However, it is also important to verify that the frame produced at lower subrate are visually perceptible. Thus, we perform the analysis on three different video sequences having low, moderate and high motion contents.

Figure 6 presents the visual results of the random center frame of each GoP selected from the reconstructed video sequence (the similar visual quality was observed for all center frames of each GoP) by using JD-TV and BCS-TV-AL3. The result shows that the visual quality of reconstructed frames by using proposed method at lower subrates is better visually perceptible than the conventional method. In addition, for low motion content video sequence (Mother-Daughter) the proposed JD-TV shows higher visual quality (due to precise frame prediction and residual reconstruction) as compared to moderate (News) and higher Motion content video sequences.



Figure 6. Cont.



Frame #65 of Mobile Calendar at different subrates

**Figure 6.** Visual quality analysis of different video sequences at multiple subrates using independent BCS-TV-AL3 and the proposed JD-TV.

It should also be noted that at a lower subrate (0.05) due to the insufficient estimation information of motion few distortions are detected (red dotted circle). Further, for video sequences with fast-moving objects (Mobile Calendar), the JD-TV is vulnerable to a few distortions as emphasized by the red spotted circle.

### 4.6. Proposed JD-TV v/s Standard CS Video Compression Schemes

In order to analyze the performance of the proposed JD-TV methodically it is compared with conventional CS-based schemes i.e., MS-residual [12], k-t FOCUSS [13], MC-BCS-SPL [17] as referred in Section 2. The results presented in Table 7 (gain of proposed and conventional schemes w.r.t independent BCS) are the average of the total frames obtained at a block size of 16 × 16 and GoP of 8.

	Subrate	0.05	0.10	0.15	0.20	0.25	0.30
	JD-TV	6.88	6.75	6.68	6.51	6.47	6.33
Hall Monitor	MC-BCS-SPL	2.64	3.88	4.96	5.77	5.98	6.21
WOIIIIOI	kt-Focuss	1.50	1.96	2.45	3.21	3.97	4.05
	MS-Residual	0.85	1.06	1.55	2.17	2.88	3.35
	JD-TV	5.96	5.83	5.69	5.57	5.41	5.27
Mother Daughter	MC-BCS-SPL	2.17	3.04	3.77	4.30	4.85	5.14
2	kt-Focuss	1.08	1.87	2.65	3.34	3.99	4.75
	MS-Residual	0.38	0.77	1.27	1.83	2.59	3.05
	JD-TV	2.76	2.65	2.55	2.42	2.31	2.22
Coast	MC-BCS-SPL	0.95	1.35	1.49	1.74	1.98	2.15
Guaru	kt-Focuss	0.45	1.01	1.21	1.40	1.60	1.89
	MS-Residual	0.25	0.54	0.75	0.95	1.20	1.49
	JD-TV	5.56	4.96	4.26	3.69	2.14	2.08
Foreman	MC-BCS-SPL	0.90	2.01	2.81	3.44	3.57	3.74
	kt-Focuss	0.45	0.67	0.81	1.07	1.23	1.59
	MS-Residual	0.15	0.24	0.55	0.77	0.95	1.19
	JD-TV	3.84	3.7	3.58	3.43	3.3	3.17
Mobile Calondar	MC-BCS-SPL	1.07	1.93	2.26	3.06	3.92	4.66
Calelludi	kt-Focuss	0.77	1.13	2	2.63	3.23	3.97
	MS-Residual	0.37	0.66	1.06	1.63	2.13	2.89
	JD-TV	4.28	4.15	4.12	4.07	3.98	3.84
Stefan	MC-BCS-SPL	0.35	0.55	0.93	1.46	2.18	2.91
	kt-Focuss	0.23	0.41	0.67	1.1	1.53	2.06
	MS-Residual	0.16	0.21	1.06	0.55	0.98	1.15

**Table 7.** Average PSNR gain (dB) analysis of the proposed JD with conventional CS schemes for various video sequences.

Note: The bold values represent the highest PSNR (dB) gain achieved at specific subrate for video sequence.

The results in Table 7 clearly show that the proposed JD provides considerable gain for various motion content video sequences at lower subrates than conventional CS-based schemes. Further, is can also be observed that the gain decreases as the subrate increases, which is due to impact of  $F_K$  subrate as discussed earlier. In addition, the focus of the proposed scheme is to provide better reconstruction at lower subrates.

## 4.7. Comparison of Proposed JD-TV with Conventional Video Compression Schemes

As discussed earlier, CS is a type of dimensionality reduction such that the signal at no time exists in the sensor in its full dimensionality. CS is based on a simple-encoder complex-decoder archetype, which is in contrast to the conventional video codecs. In order to use CS as a compression mechanism, quantization and entropy coding must be coupled with CS to generate bit from the CS measurements.

In this section, the proposed JD-TV is coupled scalar quantization- adaptive differential pulse code modulation (SQ-ADPCM) referred as SQ-ADPCM-JD to compare it with state-of-the-art video codecs i.e., DISCOVER [27], H.264 (intra, (I-P-P)) [28] and H.263 (intra, (I-P-P)) [29]. The results shown in Figure 7 are the average values of the frames obtained for each complete video sequence at group of picture (GoP) = 3 and block size = 16 for all implementations.



**Figure 7.** PSNR comparison of proposed SQ-ADPCM-JD with conventional codecs i.e., DISCOVER, H.264 (intra, (I-P-P)) and H.263 (intra, (I-P-P)) at various bitrate for different video sequences.

It can be noted from the results that the CS codec, coupled with SQ-ADPCM, performs better than the standard video codec at various bitrates. For example, when compared with H.263 (Intra) and H.264 (Intra), the CS codec outperforms them for all video sequences at various bitrates. For the case of moderate content type video (Foreman, Coastguard), the CS codec performs noteworthily better than H.263 (I-P-P) at various bitrates, while it should also be noted that the performance of the proposed SQ-ADPCM incorporated with the JD scheme performs better than DISCOVER and H.264 (I-P-P) at lower bitrates. As mentioned earlier, the proposed JD-TV make use of keyframes that contains larger set of measurements (higher subrate) to improve the non-key frames with lower measurement set (low subrates). Thus, the key frames having larger set of measurements superimposes the correlated smaller measurements set encompass by non-key frames. This, helps to decrease the non-key frame prediction errors, that arises due to smaller measurements set and predict a better version of non-keyframes. Further, DISCOVER and H.264 (I-P-P) schemes use the feedback channel to improve the reconstruction of the keyframes.

#### 4.8. Number of Bits

The proposed JD scheme is developed with the idea of benefiting low power real-time applications (VSNs) in terms of efficiency, data transmission and power utilization. This subsection observed the

bit rate savings among the independent BCS-TV-AL3 and the proposed JD-TV for different video sequences at the various reconstruction qualities (PSNR).

The results presented in Table 8 shows the reconstruction quality at different numbers of bits for different video sequences. The results obtained can also be calculated as given in [30]. It can be noted that the number of bits essential for conventional BCS and the proposed JD schemes to achieve the same PSNR varies. The proposed scheme utilizes minimum number of bits as compared to the independent BCS scheme to achieve the same PSNR i.e., for higher motion content video sequence the average saving rate is ~42%, while for lower content video sequence the average saving rate is ~42%, while for lower content video sequence the average saving rate is ~65% for all videos. This is due to the fact that the proposed JD offers improved reconstruction quality at lower bit rates. As the proposed scheme provides better reconstruction at lower bits this helps to reduce the computational burden from the encoder as the encoder will be transmitting small amounts of data. On average, the number of bits saved by the proposed scheme contrary to the independent scheme is ~50% at different reconstruction qualities.

	PSNR	Bits BCS-TV-AL3	Bits JD-TV	Bits Saved	Bits Saving (%)	Bits Average (%)
	~24.03	13068	1980	11088	84	
Hall Monitor	~25.05	17424	5148	12276	72	75
	~25.57	20196	7128	13068	68	
	~23.42	12276	5940	6336	52	
Coast Guard	~24.02	14256	9108	5148	37	40
	~24.53	17424	12276	5148	30	
	~29.37	9108	3960	5148	58	
Mother Daughter	~30.20	13068	7128	5940	45	49
	~31.59	17424	9908	7516	43	
	~18.32	9108	3168	5940	66	
Calendar	~18.78	12276	5148	7128	58	58
	~19.39	14256	7128	7128	50	

**Table 8.** Analysis of the proposed JD and BCS-TV-AL3 scheme in terms of bits saved at various reconstruction qualities (PSNR).

#### 4.9. Execution Time (Encoding/Decoding Complexity)

In this subsection, the encoding and decoding complexity (average compression and reconstruction time) of the proposed JD-TV and other conventional video schemes i.e., DISCOVER [27], and H.264 (intra) [28] with GoP size 3 and block size 16 for various video sequences is presented. Also, the reconstruction complexity of the proposed schemes is compared with other CS-based reconstruction schemes. All the schemes are implemented using MATLAB (R2019a) running on a computer equipped with an Intel<sup>®</sup> Core<sup>TM</sup> i7-6700 CPU @ 3.4 GHz and 16 GB RAM.

The results presented in Table 9 shows that the average encoding time of the proposed JD-TV, DISCOVER [27], and H.264 (intra) [28] ranges from 6.31 s–7.38 s, 31.05 s–60.08 s, and 62.40 s–125.45 s, respectively. It is also observed that at higher subrates, the proposed JD-TV requires more encoding time than lower subrates. In contrast to conventional video codecs, the proposed JD-TV codec takes lesser time to encode the video frames i.e., for all types of video sequence (low, medium, high) the average encoding time of the proposed JD-TV method is significantly lower than the DISCOVER and H.264 (intra) video codecs. For example the encoding time required by JD-TV to encode the Foreman video sequence is 6.94 s which, is 4–5 times less than DISCOVER (36.40 s) and 9–10 times less than the H.264 (Intra) (78.50 s). A similar trend can be observed for other video sequences. Thus, this validates

the fact that the conventional video schemes' encoders are exposed to heavy computation burden and are not suitable for low power real-time applications.

Sequence	Proposed (JD-TV)	DISCOVER	H.264 (Intra)
Hall-Monitor	6.31 s	31.21 s	65.66 s
Mother Daughter	6.39 s	40.65 s	86.58 s
Coast-Guard	6.53 s	31.05 s	62.40 s
Foreman	6.94 s	36.40 s	78.50 s
Mobile Calendar	7.38 s	60.08 s	125.45 s

**Table 9.** Average encoding (execution) time (sec) comparison of proposed CS based JD-TV codec withconventional video codecs (DISCOVER, H.264 (intra)) for various video sequences.

We also observed the decoding complexity (average reconstruction time) of the proposed JD-TV, DISCOVER, and H.264 (intra). The results presented in Table 10 shows that the proposed JD-TV requires more reconstruction time as compared to the conventional codecs, i.e., its decoder has a heavy computational complexity when compared with DISCOVER and H.264 (intra). This validates the fact the CS-based codecs are based on a simple encoder and complex decoder, unlike conventional codecs (DISCOVER and H.264 (intra)). Also, it should also be noted that the results shown are a trade-off between encoding/decoding complexity and rate-distortion performance and has a weak comparability aspect.

**Table 10.** Average decoding (execution) time (sec) comparison of proposed CS based JD-TV codec with conventional video codecs (DISCOVER, H.264 (intra)) for various video sequences.

Sequence	Proposed (JD-TV)	DISCOVER	H.264 (Intra)
Hall-Monitor	7.21 s	1.13 s	0.60 s
Mother Daughter	7.17 s	1.11 s	0.59 s
Coast-Guard	9.22 s	1.27 s	0.75 s
Foreman	9.19 s	1.22 s	0.67 s
Mobile Calendar	9.43 s	1.27 s	0.75 s

Additionally, the average reconstruction time of the proposed JD-TV and other CS reconstruction schemes i.e., MS-residual [12], k-t FOCUSS [13], and MC-BCS-SPL [17], is also shown in Figure 8. Overall, the proposed JD-TV executes about 2–3 times faster than MC-BCS-SPL and kt-Focuss whereas, the proposed JD-TV outperforms MS-residual. This reason is that the proposed scheme incorporates BCS-TV-AL3 (less complex) for initial reconstruction and the simplified process of predicting the non-key frames.



**Figure 8.** Average execution time (sec) comparison of proposed CS based JD-TV codec with other CS codecs for various video sequences.

#### 5. Conclusions

In this paper, a joint decoding (JD) framework is proposed that reduces the redundancies present between video frames at the decoder and reduces the computational burden on the encoder. The proposed scheme makes use of an efficient registration and fusion method to generate side information that helps in the reconstruction of final frame sequences. The simulation results are based on different arrangements with different subrates and GoP sizes (3, 5, 8). The results show the effect of different GoP sizes on the reconstruction quality i.e., smaller GoP = 3 provides improved reconstruction quality compared to larger ones. The comprehensive experimental analysis proves that the proposed JD-TV performs notably better than the independent BCS-TV-AL3 scheme at different subrates for various video sequences having low, moderate and high motion contents. In addition, the proposed scheme outperforms the conventional CS video reconstruction schemes at lower subrates. Further, when compared with conventional video codecs (DISCOVER, H-263, H264) the proposed framework shows notable performance at various bitrates. Further, the coding efficiency in terms of encoding and decoding time of the proposed and conventional video codes are also compared.

The developed scheme can be used in real-time transport safety applications such as driver warning systems for motorbikes or cars, or driver drowsiness detection. Similarly, it can be applied for real-time visual patient monitoring, MRI, patient care and remote sensing.

**Author Contributions:** Conceptualization: M.E.; Methodology: M.E.; Software: M.E. and S.H.A.; Validation: M.E., S.H.A., and K.R.; Formal analysis: M.E., S.S.A.A.; Investigation: M.E. and S.H.A.; Resources: M.E. and K.R.; Data curation: M.E. and S.H.A.; Writing—original draft preparation: M.E.; Writing—review and editing: M.E. and S.S.A.A.; Visualization: M.E., K.R. and S.S.A.A.; Supervision: S.H.A. and K.R.; Project administration: M.E. and S.H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The author would like to acknowledge the technical and administrative support of Sunway University, Malaysia, Iqra University, Pakistan and Universiti Teknologi PETRONAS, Malaysia. This research is partially supported by Iqra University, Pakistan and Ministry of Education, Malaysia under the Higher Institute Centre of Excellence (HICoE) awarded to Center for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Malaysia.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- Puri, R.; Majumdar, A.; Ishwar, P.; Ramchandran, K. Distributed Video Coding in Wireless Sensor Networks. *IEEE Signal Process. Mag.* 2006, 23, 94–106. [CrossRef]
- 2. Ebrahim, M.; Chong, C.W. A Comprehensive Review of Distributed Coding Algorithms for Visual Sensor Network (VSN). *Int. J. Commun. Netw. Inf. Secur. (IJCNIS)* **2014**, *6*, 104–117.
- 3. Donoho, D.L. Compressed Sensing. IEEE Trans. Inf. Theory 2006, 52, 1289–1306. [CrossRef]
- Ebrahim, M.; Chong, C.W.; Adil, S.H.; Raza, K. Block Compressive Sensing (BCS) Based Low Complexity, Energy Efficient Visual Sensor Platform with Joint Multi-Phase Decoder (JMD). Sensors 2019, 19, 2039. [CrossRef] [PubMed]
- Chen, X.; Frossard, P. Joint reconstruction of compressed multi-view images. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 1005–1008.
- Li, X.; Wei, Z.; Xiao, L. Compressed sensing joint reconstruction for multi-view images. *Electron. Lett.* 2010, 46, 1548–1550. [CrossRef]
- Wakin, M.B. A manifold lifting algorithm for multi-view compressive imaging. In Proceedings of the Picture Coding Symposium, Chicago, IL, USA, 6–8 May 2009.
- 8. Ebrahim, M.; Chong, C.W. Multi-view Image Block Compressive Sensing with Multi-phase Joint Decoding for Visual Sensor Network. *ACM Trans. Multimed. Comput. Commun. Appl.* **2015**, *12*, 23.

- 9. Park, J.Y.; Wakin, M.B. A geometric approach to multi-view compressive imaging. *EURASIP* 2012, 37. [CrossRef]
- 10. Marcia, R.; Willet, R. Compressive coded aperture video reconstruction. In Proceedings of the European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland, 25–29 August 2008.
- 11. Park, J.Y.; Wakin, M.B. A multiscale framework for compressive sensing of video. In Proceedings of the Picture Coding Symposium, Chicago, IL, USA, 6–8 May 2009; pp. 1–4.
- Lu, W.; Vaswani, N. Modified compressive sensing for real-time dynamic MR imaging. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 3045–3048. [CrossRef]
- 13. Jung, H.; Sung, K.; Nayak, K.S.; Kim, E.Y.; Ye, J.C. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magn. Reson. Med.* **2009**, *61*, 103–116. [CrossRef] [PubMed]
- Jung, H.; Ye, J.C. Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques. *Int. J. Imaging Syst. Technol.* 2010, 20, 81–98. [CrossRef]
- Trocan, M.; Maugey, T.; Tramel, E.W.; Fowler, J.E.; Popescu, P. Compressed sensing of multi-view images using disparity compensation. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 3345–3348. [CrossRef]
- Chang, K.; Qin, T.; Xu, W.; Men, A. A joint reconstruction algorithm for multi-view compressed imaging. In Proceedings of the 2013 IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, China, 19 May 2013; pp. 221–224. [CrossRef]
- 17. Mun, S.; Fowler, J.E. Residual reconstruction for block-based compressed sensing of video. In Proceedings of the IEEE Data Compression Conference, Snowbird, UT, USA, 29–31 March 2011; pp. 183–192.
- 18. Cen, N.; Guan, Z.; Melodia, T. Joint decoding of independently encoded compressive multi-view video streams. In Proceedings of the Picture Coding Symposium (PCS), San Jose, CA, USA, 8–11 December 2013.
- 19. Trocan, M.; Tramel, E.W.; Fowler, J.E.; Pesquet, B. Compressed sensing recovery of multi-view image and video sequences using signal prediction. *Multimed. Tools Appl.* **2014**, *72*, 95–121. [CrossRef]
- 20. Candes, E.; Romberg, J. Sparsity and incoherence in compressive sampling. *Inverse Probl.* **2007**, *23*, 969–985. [CrossRef]
- Gan, L. Block Compressed Sensing of Natural Images. In Proceedings of the 15th International Conference on Digital Signal Processing, Cardiff, UK, 1–4 July 2007; pp. 403–406.
- 22. Ebrahim, M.; Chong, C.W.; Adil, S.H.; Nawaz, D. A Performance Comparative Analysis of Block Based Compressive Sensing and Line Based Compressive Sensing. *Eng. Technol. Appl. Sci. Res.* **2018**, *8*, 2809–2813.
- 23. Chambolle, A.; Lions, P.L. Image recovery via total variation minimization and related problems. *Numer. Math.* **1997**, *76*, 167–188. [CrossRef]
- 24. Li, C. Compressive Sensing for 3d Data Processing Tasks: Applications, Models and Algorithms. Ph.D. Thesis, Rice University, Houston, TX, USA, 2013. Available online: http://hdl.handle.net/1911/70314 (accessed on 13 March 2020).
- 25. YUV Video Test Sequences, Retrieved 15 January 2015. Available online: http://www.codersvoice.com/a/webbase/video/08/152014/130.html (accessed on 21 April 2020).
- 26. HD Video Sequence Book Arrival, Newspaper, LoveBirds Multiview Video Sequence—Courtesy IRCCyN IVC DIBR. Available online: http://ivc.univ-nantes.fr/en/databases/DIBR\_Videos/ (accessed on 21 April 2020).
- 27. Artigas, X.; Ascenso, J.; Dalai, M.; Klomp, S.; Kubasov, D.; Ouaret, M. The DISCOVER codec: Architecture, techniques and evaluation. In Proceedings of the Picture Coding Symposium, Lisbon, Portugal, 7–9 November 2007.
- International Telecommunication Union (ITU). H.264: Advanced Video Coding for Generic Audiovisual Services. ITU-T Recommendations for h.264, 2005 132-International. Available online: https://www.itu.int/ rec/T-REC-H.264 (accessed on 18 October 2020).

- 29. Telecommunication Union (ITU). H.263: Video Coding for Low Bit Rate Communication. ITU-T Recommendations for h.263. 2005. Available online: https://www.itu.int/rec/T-REC-H.263/ (accessed on 18 October 2020).
- 30. Serge. Bjontegaard Metric Calculation (BD-PSNR). 2020. Available online: https://www.mathworks.com/ matlabcentral/fileexchange/41749-bjontegaard-metric-calculation-bd-psnr (accessed on 21 October 2020).

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).