

# Compressive Sensing Reconstruction for Video: an Adaptive Approach Based on Motion Estimation

Xin Ding, *Student Member, IEEE*, Wei Chen, *Member, IEEE*, and Ian J. Wassell.

**Abstract**—This paper focuses on the problem of causally reconstructing Compressive Sensing (CS) captured video. The state-of-art causal approaches usually assume the signal support is static or changing sufficiently slowly over time, where Magnetic Resonance Imaging (MRI) is widely used as a motivating example. However, such an assumption is too restrictive for many other video applications, where the signal support changes rapidly. In this paper, we propose a framework that combines Motion Estimation (ME), the Kalman Filter (KF) and CS to adapt the reconstruction process to motions in the video so that the slowly-changing assumption on the signal support is relaxed and consequently is more suitable for video reconstruction. Explicit and implicit ME are designed to provide motion aware predictions, upon which a modified KF procedure is applied. Furthermore, three CS algorithms with embedded ME and KF are developed, and theoretical analyses are conducted via reconstruction error upper bounds, to characterize the various factors that affect reconstruction accuracy. Extensive simulations utilizing actual videos are carried out and the superiority of our methods is demonstrated.

**Index Terms**—Compressive sensing; Kalman filter; Video reconstruction; Motion Estimation; Multiscale Reconstruction.

## I. INTRODUCTION

COMPRESSIVE Sensing (CS) [2], [3], which leverages the sparse structure that is present in most practical signals, has been making dramatic improvements on the acquisition, processing and reconstruction of signals in recent years. Compared to the conventional “sensing-then-compressing” signal acquisition framework, CS combines the two stages to directly obtain a compressed version of the original signal using random sampling. Besides, it allows accurate reconstruction from far fewer samples than that required by the Nyquist sampling theory. The development of the CS theory has led to various designs of CS based cameras such as the Rice single pixel camera [4], coded aperture imagers [5], CMOS CS imagers [6] and spectral imagers [7], just to name a few. These designs alleviate the constraint present in a conventional camera that full-frame sensors are required to capture an image. In other words, CS cameras are capable of acquiring an image of higher resolution with the usage of the same number of sensors. To recover the images taken by these CS cameras, various approaches to the CS reconstruction problem have been investigated in the literature [2], [3], [8]–[10].

Xin Ding and Ian J. Wassell are with the Computer Lab, University of Cambridge, UK (e-mail: xd225, ijw24@cam.ac.uk).

Wei Chen is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China, and also with the Computer Lab, University of Cambridge, UK (e-mail: wc253@cam.ac.uk).

This work is supported by EPSRC Research Grant EP/K033700/1; the National Science Foundation of China (61401018, U1334202). (Corresponding author: Wei Chen.) A part of this work has been presented in [1].

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

In addition to the sparse characteristic within an image, more structure can be exploited when it comes to the reconstruction of a video captured by an CS camera in view of the temporal correlations that exist between frames [11]. Ignoring the correlations leads to the most straightforward but least effective reconstruction method, i.e., applying CS reconstruction to each frame independently. To enhance the performance of such methods, the intra-frame correlation is exploited in [12]. An alternative approach that takes advantage of the inter-frame correlation is to treat the whole video as a 3D signal and apply batch CS processing [13]–[15]. However, the batch solution causes delay as it cannot be performed until all the frames forming the 3D signals are acquired. Also, the computational complexity increases significantly due to the high number of dimensions involved. By enforcing a structured correlation model such as the one used in Multiple Measurement Vectors (MMV) [16], it is possible to jointly reconstruct multiple frames with a relatively low computational complexity. Unfortunately, the MMV model assumes that the sparsity pattern of the whole sequence stays the same through time, which is too strict for practical video applications.

Another popular approach to exploit the inter-frame correlation relates to the problem of causal and recursive reconstruction of CS captured videos, i.e., only the previous reconstructed frame and measurements of the current frame are employed in the reconstruction of current frame. In [17], Vaswani proposes a method known as Kalman Filtered Compressed Sensing (KFCS) to recover causally a time-varying sequence. The work is extended in [18] to an actual Magnetic Resonance Imaging (MRI) sequence, which is nearly sparse (compressible), rather than ideally so. Non-KF causal methods are also developed for the noiseless case in [19] and the noisy case in [20]. All these approaches are based on the assumption that the signal support (the indices set of non-zero coefficients in a sparse signal) is slowly changing over time. It has been shown that these techniques are successful in applications where the assumption holds, e.g., MRI. However, this is not true for a large class of videos, wherein the support changes much more rapidly. In Fig. 1, some popular videos and MRI sequences are illustrated along with their Support Change Rates (SCRs). The SCR reflects how dynamically the support is changing and is defined as:  $SCR = (T - 1)^{-1} \sum_{t=2}^T |\mathcal{N}_t|^{-1} (|\mathcal{N}_t - \mathcal{N}_{t-1}| + |\mathcal{N}_{t-1} - \mathcal{N}_t|)$ , where  $T$  is the total number of frames,  $\mathcal{N}_t$  and  $\mathcal{N}_{t-1}$  are the support of the  $t$ -th frame and the  $(t - 1)$ -th frame, respectively. Clearly, the SCRs for these commonly used videos are much higher than those for the MRI sequences. Ignoring the rapid support change present in video will result in performance degradation for the previously mentioned methods.

The motion between frames in a video has encouraged us

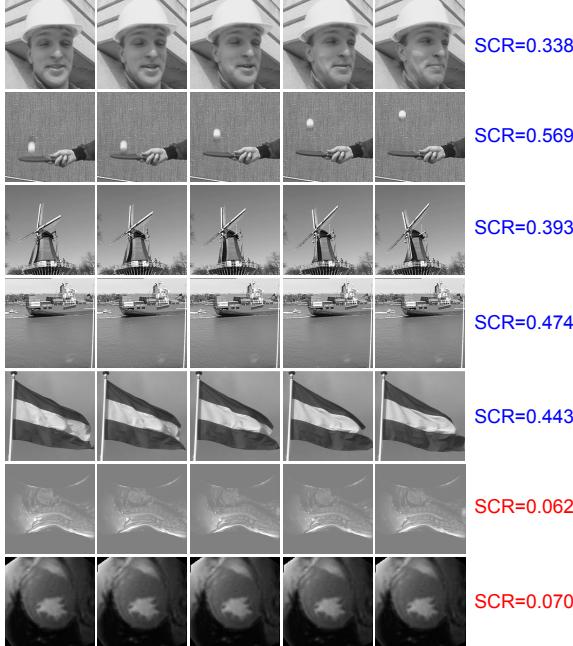


Fig. 1: SCR values for video: foreman, tennis, windmill, boats, flag; and MRI sequences: larynx, cardiac. Each sequence consists 20 frames and a few example frames are demonstrated. A 4 level 2D discrete wavelet transform is employed for sparsifying.

to investigate motion-adaptive reconstruction, thereby generalizing the aforementioned methods. However, in contrast to conventional video applications, the full resolution frames are not available in CS captured video, which makes conventional Motion Estimation (ME) methods infeasible. To overcome this difficulty, various researchers have proposed to in the first instance conduct coarse reconstruction in order to obtain an approximation of the whole video sequence, followed by ME and CS to refine the results [21]–[24]. However, these methods have high computational complexity as they involve two reconstructions, as well as a ME process for each frame. In [1], we propose a ME method that can be directly applied to CS samples when the random sampling strategy is employed. ME for CS samples is also developed in [25] for a block sensing scheme and in [26] for a particular camera design. Rather than using explicit ME, motion information has also been incorporated into CS video reconstruction implicitly via dictionary learning in [27]–[29], where the learning mechanism plays a key role concerning the reconstruction quality. Besides, classification-based methods are developed in [30], [31], which are not fully motion-adaptive as they independently reconstruct the region involving motion without utilizing temporal correlations.

In this paper, we aim to adapt the KFCS idea to account for the actual motions between frames so that it is more appropriate for the reconstruction of practical CS captured video in a noisy environment. Compared to the previous work in the literature, our contributions include:

- A novel framework that allows the adaption of motion when reconstructing CS captured video is proposed. In this framework, a motion aware prediction is first pro-

duced using CS samples, which is then incorporated into the Kalman Filter (KF) processing. Combined with the updated KF estimation process, a CS step is also included in order to improve reconstruction performance.

- Two motion aware prediction approaches are designed using either explicit or implicit Motion Estimation (ME) for particular CS systems. Both methods overcome the difficulty caused by the unavailability of full resolution frames when applying conventional ME to CS videos. Compared to the prediction methods currently in the literature, the explicit scheme has lower computational complexity since it avoids solving an under-determined problem before ME; while the implicit scheme improves their prediction accuracy by including a motion-aware dictionary learning procedure.
- A modification of the original KF is derived so that the motion aware prediction can be incorporated into the filtering process. By updating the prediction using the KF, a more accurate estimation of the frames are produced, which also benefits the following CS step.
- Various CS schemes that exploit the KF estimation are developed for CS video reconstruction. The performance of these schemes are analyzed via theoretical error upper bounds and it is shown that the reconstruction quality is affected by the sensing noise and the accuracy of the motion aware prediction. Experiments conducted using example videos demonstrate the superiority of these schemes.

The rest of the paper is organized as follows. Section II introduces the mathematical models for CS and multiscale CS, and reviews the existing non-motion-adaptive causal CS methods and motion-adaptive methods; in Section III, the proposed approach is overviewed; Section IV provides the details of our framework, that includes the proposed motion aware prediction, the utilization of KF and the design of the CS algorithms; Section V investigates the theoretical performance of the proposed approach; the performance of all the proposed algorithms is evaluated in Section VI and Section VII presents conclusions.

*Notation:* In the rest of this paper, we always use  $\hat{s}$ ,  $\bar{s}$  and  $\hat{s}$  to represent the motion aware prediction, KF estimation and final CS estimation, respectively. Boldface lower-case letters denote vectors, boldface upper-case letters denote matrices and non-boldface letters denote scalars. Calligraphic upper-case letters denote sets and the superscript  $(\cdot)^c$  represents the complement set. By  $(s_t)_{\mathcal{J}}$  we mean the sub-vector of  $s_t$  that contains the elements with index in  $\mathcal{J}$  and  $|\mathcal{J}|$  denotes the cardinality of  $\mathcal{J}$ . The matrix transpose and inverse operations are denoted by the superscripts  $(\cdot)^T$  and  $(\cdot)^{-1}$ , respectively. The superscripts with brackets  $(\cdot)^{(i)}$  denote the coefficient index  $i$ . The  $l_p$  ( $p = 0, 1, 2$ ) norm of vectors is denoted by  $\|\cdot\|_p$  and  $I$  represents an identity matrix. The operator  $\text{diag}[\cdot]$  means forming a diagonal matrix by placing the elements on the main diagonal.  $DWT(\cdot)$  and  $IDWT(\cdot)$  mean conducting a discrete wavelet transform and an inverse discrete wavelet transform, respectively. The operators  $\min(\cdot)$  and  $\max(\cdot)$  represent taking the minimum and maximum value of the contents, respectively.

## II. BACKGROUND AND RELATED WORK

### A. CS and Multiscale CS

Consider a video sequence  $\mathbf{x}_t \in \mathbb{R}^N$  ( $t = 1, \dots, T$ ) where each frame can be represented as:

$$\mathbf{x}_t = \Psi \mathbf{s}_t, \quad (1)$$

where  $\Psi \in \mathbb{R}^{N \times N}$  is a sparsifying basis that can be either pre-defined (e.g., a Discrete Wavelet Transform (DWT)) or adaptively learned [27], [32], and  $\mathbf{s}_t \in \mathbb{R}^N$  is the sparse representation which has only  $K$  ( $K \ll N$ ) non-zero coefficients, i.e.,  $\|\mathbf{s}_t\|_0 = K$ . We assume a CS camera is used for acquiring the video. The measurement vector for each frame can be represented as:

$$\mathbf{y}_t = \Phi \mathbf{x}_t + \mathbf{e}_t = \mathbf{A} \mathbf{s}_t + \mathbf{e}_t, \quad (2)$$

where  $\mathbf{y}_t \in \mathbb{R}^M$  ( $M \ll N$ ) is the  $t$ -th observation vector,  $\Phi \in \mathbb{R}^{M \times N}$  represents the sensing matrix,  $\mathbf{e}_t \in \mathbb{R}^M$  is a noise term and  $\mathbf{A} = \Phi \Psi$  is the equivalent projection matrix. It is conventionally impossible to recover  $\mathbf{s}_t$  or  $\mathbf{x}_t$  from this under-determined system. However, CS asserts that when  $\mathbf{A}$  obeys the Restricted Isometry Property (RIP) [33], this ill-posed inverse problem can be solved with an overwhelming probability of success.

*Definition 1:* A matrix  $\mathbf{A}$  satisfies the RIP of order  $K$  with a the Restricted Isometry Constant (RIC)  $\delta_K$  being the smallest number such that

$$(1 - \delta_K) \|\mathbf{s}\|_2^2 \leq \|\mathbf{As}\|_2^2 \leq (1 + \delta_K) \|\mathbf{s}\|_2^2 \quad (3)$$

holds for all  $\mathbf{s}$  with  $\|\mathbf{s}\|_0 \leq K$ . ■

The RIP can be equivalently described by stating that all subsets of  $K$  columns taken from  $\mathbf{A}$  are nearly orthogonal. This implies that all pairwise distances between  $K$  sparse signals are well preserved in the projection space. An i.i.d. Gaussian matrix or a Bernoulli matrix is commonly used as the sensing matrix  $\Phi$  because of their universal incoherence, in which case the RIP condition holds for the matrix  $\mathbf{A}$  with a high probability regardless of the choice of  $\Psi$  [34].

The CS reconstruction in the noisy case for individual images can be formulated as a Basis Pursuit De-Noising (BPDN) problem:

$$\min_{\mathbf{s}_t} \|\mathbf{s}_t\|_1 \quad s.t. \|\mathbf{As}_t - \mathbf{y}_t\|_2 \leq \varepsilon, \quad (4)$$

where  $\varepsilon$  is an estimate of the noise level. It has been proved that when the RIP is satisfied, the reconstruction accuracy of (4) can be bounded by using the following theorem [33].

*Theorem 1:* Assume that  $\delta_{2K} < \sqrt{2} - 1$  and  $\|\mathbf{e}_t\|_2 \leq \varepsilon$ . Then the solution  $\hat{\mathbf{s}}_t$  to (4) obeys

$$\|\hat{\mathbf{s}}_t - \mathbf{s}_t\|_2 \leq C_0 K^{-1/2} \|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1 + C_1 \varepsilon \quad (5)$$

where  $C_0 = \frac{2+(2\sqrt{2}-2)\delta_{2K}}{1-(\sqrt{2}+1)\delta_{2K}}$ ,  $C_1 = \frac{4\sqrt{1+\delta_{2K}}}{1-(\sqrt{2}+1)\delta_{2K}}$ ,  $\delta_{2K}$  is the RIC of matrix  $\mathbf{A}$ ,  $(\mathbf{s}_t)_K$  is an approximation of  $\mathbf{s}_t$  with all but the  $K$  largest entries set to zero. ■

When some prior information of the sparsity structure is available, CS can benefit by designing the sensing strategy rather than employing purely random sampling. A well known example, namely multiscale CS [35] offers a non-uniform sampling scheme based on the coarse-to-fine structure presented in the wavelet transforms of signals, e.g., [36]–[39].

In multiscale CS, full-rate linear sampling is applied to the coarse scale wavelet coefficients, and the fine scale coefficients are compressively sampled at various lower rates

depending on the scale levels. We denote the  $i$ -th scale wavelet coefficients as  $\mathbf{s}_t^i \in \mathbb{R}^{N_i}$  and define the following two operators:  $\mathbf{A}^i \in \mathbb{R}^{M_i \times N_i}$  is the equivalent projection matrix that functions on the  $i$ -th wavelet scale and  $\mathbf{R}^i \in \mathbb{R}^{N_i \times N}$  represents a linear restriction operator [21] that only outputs the  $i$ -th scale coefficients and omits the others. The multiscale sensing process can then be represented as follows:

$$\mathbf{y}_t^i = \mathbf{A}^i \mathbf{s}_t^i + \mathbf{e}_t^i = \mathbf{A}^i \mathbf{R}^i \mathbf{s}_t + \mathbf{e}_t^i = \Phi^i \mathbf{x}_t + \mathbf{e}_t^i, \quad (6)$$

where the superscript  $i$  denotes the  $i$ -th scale ( $i = 1, 2, \dots$ ),  $M_i = N_i$  for the specified coarse scales  $i$  and  $M_i \ll N_i$  for the others,  $\Phi^i = \mathbf{A}^i \mathbf{R}^i \Psi^{-1}$  is the sensing matrix that is actually applied to the pixel-domain signal and  $\Psi \in \mathbb{R}^{N \times N}$  is the DWT sparsifying basis. The multiscale CS reconstruction is then a combination of linear reconstruction at the coarse scales and non-linear CS reconstruction at the fine scales.

Clearly, the multiscale scheme benefits from the non-uniform structure of the sparsity of an image. It takes a larger fraction of measurements at the less sparse scales and reduces sensing energy at the more sparse scales. Besides, due to the scale-separate reconstruction in multiscale CS, the signal dimensions that we can deal with are enlarged. Another popular method for dealing with high dimensional signals is a block-based approach, e.g., dividing images into small patches and sensing then reconstructing each patch separately. However, all the previously mentioned methods do not consider any temporal correlations that would exist when applied to video reconstruction, which motivates extensions and improvements to these methods.

### B. Related Work

To exploit temporal correlations in a sequence of images and help the reconstruction of CS captured video, causal reconstruction methods have been developed. As mentioned, in [17], [18], Vaswani proposes a causal method, namely KFCS, to reconstruct a series of CS captured images. It is assumed that the support of the current image is the same as that of the previous one and a KF process is carried out to get an estimate of current signal,  $\bar{\mathbf{s}}_t$ . Then CS is applied on the KF residual as follows:

$$\min_{\mathbf{s}_{t,res}} \|\mathbf{s}_{t,res}\|_1, \quad s.t. \|\mathbf{y}_{t,res} - \mathbf{A} \mathbf{s}_{t,res}\|_2 \leq \varepsilon, \quad (7)$$

where  $\mathbf{y}_{t,res} = \mathbf{y}_t - \mathbf{A} \bar{\mathbf{s}}_t$ . By using the support of the new estimation of the current frame, i.e.,  $\hat{\mathbf{s}}_t = \bar{\mathbf{s}}_t + \hat{\mathbf{s}}_{t,res}$ , a second KF process is applied to enhance the reconstruction quality.

Under the same assumption, a non-KF causal method, namely modified CS, is proposed in [19], [20]. The signal support of the previous frame is used in the CS reconstruction of the current frame by solving the following optimization problem:

$$\min_{\mathbf{s}_t} \|(\mathbf{s}_t)_{\mathcal{J}_t}\|_1, \quad s.t. \|\mathbf{y}_t - \mathbf{A} \mathbf{s}_t\|_2 \leq \varepsilon, \quad (8)$$

where  $\mathcal{J}_t$  is the support of  $\hat{\mathbf{s}}_{t-1}$ . If the support change between frames is small, solving the optimization problem in (8) leads to an improved reconstruction quality. Note that the optimization problem in (8) can be seen as a specialized form of a weighted  $l_1$  problem:

$$\min_{\mathbf{s}_t} \|\mathbf{W}_t \mathbf{s}_t\|_1 \quad s.t. \|\mathbf{y}_t - \Phi_t \Psi \mathbf{s}_t\|_2 \leq \varepsilon, \quad (9)$$

where  $\mathbf{W}_t = \text{diag}[w^{(1)}, w^{(2)}, \dots, w^{(N)}]$  and  $w^{(i)}$  can be designed using different schemes. If  $w^{(i)} = 0$  when  $i \in \mathcal{J}_t$  and

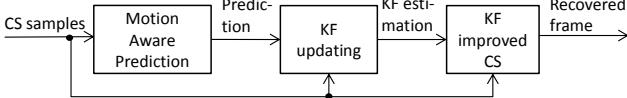


Fig. 2: Proposed framework for motion adaptive reconstruction.

$w^{(i)} = 1$  when  $i \in \mathcal{J}_t^c$ , the weighted  $l_1$  minimization problem is equivalent to (8). The optimal reconstruction accuracy of the weighted minimization is achieved if the weights are equal to the inverse of the true signal magnitudes, that in practice however are unknown.

These approaches are more effective than conventional CS due to their temporal recursive nature; however, they perform poorly when their assumptions are broken, i.e., the support changes rapidly between frames. In [21]–[24], motions between frames are incorporated into the reconstruction process to generalize the use of temporal correlations. In order to take advantage of available mature ME techniques, they all carry out the CS reconstructions first, based on which the ME and Motion Compensation (MC) are conducted to update the reconstruction and help with a second CS recovery. In particular in [22], temporal regularization terms are utilized in its second CS recovery, i.e.,  $\|\Psi^{-1}(\mathbf{F}_{t-1}\mathbf{x}_{t-1} - \mathbf{x}_t)\|_1$  and  $\|\Psi^{-1}(\mathbf{B}_{t+1}\mathbf{x}_{t+1} - \mathbf{x}_t)\|_1$ , where  $\mathbf{F}$  and  $\mathbf{B}$  are forward and backward motion operators, respectively; and in order to reduce the complexity, in [23], a weighted  $l_2$  minimization is designed to update the prediction of ME and MC.

In [27]–[29], ME is involved implicitly by employing a motion aware dictionary in the CS reconstruction. In these approaches, the dictionary is learned from adjacent frames, which normally results in a more sparse representation of the current frame than if a fixed dictionary such as DWT is employed. In particular, in the most recent work [27], for each block of each frame, a Karhunen-Loëve basis is learned, which has been shown to be beneficial for CS video reconstruction.

### III. METHOD OVERVIEW AND PROBLEM FORMULATION

In order to relax the assumption of a slowly changing support pattern in the existing causal reconstruction methods, we propose an adaptive framework that combines the steps of motion aware prediction, KF updating and an improved CS reconstruction, as shown in Fig. 2. It proceeds as follows:

- 1) Conduct a motion aware prediction step. We propose two types of approach for this step: prediction by explicit ME; prediction by motion aware reconstruction using a learned dictionary, i.e., implicit ME.
- 2) Based on the predicted frame obtained in the previous step, a KF process is carried out to update the estimation using the current measurement.
- 3) CS reconstruction is conducted in the final step where the KF estimation is exploited to enhance the reconstruction accuracy.

Overall, our method can be represented as an optimization problem as follows:

$$\begin{aligned} \min_{\mathbf{s}_t} & \|\mathbf{W}_t(\mathbf{s}_t - \bar{\mathbf{s}}_t)\|_1, \\ \text{s.t.} & \|\mathbf{y}_t - \mathbf{A}\mathbf{s}_t\|_2 \leq \varepsilon, \end{aligned}$$

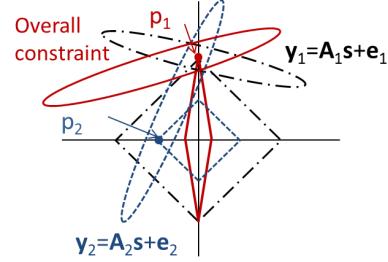


Fig. 3: Geometrical explanation of the advantage of the proposed approach.

$$\begin{aligned} \bar{\mathbf{s}}_t &= f_1(\check{\mathbf{s}}_t, \mathbf{y}_t, \mathbf{A}), \\ \check{\mathbf{s}}_t &= f_2(\hat{\mathcal{S}}_{nb}, \mathbf{y}_t, \mathbf{A}), \end{aligned} \quad (10)$$

where  $\mathbf{W}_t$  is a weight matrix;  $\hat{\mathcal{S}}_{nb}$  represents the set of the reconstructed neighbor frames that for example could be the previous frame or several frames according to the specific scheme employed;  $f_1$  denotes the process of KF updating and  $f_2$  represents the prediction process utilizing explicit or implicit ME.

The advantage of our method can be explained geometrically in Fig. 3. The black dot dash diamond and the black dot dash ellipse  $y_1 = \mathbf{A}_1 s + \mathbf{e}_1$  demonstrate the  $l_1$  pursuit process of conventional CS. In this case, there is a unique solution  $p_1$ . However, when the sensing matrix does not obey the RIP, as denoted by  $\mathbf{A}_2$ , we cannot guarantee to solve the problem with the blue ellipse  $y_2 = \mathbf{A}_2 s + \mathbf{e}_1$ , because there exists a solution  $p_2$  that has a smaller  $l_1$  norm than  $p_1$ . In this case, as our method in (10) has additional constraints other than the measurements, so the overall constraint is modified as shown by the red ellipse. Consequently solution  $p_2$  is avoided and solution  $p_1$  can still be obtained. In addition, when  $\mathbf{W}_t$  is designed as a non-identity matrix, the  $l_1$  ball can be altered in shape as shown by the red thin diamond, which also helps to guarantee the correct solution. Besides, due to the involvement of the motion aware prediction step, our method has better constraints compared to that of assuming  $\check{\mathbf{s}}_t = \hat{\mathbf{s}}_{t-1}$ , which will lead to a more accurate estimation.

### IV. AN ADAPTIVE APPROACH BASED ON MOTION ESTIMATION

In this section, we elaborate the proposed adaptive approach in Fig. 2, including the detailed steps and implementation for each stage and the algorithms summarizing the proposed methods.

#### A. Motion Aware Prediction

ME is a core step in traditional video coding and many ME techniques have been developed [40]. However, the traditional ME methods cannot be directly applied in CS videos, as the full resolution frames are not available in this case. Under specific sensing designs, we propose two approaches utilizing explicit and implicit ME, respectively, to deal with this difficulty. Both the proposed approaches can provide a motion aware prediction for the desired video sequence, which is then refined by the following KF and CS steps.

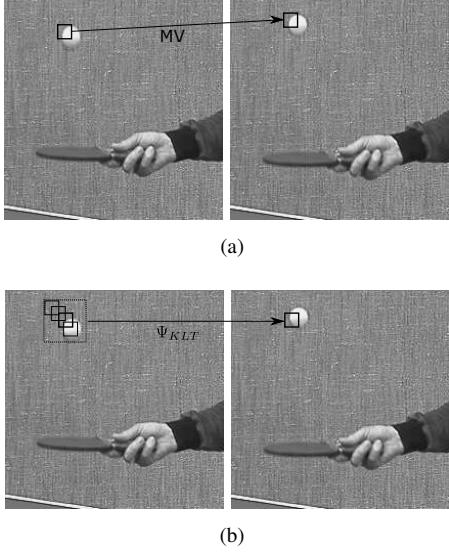


Fig. 4: Motion aware prediction procedure using (a) explicit ME/MC; (b) implicit ME.

As shown in Figure 4(a), in the explicit scheme, the prediction is obtained by moving the blocks in the previous frame according to the Motion Vector (MV) that is derived using the proposed ME. While in the implicit scheme, as shown by Figure 4(b), the prediction is obtained by reconstructing its representation in a dictionary that is generated based on the neighbor blocks in the neighbor frame, e.g., a Karhunen-Loève basis  $\Psi_{KLT}$  [27]. To improve the prediction accuracy, several neighboring frames can be employed for generating the dictionary.

1) *Prediction by Explicit ME/MC:* Our explicit ME method is based on the measurements obtained using the multiscale CS strategy [35] directly, rather than solving an under-determined problem first to get approximations of the frames as is performed [21]–[24]. Suppose the previously reconstructed reference frame and the current frame are  $\mathbf{x}_{ref}$  and  $\mathbf{x}_{cur}$ , respectively, and their wavelet representations are  $\mathbf{s}_{ref}$  and  $\mathbf{s}_{cur}$ . Following the setting in the multiscale CS approach [35], we assume that for all  $l$  scales, the scales 1 to  $j$  are fully sampled and  $j+1$  to  $l$  are compressively sampled. Using the measurements  $\mathbf{y}_t^i$  for each scale  $i$  ( $i = 1 \dots l$ ), the following steps are conducted:

- Step 1: linearly reconstruct the fully sampled scales  $\tilde{\mathbf{s}}_{cur}^i$  ( $i = 1 \dots j$ );
- Step 2: set  $\tilde{\mathbf{s}}_{cur}^i$  ( $i = j+1 \dots l$ ) to zero and do the IDWT, i.e.,  $\tilde{\mathbf{x}}_{cur} = IDWT(\tilde{\mathbf{s}}_{cur})$ ;
- Step 3: conduct ME and MC between  $\tilde{\mathbf{x}}_{cur}$  and  $\mathbf{x}_{ref}$ , and output the DWT of the MC result  $\check{\mathbf{s}}_{cur} = DWT(\tilde{\mathbf{x}}_{cur})$  to the following steps.

Using the predictions  $\check{\mathbf{s}}_{cur}^i$  ( $i = 1 \dots l$ ), the KF and CS steps are then carried out for each scale individually, which will be detailed later. The above prediction approach can also be considered as producing a low-pass filtered approximation of the current frame, based on which the ME/MC processes are carried out. As the reconstructions involved are utilizing full-rate samples, it is much less complex than conducting coarse CS reconstructions as is done in some other methods [21]–

[24].

To improve the accuracy of the prediction, one can choose to serially reconstruct each scale and repeat the previously described prediction process during each reconstruction. When doing so, all the previously reconstructed scales are involved into the ME step, i.e., when reconstructing scale  $\nu$ , the coefficients in  $\tilde{\mathbf{s}}_{cur}^i$  ( $i = 1 \dots \nu - 1$ ) have been reconstructed and only  $\tilde{\mathbf{s}}_{cur}^i$  ( $i = \nu \dots l$ ) are set to zero. In this way, several rounds of ME and MC are carried out to provide a better prediction at the expense of more computations and a longer reconstruction time. One can design the specific strategy of conducting the prediction procedure according to the specific requirements, e.g., do ME/MC for every two scales to trade off between accuracy and complexity. Note that the proposed approach allows utilization of all the available ME methods, e.g., block matching, optical flow and phase-based schemes.

2) *Prediction by Implicit ME:* As verified in [27]–[29], instead of using a fixed sparsifying basis, e.g., a DWT basis, using neighbor frames to learn a motion-aware dictionary can benefit the reconstruction of CS captured video. We thus propose to employ the CS reconstruction result based on the use of such dictionaries as the prediction in our framework. In this way, the motion information is involved implicitly via the dictionary and no explicit ME/MC is required.

For the block-based dictionary learning methods, it is required that we both take measurements and perform reconstruction in a block-based manner. Suppose the video frame  $\mathbf{x}_t$  is sampled using a block-diagonal sensing matrix with  $l$  component-matrices, i.e.,  $\Phi^i$  ( $i = 1 \dots l$ ). Equivalently, each block of the frame, i.e.,  $\mathbf{x}_t^i = \Psi_t^i \mathbf{s}_t^i$ , is measured as:  $\mathbf{y}_t^i = \Phi^i \Psi_t^i \mathbf{s}_t^i + \mathbf{e}_t^i$ , where the dictionary  $\Psi_t^i$  is learned based on the blocks in adjacent frames. Then  $\mathbf{s}_t^i$  ( $i = 1 \dots l$ ) can be recovered using CS algorithms, which forms the predicted frame. For instance, we can employ the order 2 method of [27] to generate the prediction of  $\mathbf{x}_t$  using the following steps:

- Step 1: for each block  $\mathbf{x}_t^i$  of the current frame, extract neighbor blocks of the same size as  $\mathbf{x}_t^i$ , denoted by  $\mathbf{d}_t^j$  ( $j = 1 \dots b$ ), from a window of a square region centered at  $\mathbf{x}_{t-1}^i$  and  $\mathbf{x}_{t-2}^i$ , respectively, by carrying out one-pixel shift in all directions;
- Step 2: estimate the correlation matrix:  $\Theta_t^i = \frac{1}{b} \sum_{j=1}^b \mathbf{d}_t^j (\mathbf{d}_t^j)^T$ ;
- Step 3: form the Karhunen-Loève basis for  $\mathbf{x}_t^i$  using the eigenvectors of  $\Theta_t^i$ , i.e.,  $\Psi_{t,KLT}^i = \Gamma$ , where  $\Gamma$  is the matrix with columns as the eigenvectors of  $\Theta_t^i$ ;
- Step 4: predict  $\mathbf{x}_t^i$  by  $\check{\mathbf{x}}_t^i = \Psi_{t,KLT}^i \tilde{\mathbf{s}}_t^i$ , where  $\tilde{\mathbf{s}}_t^i = \arg \min \|\mathbf{y}_t^i - \Phi^i \Psi_t^i \mathbf{s}_t^i\|_2^2 / 2 + \eta \|\mathbf{s}_t^i\|_1$ , and do DWT:  $\check{\mathbf{s}}_t^i = DWT(\check{\mathbf{x}}_t^i)$ .

For initialization, a DWT basis is utilized to reconstruct the blocks in  $\mathbf{x}_1$  and an order 1 KLT basis is learned to reconstruct the blocks in  $\mathbf{x}_2$ . Note that the order number denotes the number of previous frames that are used to generate the dictionary. From the previously described procedure, we can see that even without explicit ME, the motion between frames is still included via the dictionary, which produces the motion-aware prediction that we desire. Using the predictions  $\check{\mathbf{s}}_t^i$  ( $i = 1 \dots l$ ),

the KF and CS steps are carried out for each block individually to refine the results.

### B. ME Enhanced KF

The KF [41] is one of the most popular methods for dealing with linear dynamic systems. Employing KF in video reconstruction consists of two stages: prediction and update. The prediction is conducted by:  $\check{s}_t = \mathbf{F}_t \hat{s}_{t-1}$ , where  $\mathbf{F}_t$  is a state transition matrix that will be needed when calculating the Kalman gain. However, due to the involvement of the motion-aware prediction stage in our framework,  $\check{s}_t$  has been obtained without needing  $\mathbf{F}_t$ . In addition, the prediction step in our framework has involved the current measurement  $\mathbf{y}_t$ , which conventionally is only being used in the KF update stage. Because of these changes in the prediction model, we need to modify the KF update rules so that the motion aware prediction can be exploited to enhance the reconstruction accuracy in our framework.

Instead of modeling the state transition via multiplication by  $\mathbf{F}_t$ , we consider it as an additive correction as:

$$\mathbf{s}_t = \mathbf{s}_{t-1} + \varrho(\mathcal{S}_{nb}, \mathbf{y}_t, \mathbf{A}) + \mathbf{q}_t, \quad (11)$$

where  $\mathcal{S}_{nb}$  represents the set of the neighbor frames that are used in the motion aware prediction stage,  $\varrho$  is a non-linear additive correction function that relates to the predictable motion using  $\mathcal{S}_{nb}$  and  $\mathbf{y}_t$ , and  $\mathbf{q}_t$  represents the process noise. Then in the proposed framework, we have

$$\check{s}_t = \bar{s}_{t-1} + \varrho(\hat{\mathcal{S}}_{nb}, \mathbf{y}_t, \mathbf{A}), \quad (12)$$

where  $\hat{\mathcal{S}}_{nb}$  denotes the set of the previously reconstructed neighbor frames and the correction term  $\varrho(\hat{\mathcal{S}}_{nb}, \mathbf{y}_t, \mathbf{A})$  is obtained based on  $\hat{\mathcal{S}}_{nb}$  and  $\mathbf{y}_t$ .

Owing to the imperfectly reconstructed frames in  $\hat{\mathcal{S}}_{nb}$ , the additive correction term will contain error, i.e.,  $\mathbf{g}_t = \varrho(\mathcal{S}_{nb}, \mathbf{y}_t, \mathbf{A}) - \varrho(\hat{\mathcal{S}}_{nb}, \mathbf{y}_t, \mathbf{A})$ . Consider the process noise  $\mathbf{q}_t$ , the additive correction error  $\mathbf{g}_t$  and the state estimation error  $(\mathbf{s}_{t-1} - \bar{s}_{t-1})$  are uncorrelated, based on the model in (11) and (12), we derive the following update rules for the ME enhanced KF using a similar procedure to that in [41] for the conventional KF:

$$\check{\mathbf{P}}_t = \mathbf{P}_{t-1} + \mathbf{Q} + \mathbf{G} \quad (13)$$

$$\mathbf{K}_t = \check{\mathbf{P}}_t \mathbf{A}^T [\mathbf{A} \check{\mathbf{P}}_t \mathbf{A}^T + \mathbf{R}_t]^{-1} \quad (14)$$

$$\bar{s}_t = \check{s}_t + \mathbf{K}_t (\mathbf{y}_t - \mathbf{A} \check{s}_t) \quad (15)$$

$$\mathbf{P}_t = \check{\mathbf{P}}_t - \mathbf{K}_t \mathbf{A} \check{\mathbf{P}}_t, \quad (16)$$

where  $\mathbf{P}$  is the KF error covariance matrix,  $\check{\mathbf{P}}$  is the prediction error covariance matrix,  $\mathbf{Q}$  and  $\mathbf{G}$  are the covariance matrices associated with  $\mathbf{q}_t$  and  $\mathbf{g}_t$ , respectively,  $\mathbf{K}$  is the Kalman gain,  $\mathbf{R} = \sigma_{obs}^2 \mathbf{I}$  is the measurement noise covariance and  $\bar{s}$  is the KF estimation. The derivation of these modified KF equations is given in Appendix A.

In the KF stage of our framework, the modified KF equations are performed scale by scale or block by block for the explicit or implicit ME methods, respectively. Note that in (14), the computational complexity of the matrix inversion operation is  $\mathcal{O}(M^3)$ , which is lower than  $\mathcal{O}(M^2 N^{1.5})$ , i.e., the complexity of the  $l_1$  minimization procedure [9], since  $M < N$ . Thus the use of KF does not increase the order

of computational complexity in comparison to traditional CS reconstruction.

Now we discuss the issue of how to determine  $\mathbf{Q}$  and  $\mathbf{G}$ . Unfortunately, the error of the CS based motion aware prediction does not have a closed form expression, owing to the nonlinear process of the CS reconstruction and motion estimation. In view of the fact that  $\mathbf{g}_t$  is mainly induced by the estimation error of the frames in  $\hat{\mathcal{S}}_{nb}$ , we approximate the matrix  $\mathbf{G}$  using an estimation of the mean square error of the oracle least square (LS) estimator, which represents the best achievable performance for any unbiased CS estimator [42]. Specifically, for each scale or block  $i$ , we define  $\mathbf{G}^i = \sigma_1^2 \mathbf{I}$ , where  $\sigma_1^2$  is estimated as following [42]:

$$\sigma_1^2 = \frac{|\mathcal{J}_1|}{N} \sigma_{obs}^2 \text{Tr}\{[(\mathbf{A}^i)^T \mathbf{A}^i]^{-1}\}, \quad (17)$$

in which  $|\mathcal{J}_1|$  is an average estimation of the support cardinality for the scale or block  $i$  in the DWT basis  $\Psi$ . Then for  $\mathbf{Q}$ , as it relates closely to the prediction scheme, we utilize different strategies for the explicit and implicit ME cases. For the explicit ME scheme,  $\mathbf{Q}$  can be learned using training sequences by extending the algorithm proposed in [18] as follows:

- Step 1: for a training sequence  $\mathbf{s}_t$  ( $t = 1 \dots T$ ), keep the coefficients at the specified coarse scales, i.e., those that will be fully sampled, and set the rest to zero, where the result is denoted by  $\check{s}_t$  ( $t = 1 \dots T$ );
- Step 2: apply IDWT to both  $\mathbf{s}_t$  and  $\check{s}_t$  ( $t = 1 \dots T$ ), carry out ME/MC between every pair of  $\mathbf{x}_{t-1}$  and  $\check{\mathbf{x}}_t$ , and apply the DWT to the MC result, i.e.,  $\mathbf{s}_{t,MC} = \text{DWT}(\mathbf{x}_{t,MC})$ ;
- Step 3: for each scale in both  $\mathbf{s}_t$  and  $\mathbf{s}_{t,MC}$ , ( $t = 1 \dots T$ ), select a threshold using the method described in [18] and set all coefficients below it to zero;
- Step 4: calculate a  $\mathbf{Q}$  for each scale  $i$  as:  $\mathbf{Q}^i = \sigma_2^2 \mathbf{I}$ , where  $\sigma_2^2 = \frac{1}{\sum_{j \in \mathcal{H}(|\nu_j|)} \sum_j \sum_t (\mathbf{s}_t^{(j)} - \mathbf{s}_{t,MC}^{(j)})^2}$ ,  $\nu_j := \{t : \mathbf{s}_t^{(j)} - \mathbf{s}_{t,MC}^{(j)} \neq 0\}$  and  $\mathcal{H}$  is the set of indices in the scale.

For the implicit ME scheme, since the prediction is obtained based on a CS reconstruction, we employ a similar strategy to that for determining  $\mathbf{G}$ . That is, for each block  $i$ , we approximate  $\mathbf{Q}$  as  $\mathbf{Q}^i = \sigma_2^2 \mathbf{I}$ , in which

$$\sigma_2^2 = \frac{|\mathcal{J}_2|}{TN} \sigma_{obs}^2 \sum_t \text{Tr}\{[(\Psi_{t,KLT}^i)^T (\Phi^i)^T \Phi^i \Psi_{t,KLT}^i]^{-1}\}, \quad (18)$$

where  $|\mathcal{J}_2|$  is an average estimation of the support cardinality for the block  $i$  in the KLT basis  $\Psi_{t,KLT}$ .

### C. KF Improved CS

Due to the involvement of the motion adaptive prediction and KF updating,  $\bar{s}_t$  tends to be much closer to  $s_t$  than is the case for the non-motion-adaptive methods. We thus anticipate a sparse residual signal  $\mathbf{s}_{t,res}$ , which can then be recovered using various algorithms. We now consider applying three of the schemes mentioned previously in Section II-B for the KF improved CS stage:

- Step 1: derive the residue in the measurement domain:  $\mathbf{y}_{t,res} = \mathbf{y}_t - \mathbf{A} \bar{s}_t$ ;

**Algorithm 1** ExMA

**Input:**  $\mathbf{y}_t^i, (t = 0 \dots T, i = 1 \dots l)$ ,  $\mathbf{A}^i, (i = 1 \dots l)$ ,  $\mathbf{Q}^i, (i = 1 \dots l)$ ,  $\mathbf{G}^i, (i = 1 \dots l)$ ,  $\mathbf{P}_{0|0}^i, (i = 1 \dots l)$ ,  $\sigma_{obs}$ ,  $\varepsilon$ ,  $\lambda$ .

**Output:**  $\hat{\mathbf{x}}_t (t = 0 \dots T)$ .

**Initialization:** Get  $\hat{\mathbf{s}}_0$  by multiscale CS reconstruction.

- 1: **for**  $t = 1 : T$  **do**
- 2:   Linearly reconstruct fully sampled scales ( $i = 1 \dots v$ ).
- 3:   Predict scales ( $v + 1 \dots l$ ) by the explicit ME/MC steps.
- 4:   **for**  $i = v + 1 \dots l$  **do**
- 5:     **if** Updated ME/MC required **then**
- 6:       Repeat explicit ME/MC prediction with scales  $1 \dots v$  known.
- 7:     **end if**
- 8:     Apply ME enhanced KF steps (13)-(16) for scale  $i$ .
- 9:     Conduct KF improved CS stage for scale  $i$ .
- 10:  **end for**
- 11:   $\hat{\mathbf{x}}_t = \text{IDWT}(\hat{\mathbf{s}}_t)$ .
- 12: **end for**

**Algorithm 2** ImMA

**Input:**  $\mathbf{y}_t^i, (t = 0 \dots T, i = 1 \dots l)$ ,  $\Phi^i, (i = 1 \dots l)$ ,  $\Psi^i, (i = 1 \dots l)$ ,  $\mathbf{Q}^i, (i = 1 \dots l)$ ,  $\mathbf{G}^i, (i = 1 \dots l)$ ,  $\mathbf{P}_{0|0}^i, (i = 1 \dots l)$ ,  $\sigma_{obs}$ ,  $\eta$ ,  $\varepsilon$ ,  $\lambda$ .

**Output:**  $\hat{\mathbf{x}}_t (t = 0 \dots T)$ .

**Initialization:** Get  $\hat{\mathbf{x}}_0$  and  $\hat{\mathbf{x}}_1$  by CS reconstruction using a DWT dictionary and order 1 KLT dictionary, respectively.

- 1: **for**  $t = 2 : T$  **do**
- 2:   **for**  $i = 1 \dots l$  **do**
- 3:     Predict  $\check{\mathbf{s}}_t^i$  by the implicit ME steps.
- 4:     Apply ME enhanced KF steps (13)-(16) for block  $i$ .
- 5:     Conduct KF improved CS stage for block  $i$ .
- 6:   **end for**
- 7:   $\hat{\mathbf{x}}_t = \text{IDWT}(\hat{\mathbf{s}}_t)$ .
- 8: **end for**

Step 2: utilize one of the following schemes:

scheme 1:  $\min_{\mathbf{s}_{t,res}} \|\mathbf{s}_{t,res}\|_1$ , s.t.  $\|\mathbf{y}_{t,res} - \mathbf{A}\mathbf{s}_{t,res}\|_2 \leq \varepsilon$ ;

scheme 2:  $\min_{\mathbf{s}_{t,res}} \|(\mathbf{s}_{t,res})_{\mathcal{J}_t^c}\|_1$ , s.t.  $\|\mathbf{y}_{t,res} - \mathbf{A}\mathbf{s}_{t,res}\|_2 \leq \varepsilon$ , where  $\mathcal{J}_t$  is the estimated support of  $\bar{\mathbf{s}}_t$ ;

scheme 3:  $\min_{\mathbf{s}_{t,res}} \|\mathbf{W}_t \mathbf{s}_{t,res}\|_1$  s.t.  $\|\mathbf{y}_{t,res} - \mathbf{A}\mathbf{s}_{t,res}\|_2 \leq \varepsilon$ , where  $\mathbf{W}_t = \text{diag}[w^{(1)}, w^{(2)}, \dots, w^{(N)}]$ ,  $w^{(i)} = (|\bar{\mathbf{s}}_t^{(i)}| + \lambda)^{-1}$ ;

Step 3: get final estimation:  $\hat{\mathbf{s}}_t = \bar{\mathbf{s}}_t + \hat{\mathbf{s}}_{t,res}$ .

We observe that scheme 1 is simply a CS reconstruction on the residue; scheme 2 is a modified CS based method, in which the 99%-energy support estimation [19] is employed, i.e.,  $\mathcal{J}_t := \{i : i \in [1, N], \mathbf{s}_i^2 > \zeta\}$ , where  $\zeta$  is the largest real constant for which  $N$  contains 99% of the signal energy; scheme 3 is a weighted CS scheme, where we employ the reciprocal of the KF result as the weights with a stability parameter  $\lambda$  to avoid infinities. These schemes are applied to the scales or blocks individually according to the settings in the previous stages.

The proposed algorithms for motion adaptive reconstruction using explicit ME prediction (ExMA) and implicit ME prediction (ImMA) are summarized in Algorithm 1 and 2, respectively. In the rest of this paper, we identify the various KF improved CS schemes as described previously as ExMA-1,

ExMA-2, ExMA-3, ImMA-1, ImMA-2, ImMA-3, respectively.

## V. THEORETICAL ANALYSIS

In this section, we provide a theoretical analysis for the reconstruction accuracy of the proposed approach. Our goal is to investigate the factors that affect the performance of the proposed approach and provide more theoretical insights into the proposed framework. In order to provide a more general view, the results all consider a single reconstruction.

## A. Error Upper Bounds

As the reconstruction error of CS does not have a closed form representation, we characterize the reconstruction accuracy via error upper bounds, which is a widely used approach for investigating the performance of CS algorithms. The main results are given in the following theorem.

**Theorem 2:** Let  $\mathbf{r}_t = \mathbf{s}_{t,res} - (\mathbf{s}_{t,res})_{\mathcal{J}_t}$ ;  $w_{min}$  and  $w_{max}$  denote the minimum and maximum value of  $w^{(j)}$  ( $j \in \mathcal{T}_0^c$ ), respectively, where  $\mathcal{T}_0$  is the support of  $\mathbf{s}_{t,res}$ ;  $P_1 = w_{min}^{-1} \sum_{i \in \mathcal{T}_0} (w^{(i)})^2$  and  $P_2 = w_{min}^{-1} w_{max}$ . Assume that at each  $t$ ,  $|\mathcal{J}_t| = K_1$  and  $\mathbf{s}_t, \mathbf{s}_{t,res}, \mathbf{r}_t$  are  $K_2, K_3, K_4$  sparse, respectively, and  $\|\mathbf{e}_t\|_2 \leq \varepsilon$ . Then

- 1) the estimation error after the KF step, i.e.,  $\|\mathbf{s}_t - \bar{\mathbf{s}}_t\|_2 = \|\mathbf{s}_{t,res}\|_2$  is bounded by:

$$\|\mathbf{s}_{t,res}\|_2 \leq a_t b_t [\|\mathbf{s}_t - \bar{\mathbf{s}}_t\|_2 + \|\mathbf{A}^T \mathbf{e}_t\|_2], \quad (19)$$

where  $a_t = [c + ((\|\mathbf{M}_{t-1}\|_2 + r)^{-1})^{-1}]^{-1}$ ,  $b_t = [((\|\mathbf{M}_{t-1}\|_2)^{-1} + r)^{-1}]^{-1}$ ,  $\mathbf{M}_t = \mathbf{A}^T \mathbf{A} + \sigma_{obs}^2 \tilde{\mathbf{P}}_t^{-1}$ ,  $r = \sigma_{sys}^2 / \sigma_{obs}^2$ ,  $\sigma_{sys}^2$  is the variance of  $\mathbf{Q} + \mathbf{G}$  and  $c$  is the minimum eigenvalue of  $\mathbf{A}^T \mathbf{A}$ .

- 2) when  $\delta_{2K_3} < \sqrt{2} - 1$ , the reconstruction error of ExMA-1 and ImMA-1 is bounded by:

$$\|\hat{\mathbf{s}}_t - \mathbf{s}_t\|_2 \leq C_0 K_3^{-1/2} \|\mathbf{s}_{t,res} - (\mathbf{s}_{t,res})_{K_3}\|_1 + C_1 \varepsilon, \quad (20)$$

where  $C_0 = [2 + (2\sqrt{2} - 2)\delta_{2K_3}] [1 - (\sqrt{2} + 1)\delta_{2K_3}]^{-1}$ ,  $C_1 = 4\sqrt{1 + \delta_{2K_3}} [1 - (\sqrt{2} + 1)\delta_{2K_3}]^{-1}$ .

- 3) when  $\delta_{2K_4}^2 + 2\delta_{K_1+2K_4} < 1$ , the reconstruction error of ExMA-2 and ImMA-2 is bounded by:

$$\|\hat{\mathbf{s}}_t - \mathbf{s}_t\|_2 \leq C_2 K_4^{-1/2} \|\mathbf{r}_t - (\mathbf{r}_t)_{K_4}\|_1 + C_3 \varepsilon, \quad (21)$$

where  $C_2 = [2 - 2\delta_{K_1+2K_4} + 2u] [1 - \delta_{K_1+2K_4} - u]^{-1}$ ,  $C_3 = 4\sqrt{1 + \delta_{K_1+2K_4}} [1 - \delta_{K_1+2K_4} - u]^{-1}$ ,  $u = \sqrt{\delta_{2K_4}^2 + \delta_{K_1+2K_4}^2}$ .

- 4) when  $\delta_{2K_3} < (\sqrt{2}K_3^{-1/2}P_1 + 1)^{-1}$ , the reconstruction error of ExMA-3 and ImMA-3 is bounded by:

$$\|\hat{\mathbf{s}}_t - \mathbf{s}_t\|_2 \leq C_4 K_3^{-1/2} \|\mathbf{s}_{t,res} - (\mathbf{s}_{t,res})_{K_3}\|_1 + C_5 \varepsilon, \quad (22)$$

where  $C_4 = [2P_2 + (\sqrt{2} - 2P_2)\delta_{2K_3}] [1 - (1 + \sqrt{2}K_3^{-1/2}P_1)\delta_{2K_3}]^{-1}$ ,  $C_5 = 2(1 + K_3^{-1/2}P_1)\sqrt{1 + \delta_{2K_3}} [1 - (\sqrt{2}K_3^{-1/2}P_1 + 1)\delta_{2K_3}]^{-1}$ . ■

*Proof:* Proofs are given in Appendix B.

**Remark 1:** Statement 1 of Theorem 2 indicates that the accuracy of the KF estimation  $\bar{\mathbf{s}}_t$  is closely related to the accuracy of the motion aware prediction, i.e.,  $\|\mathbf{s}_t - \check{\mathbf{s}}_t\|_2$ . More investigations concerning this prediction error are given in Section V-B.

*Remark 2:* The statements 2, 3 and 4 of Theorem 2 provide error bounds for different motion adaptive algorithms. All the bounds consist of two parts, i.e., a noise term and an  $l_1$  term, that relates to  $\mathbf{s}_{t,res}$ , i.e., the estimation error of the KF step. This indicates that the performance of these algorithms depends upon the sensing error and the accuracy of the KF estimation, which is in turn related to the accuracy of the motion aware prediction step, as implied in statement 1.

*Remark 3:* Comparing the statement 2, 3 and 4, we observe that different RICs are required to yield valid performance guarantees for different algorithms. These performance bounds are valid simultaneously if  $\delta_K < \min(\sqrt{2} - 1, (\sqrt{2}K_3^{-1/2}P_1 + 1)^{-1})$ , where  $K = \max(2K_3, K_1 + 2K_4)$ .

*Remark 4:* If the RIC is satisfied for both ExMA-1/ImMA-1 and ExMA-2/ImMA-2, we have  $C_2 < C_0$ ,  $C_3 \leq C_1$  since  $u \leq \sqrt{2}\delta_K < \sqrt{2}$ . Therefore, the upper error bound of ExMA-2/ImMA-2 is lower than that of ExMA-1/ImMA-1 when the  $l_1$  term obeys  $[K_4^{-1/2}||\mathbf{r}_t - (\mathbf{r}_t)_{K_4}||_1] \leq [K_3^{-1/2}||\mathbf{s}_{t,res} - (\mathbf{s}_{t,res})_{K_3}||_1]$ . Such an observation leads to the use of ExMA-2/ImMA-2 instead of ExMA-1/ImMA-1 in appropriate cases.

*Remark 5:* If the RIC is satisfied for both ExMA-1/ImMA-1 and ExMA-3/ImMA-3, we have  $C_4 \leq C_0$ ,  $C_5 \leq C_1$  if  $P_1 \leq K_3^{1/2}$ ,  $P_2 \leq 1 + \sqrt{2}\delta_K(2 - 2\delta_K)^{-1}$ . Thus under this condition, the upper bound of ExMA-3/ImMA-3 is lower than that of ExMA-1/ImMA-1, which implies that ExMA-3/ImMA-3 can be superior to ExMA-1/ImMA-1.

### B. Motion Aware Prediction Error

As discussed before, the reconstruction accuracy of the proposed algorithms closely depends on the accuracy of the KF estimation, which is in turn related to the motion aware prediction error, i.e.,  $\|\mathbf{s}_t - \tilde{\mathbf{s}}_t\|_2$ .

In the explicit ME/MC scheme, if we denote the MC result based on conventional ME using the true frames by  $\tilde{\mathbf{s}}_t$ , the prediction error can then be bounded by:

$$\|\mathbf{s}_t - \tilde{\mathbf{s}}_t\|_2 \leq \|\mathbf{s}_t - \tilde{\mathbf{s}}_t\|_2 + \|\tilde{\mathbf{s}}_t - \check{\mathbf{s}}_t\|_2, \quad (23)$$

where  $\|\mathbf{s}_t - \tilde{\mathbf{s}}_t\|_2$  measures the accuracy of the conventional ME and MC process; while  $\|\tilde{\mathbf{s}}_t - \check{\mathbf{s}}_t\|_2$  represents the additional error in the motion aware prediction step caused by the approximation of the true frame. It has been shown in [43]–[45] that the estimation error in the conventional ME and MC process, i.e.,  $\|\mathbf{s}_t - \tilde{\mathbf{s}}_t\|_2$ , depends on a number of factors such as the motion and light variation. Apart from the application-dependent factors,  $\|\tilde{\mathbf{s}}_t - \check{\mathbf{s}}_t\|_2$  describes the additional error owing to the low-pass filtering operation in our approach. We now evaluate the effect of this term by performing the following experimental study.

The study is conducted using the foreman sequence shown in Fig. 1. We first carry out conventional ME when the full resolution frames are known and from the evaluated MVs, we obtain the MC frames, i.e.,  $\tilde{\mathbf{s}}_t$ . Then to test the additional error in the explicit ME process, each of the test frames is sampled using the multiscale sampling strategy and the predicted frames, i.e.,  $\check{\mathbf{s}}_t$ , are obtained using the explicit ME/MC scheme. The original previous frame is used as the reference frame. Then the Normalized Squared Error (NSE) between  $\check{\mathbf{s}}_t$  and

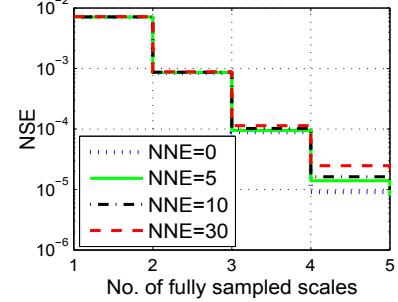


Fig. 5: NSEs of the explicit ME/MC prediction method with respect to the conventional ME for the foreman sequence.

TABLE I: Processing steps utilized in the evaluated approaches.

	Dictionary learning	Pre-CS Recon	ME/MC	KF	CS Recon
ExMA	no	no	yes	yes	yes
ImMA	yes	yes	no	yes	yes
VCS	no	yes	yes	no	yes
MAD	yes	no	no	no	yes
Ktfocuss	no	yes	yes	no	yes
ModCS	no	no	no	no	yes
KFCS	no	no	no	yes	yes

$\check{\mathbf{s}}_t$  is calculated as:  $NSE = \|\tilde{\mathbf{s}}_t - \check{\mathbf{s}}_t\|_2^2 / \|\tilde{\mathbf{s}}_t\|_2^2$ . This process is performed for 20 frames and the mean NSE values are recorded. We employ a 4 level 2D-DWT in our experiments and introduce Gaussian additive noise  $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_{obs}^2 \mathbf{I})$  during the measurement taking process. To keep the noise energy constant for each plot, we define the Normalized Noise Energy (NNE) as:  $NNE = \sigma_{obs}^2 M/N$ .

The results are shown in Fig. 5. Note that when the number of fully sampled scales is unity, only measurements of the approximate band are taken. Additional measurements from the next finer scales are introduced as the value on the x-axis of Fig. 5 is increased. As expected, the quality of the prediction improves when more scales are fully sampled and the noise power is reduced.

In the implicit ME scheme, the prediction error, i.e.,  $\|\mathbf{s}_t - \check{\mathbf{s}}_t\|_2$ , is in fact the CS reconstruction error. Therefore, according to Theorem 1, it is determined by the measurement noise and the sparsity of the frames when being represented in the motion-aware dictionaries. In particular, the reconstruction error of [27] represents the prediction error of the implicit ME scheme presented in Section IV-A2 and we will evaluate it in Section VI-C.

## VI. PERFORMANCE EVALUATION

In this section, we evaluate the proposed motion adaptive approach and compare its performance with various state-of-art approaches. Specifically, three motion-adaptive methods including Video Compressive Sensing (VCS) [22], Motion-Aware Decoding (MAD) [27], Ktfocuss [23], and two non-motion-adaptive causal methods including Modified CS (ModCS) [20] and KFCS [18] are considered in our simulations. These methods employ various process steps, that are identified explicitly in Table I.

### A. Implementation Details

As one may notice, the ExMA and ImMA approaches require different sampling schemes, i.e., multiscale sensing and block-based sensing, respectively. We thus examine performance for these schemes separately. To enable fair comparison, we implement the other chosen schemes using multiscale or block-based sensing as appropriate. That is, when evaluating the motion adaptive method using explicit ME, we implement all the approaches using multiscale sensing and reconstruction; while they are implemented using block-base schemes for evaluations of the implicit motion adaptive methods. In the following sections, we choose a naming convention where the initial letter, i.e., either “m” or “b”, indicate the sensing schemes. For instance, “bModCS” denotes the ModCS approach using the block-based scheme and “mModCS” is ModCS using the multiscale scheme.

For the MAD approach, the order-2 algorithm is employed and only a block-based scheme is implemented owing to the design of their system. Also, we only implement VCS using the multiscale scheme and the regularizers present in [22] are all included in our implementation. By “Ktfocuss” we refer to the method of Ktfocuss with ME/MC in [23] and the reconstruction before ME utilizes the Ktfocuss algorithm itself.

For evaluations of the explicit case, a 4 level 2D DWT is utilized for sparsifying and each scale is sampled by a i.i.d Gaussian sensing matrix. The first frame is reconstructed using multiscale CS reconstruction for initialization. For the high dimensional scales, block-paralleled KF and CS steps are conducted to ease implementation issues. We employ the block matching approach for all the ME processing involved in the experiments.

For the evaluations of the implicit case, a motion aware dictionary is employed in the prediction stage, whilst a 4 level 2D DWT is used for the KF and CS stages. The block size is set as  $32 \times 32$  and the searching window for the learning dictionaries is  $64 \times 64$ . Each block is sampled by a i.i.d Gaussian matrix. The first frame is reconstructed using block-based CS reconstruction for initialization. Each block is vectorized in the KF and CS stages.

In all the experiments, the videos in Fig. 1 are employed and their dimensions are: foreman  $256 \times 256$ , tennis  $1024 \times 1024$ , boats  $1024 \times 1024$ , windmill  $1024 \times 1024$  and flag  $512 \times 512$ . The Peak Signal to Noise Ratio (PSNR) is used for evaluation. The parameters are set as:  $\mathbf{P}_{0|0} = \mathbf{0}$ ,  $\varepsilon = \eta = \sigma_{obs}$ ,  $\lambda = 5$ . For the various settings of the sample ratio and noise energy, each experiment is repeated to get an overall or average result over 50 trials.

### B. Case I: Motion Adaptive Approach using Explicit ME/MC

1) *Reconstruction Accuracy:* We first carry out experiments on the foreman sequence. The sample ratio is set as  $M/N = 0.2$  and the noise energy is kept at  $NNE = 10$ . Fig. 6 shows a histogram of the PSNRs for frames 2 to 20 in all the 50 trials (i.e.,  $50 \times 19 = 950$  reconstructions) for each scheme. It is clear that the ExMA algorithms outperform the others, with ExMA-3 the best of all in terms of the mean PSNR. The

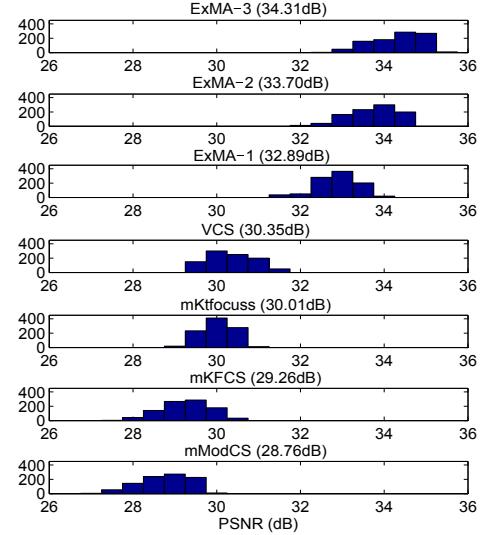


Fig. 6: Histograms of the PSNRs for each of the multiscale schemes for the foreman sequence. The mean PSNR for each algorithm is shown in the title. ( $M/N = 0.2$ ,  $NNE = 10$ )

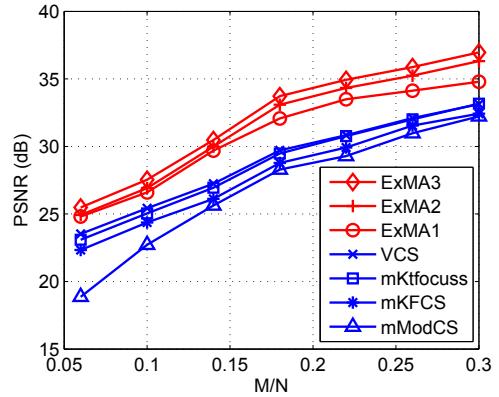


Fig. 7: PSNRs for the foreman sequence for the multiscale schemes when various sample ratios are employed. ( $NNE = 10$ )

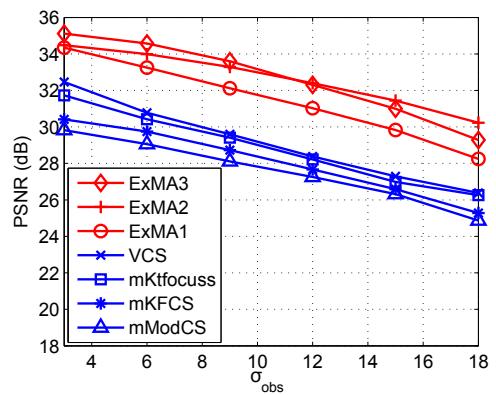


Fig. 8: PSNRs for the foreman sequence for the multiscale schemes when various amounts of noise are added. ( $M/N = 0.2$ )

TABLE II: PSNRs for various videos for the multiscale schemes.

(a) PSNRs (dB) for video Tennis										(b) PSNRs (dB) for video Windmill									
<i>NNE</i>	2	2	2	10	10	10	30	30	30	<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
ExMA-3	<b>30.5</b>	<b>35.8</b>	39.5	29.1	<b>33.3</b>	<b>37.0</b>	28.4	<b>30.9</b>	<b>32.4</b>	ExMA-3	<b>26.2</b>	<b>32.3</b>	<b>36.1</b>	<b>25.2</b>	<b>31.7</b>	<b>35.8</b>	24.8	<b>29.8</b>	<b>33.4</b>
ExMA-2	30.3	35.8	<b>39.6</b>	<b>29.5</b>	33.3	36.8	<b>28.6</b>	30.5	32.3	ExMA-2	26.0	32.2	36.0	25.0	31.6	35.4	24.6	29.7	33.3
ExMA-1	30.4	35.7	39.5	29.2	33.0	36.9	28.5	30.5	32.3	ExMA-1	26.0	32.2	36.0	25.2	31.7	35.4	<b>25.0</b>	29.6	33.4
VCS	29.0	33.9	38.0	27.5	31.3	35.2	26.5	28.8	30.8	VCS	23.3	31.0	34.9	22.0	30.2	33.9	21.6	27.9	31.9
mKtfocuss	28.7	33.8	37.9	27.0	31.0	35.4	26.2	28.6	30.7	mKtfocuss	22.8	30.6	34.8	21.8	30.0	34.0	21.2	27.4	32.0
mModCS	26.1	31.0	34.1	25.2	29.7	33.8	22.8	27.4	29.3	mModCS	21.2	29.3	33.0	20.5	27.8	32.0	19.6	26.7	31.3
mKFCS	27.2	31.9	35.0	26.5	30.5	34.2	25.4	28.1	29.8	mKFCS	21.7	26.3	31.1	21.1	26.0	30.5	20.5	25.8	30.0

(c) PSNRs (dB) for video Boats										(d) PSNRs (dB) for video Flag									
<i>NNE</i>	2	2	2	10	10	10	30	30	30	<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
ExMA-3	<b>28.0</b>	<b>35.7</b>	<b>39.3</b>	26.5	33.8	<b>37.8</b>	25.7	<b>32.7</b>	<b>35.6</b>	ExMA-3	<b>29.5</b>	33.9	36.1	<b>29.2</b>	33.4	<b>35.5</b>	27.0	29.6	<b>33.0</b>
ExMA-2	28.0	35.5	39.1	26.3	33.8	37.7	25.7	32.7	35.3	ExMA-2	29.3	34.0	<b>36.2</b>	28.9	<b>33.5</b>	35.4	27.2	29.6	32.9
ExMA-1	27.9	35.5	39.0	<b>26.7</b>	<b>34.0</b>	37.7	<b>25.9</b>	32.5	35.5	ExMA-1	29.3	<b>34.2</b>	36.2	29.0	33.5	35.4	<b>27.3</b>	<b>30.0</b>	32.9
VCS	26.3	34.8	37.8	25.5	32.5	36.6	24.8	31.1	33.8	VCS	28.0	33.1	34.8	27.5	32.2	34.2	26.0	28.4	31.4
mKtfocuss	26.0	34.8	37.9	25.4	32.3	36.5	24.4	30.6	33.7	mKtfocuss	27.6	32.8	34.9	27.0	32.0	34.0	25.4	28.0	31.1
mModCS	24.2	32.3	35.8	23.5	31.5	34.4	21.5	29.4	33.2	mModCS	23.8	30.9	34.0	23.0	29.4	33.6	21.0	26.9	30.2
mKFCS	25.0	31.0	34.4	24.2	30.1	33.1	23.8	29.8	32.8	mKFCS	27.0	31.8	34.0	26.4	30.8	33.7	24.8	27.8	30.2

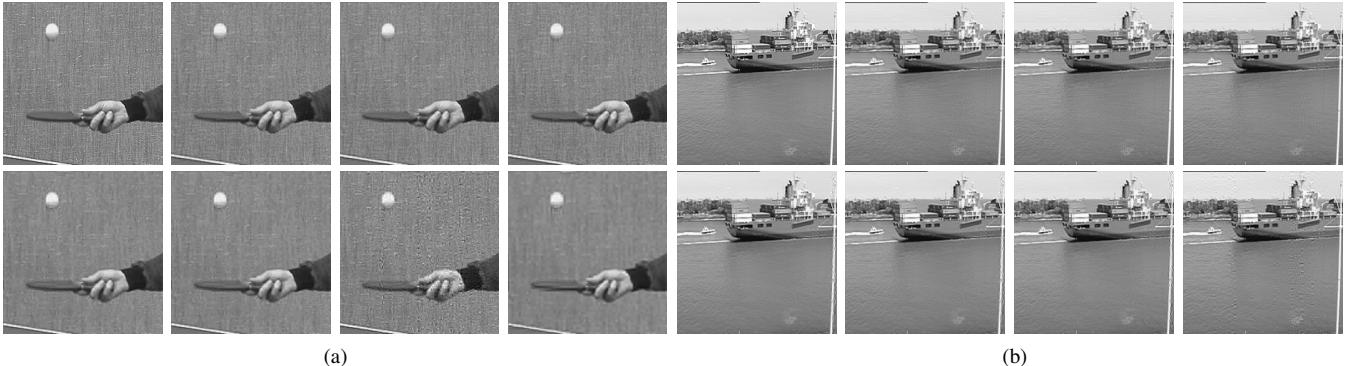


Fig. 9: Some example reconstructed frames of video (a) Tennis (6-th frame,  $M/N = 0.08$ ,  $NNE = 2$ ); (b) Boats (6-th frame,  $M/N = 0.15$ ,  $NNE = 2$ ). In each figure, the frames in the first row from left to right are that of: original video, ExMA-3, ExMA-2, ExMA-1; and the second row from left to right displays that of: VCS, mKtfocuss, mModCS, mKFCS.

ModCS approach performs worst with a mean PSNR about 6dB less than ExMA-3.

In Fig. 7, the performance of the algorithms are evaluated under various sample ratios. The result is averaged over 20 frames and 50 trials. It is observed that the ExMA algorithms are superior to the others for different sampling rates. To achieve a PSNR of 30dB, the ExMA algorithms require a sample ratio of about 14%; whilst the other approaches need sample ratios from about 19% to 24%.

The performance of all the algorithms with various amounts of added noise is shown in Fig. 8. We maintain the sample ratio at 0.2 during the experiment. The advantage of the ExMA algorithms can still be observed. It can be seen that when the noise variance increases, ExMA-2 gradually becomes the best of all followed by ExMA-3.

Similar results are evident for the experiments conducted using the other videos. The PSNR results are presented in Table II, where the highest PSNR values are highlighted in bold. We can see the ExMA algorithms perform best for all the experiments. Some example reconstructed frames are shown in Fig. 9.

TABLE III: Running time (per frame) in seconds for the foreman sequence for the multiscale schemes. ( $NNE = 10$ )

	$M/N$	Pre-Stages	ME/MC	KF&CS	Total
ExMA-3	0.1	N/A	10.01	36.78	46.79
ExMA-2	0.1	N/A	10.05	29.52	39.57
ExMA-1	0.1	N/A	10.07	35.75	45.82
VCS	0.1	21.04	9.90	21.65	52.59
mKtfocuss	0.1	13.23	9.91	11.69	34.83
mModCS	0.1	N/A	N/A	2.53	2.53
mKFCS	0.1	N/A	N/A	27.57	27.57
ExMA-3	0.3	N/A	8.04	36.60	44.64
ExMA-2	0.3	N/A	7.92	26.63	34.55
ExMA-1	0.3	N/A	7.96	35.36	43.32
VCS	0.3	20.40	8.75	20.87	50.02
mKtfocuss	0.3	12.26	8.71	10.57	31.54
mModCS	0.3	N/A	N/A	1.02	1.02
mKFCS	0.3	N/A	N/A	25.73	25.73

2) *Running Time:* Running time comparisons are also conducted using the foreman sequence. Due to the different stages employed in the various approaches, in Table III, we separately list the running time of the stages (with all the stages prior to ME/MC combined as Pre-stages), as well as the total running time. The results are averaged to yield a per frame running time and they are all obtained on a Macbook pro with a 2.6GHz Intel Core i5 CPU and 8GB RAM. To enable fair

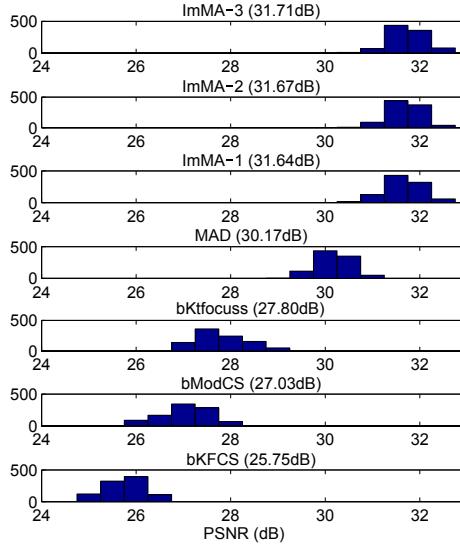


Fig. 10: Histograms of the PSNRs for each of the block-based schemes for the foreman sequence. The mean PSNR for each algorithm is shown in the title. ( $M/N = 0.2$ ,  $NNE = 10$ )

comparisons, the CS steps involved in all the approaches are set to stop when the norm of the difference between two consecutive estimates (normalized by the norm of the current estimate) is less than  $10^{-5}$ .

From Table III, we observe that the reconstructions in general run faster when the sampling rate increases, since more scales are fully sampled. Compared to the non-motion-adaptive methods, i.e., mKFCS and mModCS, the motion-adaptive approaches produce more accurate reconstruction at the cost of more computational burden that comes from the additional stages involved. We notice that the mKtfocuss approach runs faster than the other motion-adaptive methods. This is because it utilizes the less complex weighted  $l_2$  minimization during reconstruction and also we implement it in a block-parallel manner for high dimensional scales. Among all the motion-adaptive methods, VCS is the slowest. In addition, by comparing the running time of the prediction stages, i.e., pre-stages plus ME/MC stage, we observe that our prediction approach has lower computational complexity than the other motion adaptive approaches.

### C. Case II: Motion Adaptive Approach using Implicit ME

1) *Reconstruction Accuracy:* The previously described experiments are now carried out using the block-based sensing scheme to evaluate the performance of the motion adaptive approach when the implicit ME is used. Fig. 10 shows a histogram of PSNRs for each scheme when the foreman sequence is employed. The sample ratio is 0.2 and the noise energy is 10. We can see that the motion-adaptive approaches are superior to the non-motion-adaptive methods and the ImMA algorithms outperform all of the other approaches.

In Fig. 11, the sample ratio is varied to examine reconstruction performance. It is observed that the ImMA algorithms have better reconstruction quality than the others over all the tested sample ratios. For a sample ratio of 20%, all the ImMA algorithms can achieve a PSNR of about 32dB, followed by

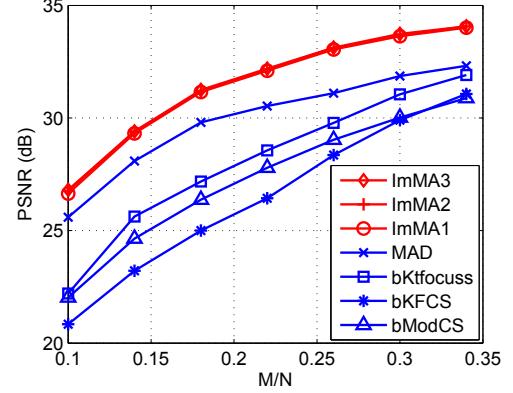


Fig. 11: PSNRs for the foreman sequence for the block-based schemes when various sample ratios are employed. ( $NNE = 10$ )

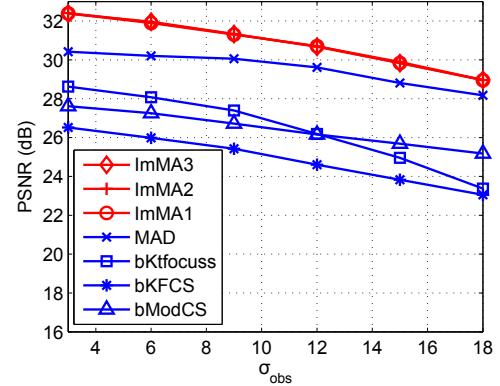


Fig. 12: PSNRs for the foreman sequence for the block-based schemes when various amounts of noise are included. ( $M/N = 0.2$ )

MAD with a PSNR of about 30dB, whilst the PSNRs of all the other schemes are all below 28dB.

When the noise energy varies, the performance of all the approaches are illustrated in Fig. 12. The advantage of the ImMA methods over all the others is clear, although the improvement over the MAD method decreases when the noise is larger. When the noise variance is above 12, we observe that the ModCS algorithm, that is non-motion-adaptive, outperforms the motion-adaptive approach Ktfocuss.

In addition, from Fig. 10 - 12, we observe that the various ImMA algorithms have very similar performance. The reason is that the implicit motion aware prediction already exploits CS reconstruction (while the explicit prediction procedure does not), which together with the KF steps lead to a good estimation of the current frame, and the varying degrees of performance improvements offered by an additional CS reconstruction for the ImMA schemes are not sufficiently significant to affect the overall performance of these schemes.

The experiments are also conducted using the other videos yielding the results shown in Table IV, where the highest PSNR values appear in bold. Some example reconstructed frames are shown in Fig 13.

2) *Running Time:* Using the same computing platform as in the section VI-B 2, we now compare the running time

TABLE IV: PSNRs for various videos for the block based schemes.

(a) PSNRs (dB) for video Tennis										(b) PSNRs (dB) for video Windmill									
<i>NNE</i>	2	2	2	10	10	10	30	30	30	<i>NNE</i>	2	2	2	10	10	10	30	30	30
<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30	<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
ImMA-3	<b>28.9</b>	<b>33.3</b>	<b>36.9</b>	<b>28.3</b>	32.2	34.2	<b>26.6</b>	29.2	<b>32.3</b>	ImMA-3	<b>28.5</b>	<b>32.1</b>	<b>35.3</b>	<b>27.5</b>	<b>31.8</b>	<b>33.2</b>	26.6	31.1	<b>32.6</b>
ImMA-2	28.9	33.1	36.9	28.2	<b>32.3</b>	<b>34.3</b>	26.6	<b>29.3</b>	32.3	ImMA-2	28.3	32.0	35.3	27.4	31.8	33.2	<b>26.7</b>	<b>31.3</b>	32.6
ImMA-1	28.8	33.2	36.9	28.3	32.2	34.3	26.6	29.2	32.3	ImMA-1	28.4	32.0	35.3	27.4	31.8	33.2	26.6	31.2	32.6
MAD	27.9	31.8	34.9	27.7	31.2	32.4	25.2	28.1	30.4	MAD	26.9	30.7	33.2	26.1	30.6	32.1	25.7	30.3	31.8
bKtfocuss	26.4	30.8	33.6	26.2	29.6	31.8	24.4	27.4	29.9	bKtfocuss	22.9	28.4	32.8	22.5	27.9	31.4	20.9	26.0	28.5
bModCS	25.8	30.1	33.1	25.7	29.3	31.6	25.2	28.1	29.9	bModCS	21.9	27.9	32.2	21.8	27.8	31.2	21.6	27.0	29.6
bKFCS	24.2	29.0	32.6	23.9	28.6	31.9	23.5	27.3	30.4	bKFCS	20.1	25.4	28.8	19.7	24.8	28.5	18.9	23.6	27.9

(c) PSNRs (dB) for video Boats										(d) PSNRs (dB) for video Flag									
<i>NNE</i>	2	2	2	10	10	10	30	30	30	<i>NNE</i>	2	2	2	10	10	10	30	30	30
<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30	<i>M/N</i>	0.10	0.20	0.30	0.10	0.20	0.30	0.10	0.20	0.30
ImMA-3	<b>28.3</b>	<b>35.5</b>	<b>39.3</b>	27.6	33.3	<b>35.3</b>	27.3	31.8	32.7	ImMA-3	<b>28.9</b>	<b>32.3</b>	<b>35.2</b>	<b>28.2</b>	<b>31.6</b>	<b>34.7</b>	<b>27.8</b>	<b>30.8</b>	<b>31.8</b>
ImMA-2	28.2	35.5	39.3	27.6	<b>33.4</b>	35.3	<b>27.4</b>	31.8	<b>32.8</b>	ImMA-2	28.8	32.3	35.2	28.2	31.6	34.7	27.8	30.8	31.8
ImMA-1	28.3	35.5	39.3	<b>27.7</b>	33.3	35.3	27.3	<b>31.9</b>	32.8	ImMA-1	28.9	32.3	35.2	28.2	31.6	34.7	27.7	30.8	31.8
MAD	27.1	33.4	37.3	26.9	32.1	34.3	25.9	30.9	32.1	MAD	27.6	31.0	33.8	26.1	30.7	33.7	25.8	29.7	30.5
bKtfocuss	24.5	31.2	35.0	24.2	30.3	33.2	22.3	28.1	30.4	bKtfocuss	25.2	29.5	33.1	24.5	29.1	32.2	22.9	27.6	29.7
bModCS	23.6	30.6	34.1	23.5	29.8	32.7	23.5	28.6	30.5	bModCS	23.6	28.7	32.2	23.2	28.1	31.1	22.9	27.5	29.3
bKFCS	21.6	26.7	31.6	21.0	26.3	31.2	20.7	25.8	30.0	bKFCS	21.9	26.2	30.6	21.7	26.0	30.4	21.2	25.9	29.0



Fig. 13: Some example reconstructed frames of video (a) Windmill (8-th frame,  $M/N = 0.30$ ,  $NNE = 2$ ); (b) Flag (14-th frame,  $M/N = 0.30$ ,  $NNE = 10$ ). In each figure, the frames in the first row from left to right are that of: original video, ExMA-3, ExMA-2, ExMA-1; and the second row from left to right displays that of: VCS, mKtfocuss, mModCS, mKFCS.

of the foreman sequence for the block-based algorithms in Table V. We observe that when the sampling rate increases, all the approaches become slower due to the larger sensing matrices involved. Kfocuss takes the longest time when the sampling rate is 0.3. The motion-adaptive methods are not necessarily slower than the non-motion-adaptive ones, e.g., ModCS, because the representations of the frames are more sparse in the motion-adaptive case owing to the utilization of the motion-aware dictionaries.

By comparing Table III and Table V, we notice that the prediction stage of ExMA (ME/MC stage in Table III) is much faster than that of ImMA (Pre-Stages in Table V), which implies that predicting using the explicit ME/MC is less complex than using the implicit scheme. Furthermore, it can also be observed that the speed of Ktfocuss, ModCS and KFCS change significantly when the different implementations are used. This is because the dimensions involved in the reconstruction vary due to the different sensing schemes. Meanwhile, the signals involved in their CS steps are related to either predictions or previous frames, which are different depending upon the sensing schemes used. It means that the sparsity of these signals in the CS stage varies and thus the

TABLE V: Running time (per frame) in seconds for the foreman sequence for the block-based schemes. ( $NNE = 10$ )

$M/N$	Pre-Stages	ME/MC	KF&CS	Total
ImMA-3	0.1	25.07	N/A	48.44
ImMA-2	0.1	24.72	N/A	46.84
ImMA-1	0.1	24.91	N/A	47.20
MAD	0.1	22.40	N/A	25.01
bKtfocuss	0.1	18.69	6.20	17.89
bModCS	0.1	N/A	N/A	66.97
bKFCS	0.1	N/A	N/A	7.03
ImMA-3	0.3	24.34	N/A	28.07
ImMA-2	0.3	24.05	N/A	27.56
ImMA-1	0.3	24.25	N/A	28.01
MAD	0.3	21.77	N/A	2.38
bKtfocuss	0.3	59.22	6.45	57.88
bModCS	0.3	N/A	N/A	76.73
bKFCS	0.3	N/A	N/A	9.80

convergence rate, i.e., the time taken to achieve the stopping criteria changes.

## VII. CONCLUSIONS

In this paper, we consider the problem of causally reconstructing videos captured by a CS camera. A framework that adapts the reconstruction process to motion in the video is proposed and it incorporates motion aware prediction, KF

updating and CS reconstruction to improve the CS video reconstruction. Under this framework, two motion aware prediction methods using explicit ME/MC and implicit ME are proposed for block-based and multiscale CS systems, respectively, in which the explicit ME prediction scheme has lower computational complexity. The reconstruction accuracy and performance guarantees for various algorithms based on the proposed framework are analyzed theoretically via upper error bounds, which shows that the performance of our method depends on the sensing noise and the accuracy of the ME methods. Under various settings of the sample ratios and sensing noise, comprehensive simulations using different videos demonstrate that the proposed motion-adaptive approach has superior reconstruction accuracy in comparison to the state-of-art causal and motion adaptive methods.

## APPENDIX A DERIVATION OF ME ENHANCED KF

Using the models defined in (11), (12), and considering the process noise  $\mathbf{q}_t$ , the error of the additive corrections  $\mathbf{g}_t$  and the state estimation error  $(\mathbf{s}_{t-1} - \bar{\mathbf{s}}_{t-1})$  are uncorrelated, we derive the prediction error covariance matrix by:

$$\begin{aligned}\check{\mathbf{P}}_t &= \mathbb{E}[(\mathbf{s}_t - \check{\mathbf{s}}_t)(\mathbf{s}_t - \check{\mathbf{s}}_t)^T] \\ &= \mathbb{E}[(\mathbf{s}_{t-1} - \bar{\mathbf{s}}_{t-1})(\mathbf{s}_{t-1} - \bar{\mathbf{s}}_{t-1})^T] \\ &\quad + \mathbb{E}[(\mathbf{g}_t + \mathbf{q}_t)(\mathbf{g}_t + \mathbf{q}_t)^T] \\ &= \mathbf{P}_{t-1} + \mathbb{E}(\mathbf{q}_t \mathbf{q}_t^T) + \mathbb{E}(\mathbf{g}_t \mathbf{g}_t^T) \\ &= \mathbf{P}_{t-1} + \mathbf{Q} + \mathbf{G}.\end{aligned}\quad (24)$$

Then with  $\check{\mathbf{s}}_t$  and  $\check{\mathbf{P}}_t$  obtained, as in conventional KF [41], we consider the distribution of the prediction as a Gaussian function with a mean  $\check{\mathbf{s}}_t$  and a variance  $\check{\mathbf{P}}_t$ . In [41], the conventional KF is derived using a fusion process of Gaussian functions. Following the same procedure, we can acquire a fused distribution of the motion aware prediction and the measurement, which is the distribution of the state estimation of the ME enhanced KF, and its mean value  $\bar{\mathbf{s}}_t$  and variance  $\check{\mathbf{P}}_t$  can then be derived as shown in (14) - (16).

## APPENDIX B PROOF OF THEOREM 2

For statement 1, we adapt the derivation of [46] to our settings. Define  $\beta_t = \mathbf{s}_t - \bar{\mathbf{s}}_t$ . As  $\bar{\mathbf{s}}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{A})\check{\mathbf{s}}_t + \mathbf{K}_t \mathbf{y}_t$ , we can get:

$$\beta_t = (\mathbf{I} - \mathbf{K}_t \mathbf{A})(\mathbf{s}_t - \check{\mathbf{s}}_t) - \mathbf{K}_t \mathbf{e}_t. \quad (25)$$

Let  $\mathbf{M}_t = \mathbf{A}^T \mathbf{A} + \sigma_{obs}^2 \check{\mathbf{P}}_t^{-1}$ ,  $\mathbf{J}_t = \mathbf{I} - \mathbf{K}_t \mathbf{A}$ , we can derive the following equations:

$$\begin{aligned}\mathbf{K}_t &= \mathbf{M}_t^{-1} \mathbf{A}^T, \quad \mathbf{J}_t = \sigma_{obs}^2 \mathbf{M}_t^{-1} \check{\mathbf{P}}_t^{-1}, \\ \check{\mathbf{P}}_t &= \mathbf{P}_{t-1} + \sigma_{sys}^2 \mathbf{I}, \\ \mathbf{P}_{t-1} &= \sigma_{obs}^2 \mathbf{M}_{t-1}^{-1},\end{aligned}\quad (26)$$

where  $\sigma_{sys}^2$  denotes the variance of  $\mathbf{Q} + \mathbf{G}$ . It can be observed that to bound  $\|\beta_t\|_2$ , we need to bound  $\|\mathbf{J}_t\|_2$ , which in turn requires bounding  $\|\mathbf{M}_t^{-1}\|_2$  and  $\|\check{\mathbf{P}}_t^{-1}\|_2$ .

Define  $r = \sigma_{sys}^2 / \sigma_{obs}^2$  and  $\lambda_{min}(\mathbf{X})$ ,  $\lambda_{max}(\mathbf{X})$  as the minimum and maximum eigenvalue of a matrix  $\mathbf{X}$ , respectively. The following properties for symmetric positive definite matrices  $\mathbf{X}$ ,  $\mathbf{Y}$  are used:  $\|\mathbf{X}\|_2 = \lambda_{max}(\mathbf{X}) = 1/\lambda_{min}(\mathbf{X}^{-1})$ ,

$\lambda_{min}(\mathbf{X} + \mathbf{Y}) \geq \lambda_{min}(\mathbf{X}) + \lambda_{min}(\mathbf{Y})$ ,  $\lambda_{max}(\mathbf{X} + \mathbf{Y}) \leq \lambda_{max}(\mathbf{X}) + \lambda_{max}(\mathbf{Y})$ . Denote  $\lambda_{min}(\mathbf{A}^T \mathbf{A})$  with a constant  $c$ . Using (26), we can get:

$$\begin{aligned}\|\mathbf{M}_t^{-1}\|_2 &= [\lambda_{min}(\mathbf{A}^T \mathbf{A} + \sigma_{obs}^2 \check{\mathbf{P}}_t^{-1})]^{-1} \\ &\leq [c + \frac{\sigma_{obs}^2}{\lambda_{max}(\check{\mathbf{P}}_t)}]^{-1} \\ &\leq [c + \frac{1}{\|\mathbf{M}_{t-1}^{-1}\|_2 + r}]^{-1}.\end{aligned}\quad (27)$$

Then we can also get:

$$\begin{aligned}\sigma_{obs}^2 \|\check{\mathbf{P}}_t^{-1}\|_2 &\leq \frac{\sigma_{obs}^2}{\sigma_{obs}^2 \lambda_{min}(\mathbf{M}_{t-1}^{-1}) + \sigma_{sys}^2} \\ &= [\frac{1}{\|\mathbf{M}_{t-1}^{-1}\|_2} + r]^{-1}.\end{aligned}\quad (28)$$

If we denote the bounds in (27) and (28) as  $a_t$  and  $b_t$ , respectively, we can derive the bound of  $\|\beta_t\|_2$  as:

$$\begin{aligned}\|\beta_t\|_2 &= \|\mathbf{M}_t^{-1}[\sigma_{obs}^2 \check{\mathbf{P}}_t^{-1}(\mathbf{s}_t - \check{\mathbf{s}}_t) - \mathbf{A}^T \mathbf{e}_t]\|_2 \\ &\leq a_t b_t [\|\mathbf{s}_t - \check{\mathbf{s}}_t\|_2 + \|\mathbf{A}^T \mathbf{e}_t\|_2].\end{aligned}$$

The next two statements are straight forward applications of the error bounds in CS [33] and modified CS [47]. For both of the statements, we can get  $\|\mathbf{s}_t - \hat{\mathbf{s}}_t\|_2 = \|\bar{\mathbf{s}}_t + \mathbf{s}_{t,res} - (\bar{\mathbf{s}}_t + \hat{\mathbf{s}}_{t,res})\|_2 = \|\mathbf{s}_{t,res} - \hat{\mathbf{s}}_{t,res}\|_2$ . Therefore, the error bounds are the same as applying CS or modified CS to  $\mathbf{s}_{t,res}$ , which leads to the results in Theorem 2.

We now derive the error bound for the following problem:

$$\min_{\mathbf{s}_t} \|\mathbf{W}_t \mathbf{s}_t\|_1 \text{ s.t. } \|\mathbf{y}_t - \mathbf{A} \mathbf{s}_t\|_2 \leq \varepsilon,$$

where  $\mathbf{W}_t = \text{diag}[w^{(1)}, w^{(2)}, \dots, w^{(N)}]$ .

The method of [33] is employed. Set  $\mathbf{s}_t^* = \mathbf{s}_t + \mathbf{h}_t$  and decompose  $\mathbf{h}_t$  into a sum of vectors  $(\mathbf{h}_t)_{\mathcal{T}_0}, (\mathbf{h}_t)_{\mathcal{T}_1}, (\mathbf{h}_t)_{\mathcal{T}_2}, \dots$ , each of sparsity at most  $K$ .  $\mathcal{T}_0$  corresponds to the locations of the  $K$  largest coefficients of  $\mathbf{s}_t$ ;  $\mathcal{T}_1$  to the locations of the  $K$  largest coefficients of  $(\mathbf{h}_t)_{\mathcal{T}_0^c}$ ;  $\mathcal{T}_2$  to the locations of the next  $K$  largest coefficients of  $(\mathbf{h}_t)_{\mathcal{T}_0^c}$  and so on. Then we can get:

$$\begin{aligned}\|\mathbf{W}_t \mathbf{s}_t\|_1 &\geq \|\mathbf{W}_t(\mathbf{s}_t + \mathbf{h}_t)\|_1 \\ &= \sum_{i \in \mathcal{T}_0} |w^{(i)}(\mathbf{s}_t^{(i)} + \mathbf{h}_t^{(i)})| + \sum_{i \in \mathcal{T}_0^c} |w^{(i)}(\mathbf{s}_t^{(i)} + \mathbf{h}_t^{(i)})| \\ &\geq \|(\mathbf{W}_t \mathbf{s}_t)_{\mathcal{T}_0}\|_1 - \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0}\|_1 \\ &\quad + \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0^c}\|_1 - \|(\mathbf{W}_t \mathbf{s}_t)_{\mathcal{T}_0^c}\|_1. \\ \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0^c}\|_1 &\leq \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0}\|_1 + 2 \|(\mathbf{W}_t \mathbf{s}_t)_{\mathcal{T}_0^c}\|_1.\end{aligned}\quad (29)$$

According to the Cauchy-Schwarz inequality, we can get:

$$\begin{aligned}\|(\mathbf{h}_t)_{\mathcal{T}_0}\|_1^2 &\leq K \|(\mathbf{h}_t)_{\mathcal{T}_0}\|_2^2, \\ \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0}\|_1 &\leq \sum_{i \in \mathcal{T}_0} [(w^{(i)})^2] \|(\mathbf{h}_t)_{\mathcal{T}_0}\|_2.\end{aligned}\quad (30)$$

We define  $w_{min}$  and  $w_{max}$  as the minimum and maximum value of  $w^{(i)} (i \in \mathcal{T}_0^c)$ , respectively. Additionally, we define  $P_1 = w_{min}^{-1} \sum_{i \in \mathcal{T}_0} (w^{(i)})^2$ ,  $P_2 = w_{min}^{-1} w_{max}$ . Using (29)(30) and the following facts:  $\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)^c}\|_2 \leq K^{-1/2} \|(\mathbf{h}_t)_{\mathcal{T}_0^c}\|_1$ ,  $\|(\mathbf{s}_t)_{\mathcal{T}_0^c}\|_1 = \|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1$ , we can derive:

$$\begin{aligned}\|(\mathbf{h}_t)_{\mathcal{T}_0^c}\|_1 &\leq w_{min}^{-1} \|(\mathbf{W}_t \mathbf{h}_t)_{\mathcal{T}_0^c}\|_1 \\ &\leq P_1 \|(\mathbf{h}_t)_{\mathcal{T}_0}\|_2 + 2P_2 \|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1,\end{aligned}\quad (31)$$

$$\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)^c}\|_2 \leq K^{-\frac{1}{2}} P_1 \|(\mathbf{h}_t)_{\mathcal{T}_0}\|_2 + 2K^{-\frac{1}{2}} P_2 \|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1. \quad (32)$$

Following the same steps as in [33], we have:  $\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2 \leq \alpha \varepsilon + \beta K^{-1/2} \|(\mathbf{h}_t)_{\mathcal{T}_0^c}\|_1$ , where

$\alpha = 2\sqrt{1 + \delta_{2K}}(1 - \delta_{2K})^{-1}$ ,  $\beta = \sqrt{2}\delta_{2K}(1 - \delta_{2K})^{-1}$ . By substituting (31), we get  $\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2 \leq \alpha\varepsilon + \beta K^{-1/2}P_1\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2 + 2\beta K^{-1/2}P_2\|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1$ . Assume that  $1 - \beta K^{-1/2}P_1 > 0$ , then

$$\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2 \leq \frac{\alpha\varepsilon + 2\beta K^{-1/2}P_2\|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1}{1 - \beta K^{-1/2}P_1}. \quad (33)$$

We now conclude from (32) and (33) that

$$\begin{aligned} \|(\mathbf{h}_t)_{\mathcal{T}_0}\|_2 &\leq (1 + K^{-\frac{1}{2}}P_1)\|(\mathbf{h}_t)_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2 + 2K^{-\frac{1}{2}}P_2\|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1 \\ &\leq \frac{(1 + \beta)2P_2}{1 - \beta K^{-\frac{1}{2}}P_1}K^{-1/2}\|\mathbf{s}_t - (\mathbf{s}_t)_K\|_1 + \frac{(1 + K^{-\frac{1}{2}}P_1)\alpha}{1 - \beta K^{-\frac{1}{2}}P_1}\varepsilon. \end{aligned}$$

Then, for ExMA-3 and ImMA-3, as  $\|\mathbf{s}_t - \hat{\mathbf{s}}_t\|_2 = \|\mathbf{s}_{t,res} - \hat{\mathbf{s}}_{t,res}\|_2$ , we can derive the last statement in Theorem 2 by appropriate substitution.

## REFERENCES

- [1] X. Ding, W. Chen, and I. J. Wassell, "Generalized-KFCS: Motion estimation enhanced Kalman filtered compressive sensing for video," in *Proc. IEEE ICIP*, Oct. 2014, pp. 1297–1301.
- [2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.
- [3] D. Dohono, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52(4), pp. 1289–1306, 2006.
- [4] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Sig. Process. Mag.*, vol. 25, no. 2, pp. 83 – 91, 2008.
- [5] R. F. Marcia, Z. T. Harmany, and R. M. Willett, "Compressive coded aperture imaging," *SPIE 7246, Comput. Imag. VII*, p. 72460G, 2009.
- [6] R. Robucci, L. K. Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson, "Compressive sensing on a CMOS separable transform image sensor," in *Proc. IEEE ICASSP*, Mar. 2008, pp. 5125–5128.
- [7] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15 (21), pp. 14013–14027, 2007.
- [8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655 – 4666, Dec. 2007.
- [9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.
- [10] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1094 –1121, Feb. 2012.
- [11] Z. Liu, A.Y. Elezzabi, and H.V. Zhao, "Maximum frame rate video acquisition using adaptive compressed sensing," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, no. 11, pp. 1704–1718, 2011.
- [12] J. Xu, J. Ma, D. Zhang, Y. Zhang, and S. Lin, "Improved total variation minimization method for compressive sensing by intra-prediction," *Signal Process.*, vol. 92, no. 11, pp. 2614–2623, 2012.
- [13] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Mag. Res. in Medicine*, vol. 59(2), pp. 365–373, 2008.
- [14] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Proc. PCS*, 2006.
- [15] J. Ma, G. Plonka, and M.Y. Hussaini, "Compressive video sampling with approximate message passing decoding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 22, no. 9, pp. 1354–1364, 2012.
- [16] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477 – 2488, July 2005.
- [17] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE ICIP*, 2008, pp. 893–896.
- [18] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MR image reconstruction using Kalman filtered compressed sensing," in *Proc. IEEE ICASSP*, 2009, pp. 393–396.
- [19] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [20] W. Lu and N. Vaswani, "Modified basis pursuit denoising (modified-BPDN) for noisy compressive sensing with partially known support," *CoRR*, 2009.
- [21] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proc. PCS*, 2009, pp. 197–200.
- [22] M.S. Asif, F. Fernandes, and J. Romberg, "Low-complexity video compression and compressive sensing," in *proc. ACSSC*, Nov 2013.
- [23] H. Jung and J. C. Ye, "Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques," *Inter. J. IST*, vol. 20, pp. 81–98, 2010.
- [24] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "K-T FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Mag. Res. in Medicine*, vol. 35, pp. 2313–2351, 2007.
- [25] J. Guo, B. Song, H. Liu, and H. Qin, "Motion estimation in measurement domain for compressed video sensing," in *proc. IEEE CIT*, Sept 2014.
- [26] N. Jacobs, S. Schuh, and R. Pless, "Compressive sensing and differential image-motion estimation," in *proc. IEEE ICASSP*, March 2010.
- [27] Y. Liu, M. Li, and D.A. Pados, "Motion-aware decoding of compressed-sensed video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 3, pp. 438–444, Mar. 2013.
- [28] T. T. Do, Y. Chen, D.T. Nguyen, N. Nguyen, L. Gan, and T.D. Tran, "Distributed compressed video sensing," in *Proc. IEEE ICIP*, 2009.
- [29] H. Chen, L. Kang, and C. Lu, "Dictionary learning-based distributed compressive video sensing," in *Proc. PCS*, Dec 2010, pp. 210–213.
- [30] J. Xu, J. Ma, D. Zhang, Y. Zhang, and S. Lin, "Compressive video sensing based on user attention model," in *Proc. IEEE PCS*. IEEE, 2010, pp. 90–93.
- [31] X. Ding, W. Chen, and I. Wassell, "Block-based feature adaptive compressive sensing for video," in *Proc. IEEE CIT/IUCC/DASC/PICOM*, 2015, pp. 1675–1680.
- [32] W. Chen and M.R.D. Rodrigues, "Dictionary learning with optimized projection design for compressive sensing applications," *IEEE Signal Process. Letters*, vol. 20, no. 10, pp. 992–995, Oct 2013.
- [33] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9 - 10, pp. 589 – 592, 2008.
- [34] E.J. Candès and M.B. Wakin, "An introduction to compressive sampling," *IEEE Sig. Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [35] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 3, pp. 549–571, Mar. 2006.
- [36] S. Ji and L. Carin, "Bayesian compressive sensing and projection optimization," in *Proc. ICML*, 2007.
- [37] J. Wu, F. Liu, L.C. Jiao, and X. Wang, "Compressive sensing SAR image reconstruction based on Bayesian framework and evolutionary computation," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1904–1911, July 2011.
- [38] X. Bi, X. Chen, Y. Zhang, and B. Liu, "Image compressed sensing based on wavelet transform in contourlet domain," *Signal Process.*, vol. 91 (5), no. 5, pp. 1085 – 1092, 2011.
- [39] X. Zhang, J. Wen, Y. Han, and J. Villasenor., "An improved compressive sensing reconstruction algorithm using linear/non-linear mapping," in *ITA workshop*, Feb 2011, pp. 1–7.
- [40] M. Chriqui and P. Sinha, "Survey of motion estimation techniques for video compression," in *Low-Light-Level and Real-Time Imaging Systems, Components, and Applications*, 2003, vol. 4796, pp. 218–226.
- [41] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation [lecture notes]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 128–132, 2012.
- [42] W. Chen, M.R.D. Rodrigues, and I.J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, April 2013.
- [43] K. C. Hui and W. C. Siu, "Extended analysis of motion-compensated frame difference for block-based motion prediction error," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1232–1245, May 2007.
- [44] W. Niehsen and M. Brunig, "Covariance analysis of motion-compensated frame differences," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 536–539, Jun 1999.
- [45] C. F. Chen and K. K. Pang, "The optimal transform of motion-compensated frame difference images in a hybrid coder," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Process.*, vol. 40, no. 6, pp. 393–397, Jun 1993.
- [46] N. Vaswani, "KF-CS: Compressive sensing on Kalman filtered residual," *Proc. CoRR*, vol. abs/0912.1628, 2009.
- [47] L. Jacques, "A short note on compressed sensing with partially known signal support," *proc. CoRR*, 2009.