

Compressed-sensing recovery of multiview image and video sequences using signal prediction

Maria Trocan · Eric W. Tramel ·
James E. Fowler · Beatrice Pesquet

Published online: 4 January 2013
© Springer Science+Business Media New York 2013

Abstract In the compressed sensing of multiview images and video sequences, signal prediction is incorporated into the reconstruction process in order to exploit the high degree of interview and temporal correlation common to multiview scenarios. Instead of recovering each individual frame independently, neighboring frames in both the view and temporal directions are used to calculate a prediction of a target frame, and the difference is used to drive a residual-based compressed-sensing reconstruction. The proposed approach demonstrates a significant gain in reconstruction quality relative to the straightforward compressed-sensing recovery of each frame independently of the others in the multiview set, as well as a significant performance advantage as compared to a pair of benchmark multiple-frame compressed-sensing reconstructions.

Keywords Compressed sensing · Multiviews · Signal prediction

M. Trocan (✉)
Institut Supérieur d'Électronique de Paris,
28 rue Notre Dame des Champs, 75006 Paris, France
e-mail: maria.trocan@isep.fr

E. W. Tramel · J. E. Fowler
Geosystems Research Institute, Mississippi State University,
Box 9571, Mississippi State, MS 39762, USA

E. W. Tramel
e-mail: ewt16@msstate.edu

J. E. Fowler
e-mail: fowler@ece.msstate.edu

B. Pesquet
Télécom ParisTech, 46 rue Barrault, 75634 Paris Cédex 13, France
e-mail: beatrice.pesquet@telecom-paristech.fr

1 Introduction

The falling cost of high-quality video sensors coupled with their increasingly widespread use in surveillance, defense, and entertainment applications has led to heightened demand for multi-sensor video-acquisition systems. In surveillance applications, for example, the use of video-sensor networks has been widely investigated, but the memory and computation burden of capturing and encoding high-quality video for transmission and storage has served as an impediment to the adoption of multi-sensor technology in many applications [27]. In the area of entertainment, much work has been done recently to promote the production and consumption of 3D video content, which so far has largely taken the form of stereoscopic video-display systems with a fixed viewpoint. Future display technologies, such as holography, promise a more realistic and engaging viewing experience by permitting many different viewing angles; however, capturing such multiview-image data requires a system more sophisticated than the two-camera approach widely used today. In these, and in other applications, the excessively voluminous nature of multiview-image and video data causes a serious impediment to the continued development of these fields.

Compressed sensing (CS) (e.g., [5]) is a recent paradigm which permits linear projection of a signal into a dimension much lower than that of the original signal while still providing a method of recovery which, under certain strict constraints, incurs little to no loss. The CS methodology effectively combines signal acquisition and dimensionality reduction into a single step, thereby reducing memory and computational requirements within the sensing device as well as transmission bandwidth. Particular interest has centered on CS for images and video, and physical implementations based upon CS have been created, such as the “single-pixel” camera of [10].

In the context of multiview images and multiview video, CS has the potential to greatly enhance multiview signal acquisition not only by decreasing the inherent memory cost by lowering the number of measurements taken, but also by decreasing the computational burden on the sensor. As the dimensionality-reduction aspect of CS can be accomplished via modulation and projection of the analog light signal onto a single sensor [10], essentially zero computation is needed on-board the actual sensing device. Therefore, it is hoped that this acquisition process will be realizable with video sensors that are much cheaper and energy efficient than classical camera architectures, permitting thus longer operation time in wireless environments.

In this paper, we consider CS recovery of multiview image and multiview video sequences wherein we assume that each frame in each view is acquired directly in a reduced dimensionality via a CS-based image sensor. Moreover, knowing that multiple images within a multiview dataset are highly correlated, we exploit this correlation in the CS reconstruction process. Specifically, in the case of multiview images, we capitalize on disparity estimation (DE) and disparity compensation (DC) between adjacent views to provide a prediction of the current image to be reconstructed. The DE/DC prediction drives a residual-based CS reconstruction of the current view. This process is further extended to the case of multiview video, wherein DE/DC is coupled with motion estimation (ME) and motion compensation (MC) such that predictions for the current view are created both from adjacent views as well as from temporally neighboring frames. Experimental results show that the incorporation of DE/DC and, in the case of multiview video, ME/MC, into the CS

recovery process provides a significant increase in reconstruction quality as compared to the straightforward CS reconstruction of each individual view independently of the others.

Different from our own preliminary work [31–33] wherein we proposed the general use of simple block-based DE/DC within a CS framework for multiview images, this present work focuses on DE/DC methods driven by optical flow in a complete CS acquisition and reconstruction system. In this work, we also propose the CS recovery of multiview video by simultaneous use of DE/DC and ME/MC and present a comprehensive and complete discussion on the prediction-based CS-reconstruction approach. We note that this work expands on preliminary results for CS video initially reported in [13]; again, the more thorough discussion presented here employs optical flow for DE/DC and reports results in greater detail. Additionally, two alternative approaches to prediction-aided CS multiview recovery are considered in order to demonstrate the efficacy of our proposed prediction strategy for multiview signal recovery in several reconstruction settings.

The remainder of this paper is organized as follows. First, in Section 2, we briefly overview CS theory and the prior use of CS for image reconstruction. Then, in Section 3, we describe the general procedure of CS recovery based on signal prediction. In Section 4, we describe the specific algorithms we use for multiview-image and multiview-video reconstruction, and, in Section 6, we present a battery of experimental results that evaluate the performance of these prediction-based recoveries. Finally, we make several concluding remarks in Section 7.

2 Background

One of the main advantages of the CS paradigm is the very low computational burden placed on the signal-acquisition process, which effectively acquires the desired signal directly in a reduced dimensionality. Specifically, CS requires only the projection of the signal $x \in \mathbb{R}^N$, which is sparse in some transform basis Ψ , onto some measurement basis Φ of size $N \times M$ where $M \ll N$. The result of this signal-acquisition process is the M -dimensional vector of measurements, $y = \Phi x$. Φ is often chosen to be a random matrix because it satisfies the incoherency and isometry requirements of CS reconstruction for any structured signal transform Ψ with high probability [5]. For simplicity, we assume Φ is orthonormal ($\Phi^T \Phi = I$). We define the subsampling rate, or *subrate*, of the CS acquisition process to be M/N .

This computationally-light signal-acquisition procedure offloads most the computation associated with CS onto the signal-reconstruction process. Because the inverse of the projection $\hat{x} = \Phi^{-1} y$ is ill-posed, we cannot directly solve the inverse problem to find the original signal from the given measurements.

As x is assumed to be sparse with respect to some transform basis Ψ , the reconstruction process entails the production of a sparse set of significant transform coefficients, $\hat{x} = \Psi x$. The recovery procedure searches for \hat{x} with the smallest l_0 norm that is consistent with the observed y ; i.e.,

$$\hat{x} = \arg \min_{\hat{x}} \|\hat{x}\|_0, \quad \text{such that } y = \Phi \Psi^{-1} \hat{x}, \quad (1)$$

where Ψ^{-1} represents the inverse transform. Due to NP-completeness of this l_0 optimization, alternative procedures have been proposed for sparse reconstructions using l_1 or l_2 -norms.

Indeed, one resorts to using any of a number of CS-reconstruction approaches that have appeared in recent literature and include convex programming [8], gradient-descent [11], greedy-pursuit [34], and iterative-thresholding [2] implementations, for solving the l_1 or l_2 relaxations of (1). Unfortunately, many of these CS reconstruction algorithms tend to be rather computationally complex, and large-dimensionality CS reconstructions such as those needed for natural images significantly exacerbate the problem.

To face this issue, Gan [15] proposed partitioning natural-image CS acquisition into distinct blocks, tantamount to imposing a block-diagonal structure on Φ . A Wiener smoothing step was used within an iterative-threshold recovery to remove blocking artifacts resulting from the discontinuous nature of the partitioning. One other advantage to this method is that the storage requirement for Φ at the signal-acquisition platform is also reduced by orders of magnitude, as the same projection can be used for each block within the image. This method was extended in [25] by making use of directional transforms for sparsity basis Ψ coupled with a thresholding based on statistical wavelet models. The resulting algorithm was called block compressed sensing with smoothed projected Landweber (BCS-SPL) reconstruction in [25].

Another popular approach to the CS reconstruction of images, total variation (TV) minimization [6, 7, 30], uses piece-wise smooth characteristics of natural signals to great effect. Instead of finding the sparsest solution within the domain of transform Ψ , TV minimization finds the “smoothest” solution within the space of possible solutions. Anisotropic TV minimization makes use of the l_1 norm to enforce sparsity upon the gradient of the solution, creating a penalty function of the form

$$TV(x) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|. \quad (2)$$

Using the penalty above, the CS recovery problem can be stated as

$$\hat{x} = \arg \min_x \|y - \Phi x\|_2 + \lambda TV(x). \quad (3)$$

TV minimization has been widely used in CS recovery; however, to date, many of the methods used to solve (3) (such as second-order-cone programs using interior-point or log-barrier methods), are too computationally complex to be of practical use except for exceedingly small image sizes. Indeed, the cost of reconstruction using such approaches has prevented the use of TV minimization for CS reconstruction in many applications where large volumes of data must be processed, such as multiview image and video. Other, more computationally efficient, approaches to solving (3) have been proposed, such as iterative soft thresholding [1] and alternating minimization [41].

In [21] an augmented Lagrangian formulation coupled with an alternating direction algorithm (TV-AL3) was proposed for solving (3) for both its anisotropic and isotropic forms. The TV-AL3 method retains the same reconstruction accuracy afforded by TV minimization for CS image recovery while decreasing the

computation time by orders of magnitude over other techniques. Because of the decreased computational burden, TV-AL3 permits us to make use of high-quality TV minimization for the reconstruction of multiview data. Below, we use TV-AL3, along with BCS-SPL, to demonstrate the efficacy of our proposed prediction strategy for multiview signal recovery in several reconstruction settings.

3 CS reconstruction using signal prediction

In traditional source coding, it has long been known that signal prediction can play a significant role in increasing signal compressibility. For example, DPCM methods have been used to code many different forms of data and have, in particular, seen extensive use in video-coding algorithms. In fact, frame prediction comprises the core functionality of many video-coding standards such as MPEG-2 and H.264/AVC. By creating a frame prediction from highly correlated, temporally neighboring frames by way of some form of ME and MC, a temporally decorrelated and compressible residual frame can be calculated by the subtraction of the predicted frame from the original at the encoder.

Whereas in traditional video coding, ME/MC is used at the encoder side in order to produce a highly compressible residual, in the CS paradigm, prediction must take place at the opposite end of the system, i.e., in the signal-reconstruction process. Decorrelation achieved via prediction aids CS recovery by increasing the compressibility of the signal. Here, as is common in CS literature, signal “compressibility” refers specifically to the case in which coefficient magnitudes exhibit a power-law decay in some transform domain [3]. The more compressible a signal is, the closer the sparse reconstruction resulting from CS recovery will approximate it. (e.g., see Theorem 3.2 of [4]).

When one is considering the CS reconstruction of multiple images—either multiple views or frames in a set of multiview video sequences—it is desirable to include some form of inter-image decorrelation into the CS-recovery process. Many techniques have approached this problem through 3D transforms, treating a collection of frames as a volume (e.g., [39, 40]). This method, however, is problematic because of the limited exploitation of frame-to-frame object motion; this is the case in which the multiple frames are either temporally consecutive video frames or different views from a dataset of multiview images. The alternative is decorrelation that takes advantage of motion or disparity between the frames and views.

Specifically, suppose we have some frame x_d for which we calculate a prediction x_p from other frames in the collection of frames. Here, our collection of frames may be a temporal set from a video sequence, or a set of multiple views in a multiview dataset. In either case, we can substitute the original problem of CS reconstruction of x_d from its measurements $y_d = \Phi x_d$ with the recovery of the residual between x_d and its prediction x_p due to the linear nature of CS acquisition, i.e.:

$$r = \Phi r_d = \Phi(x_d - x_p) = y_d - \Phi x_p. \quad (4)$$

From here, we can see that it is possible to recover the residual frame r_d by CS reconstruction of the difference, r , between the original measurements (y_d) and the

projection of the prediction into the measurement domain. The final recovery of \hat{x}_d is then

$$\hat{x}_d = x_p + \text{CSRecovery}(r, \Phi). \quad (5)$$

If the prediction process producing x_p is reasonably accurate, the residual frame r_d should be more compressible than the original image x_d . This is demonstrated empirically for a video sequence in Fig. 1 wherein it is seen that the transform-coefficient magnitudes decay more quickly for a motion-compensated residual frame than for the original video frame.

In the sequel, we investigate the adaptation of this approach to the CS recovery of multiview images as well as multiview video. We propose to further improve the quality of the CS recovery in an iterative manner, by jointly reconstructing the frames and repeating the process described in (4) and (5) on higher fidelity images. This way, a more precise prediction x_p for the current frame is obtained. As the resulting image residual is more compressible with each iteration, a high-quality recovery \hat{x}_d of the current image is obtained at the end of the reconstruction cycle.

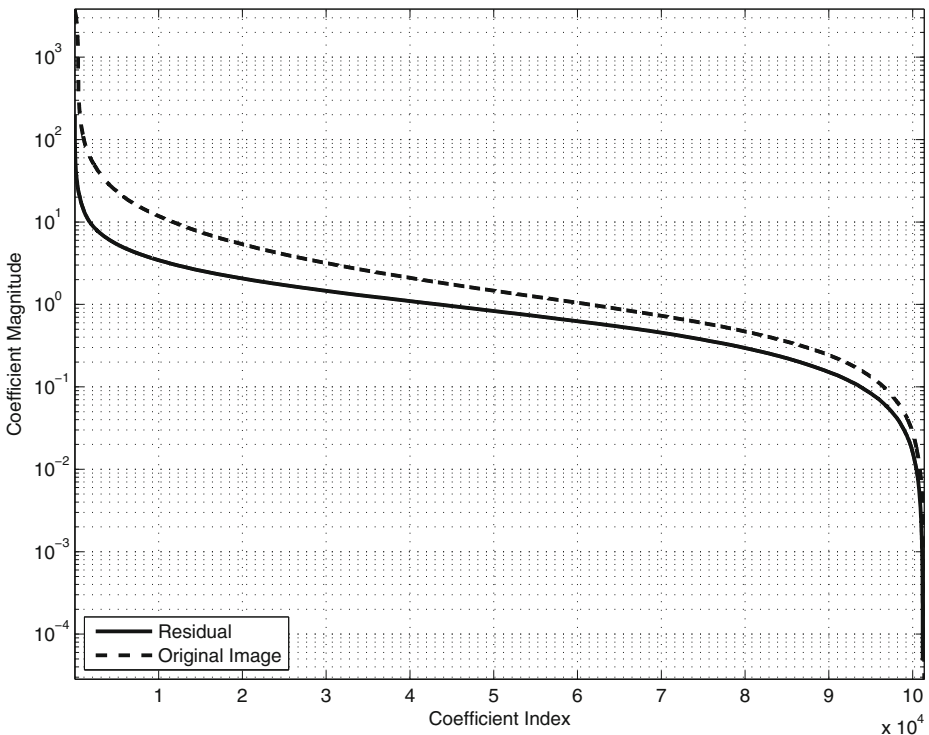


Fig. 1 Decay of the magnitudes of the transform coefficients for frame 1 of the “Foreman” video sequence as compared to that of the motion-compensated residual between frames 1 and 0. ME/MC is based on 16×16 blocks with quarter-pixel accuracy over a window of size 15×15 pixels; ME/MC is performed between the original frames of the sequence. The transform is a 4-level biorthogonal 9/7 DWT

4 CS reconstruction of multiview images and multiview video sequences

In the sequel, we adapt the prediction-driven CS reconstruction of (4) and (5) to the case of multiview images in which the collection of frames in question is a set of highly correlated images of a single subject taken from slightly different perspectives. Different from [13] wherein a simple block-based DE is performed, in this work, we propose DE using optical flow [23] to calculate dense disparity fields, along with warping DC to produce a high-accuracy prediction x_p of the current view x_d from adjacent views which are likely to be highly correlated with x_d . We first consider the reconstruction of a single view within a set of multiview images in Section 4.1 before extending the process into a multistage reconstruction of the entire set of multiview images in Section 4.2. Finally, in Section 4.3, we deploy the proposed multiview reconstruction on multiview video, incorporating both DE/DC as well traditional block-based ME/MC into the reconstruction process.

4.1 Single-view reconstruction

In order to provide a CS reconstruction of a single view within a set of multiview images, we couple a still-image recovery with a DE/DC-driven prediction process. We call the resulting algorithm DC-CS. In our test framework we consider two CS reconstruction methods, namely BCS-SPL of [13] and TV-AL3 of [21].

Initially, all the multiview images are CS recovered independently from one another; i.e.,

$$\hat{x}_d^{\text{init}} = \text{CSRecovery}(y_d, \Phi_d). \quad (6)$$

In the sequel, this first reconstruction will be referred to as “initial” recovery stage. Subsequently, our DC-CS algorithm is partitioned into two phases. In the first phase, an initial prediction x_p^{init} for the current view x_d is created by bidirectionally interpolating the closest adjacent views from the initial recovery,

$$x_p^{\text{init}} = \text{ImageInterpolation}(\hat{x}_{d-1}^{\text{init}}, \hat{x}_{d+1}^{\text{init}}), \quad (7)$$

where $\hat{x}_{d-1}^{\text{init}}$ and $\hat{x}_{d+1}^{\text{init}}$ are the “left” and “right” neighbors of x_d , respectively. In this interpolation, we use as references the reconstructions obtained in the initial stage, i.e. $\hat{x}_{d-1}^{\text{init}}$ and $\hat{x}_{d+1}^{\text{init}}$.

The image interpolation is performed as in [17]: the neighboring views $\hat{x}_{d-1}^{\text{init}}$ and $\hat{x}_{d+1}^{\text{init}}$ are firstly spatially low-pass filtered, then a classical forward block-matching DE is performed between them, which will be further refined in order to obtain a bidirectional view interpolation. Next, we calculate the residual r between the original observation y_d and the observation resulting from the projection of x_p using the *same* measurement matrix, Φ_d , i.e.:

$$r = y_d - y_p, \quad \text{s.t.} \quad y_p = \Phi_d x_p^{\text{init}}. \quad (8)$$

This residual then drives the CS reconstruction, $\hat{r} = \text{CSRecovery}(r, \Phi_d)$. Note that the use of the original Φ_d , used at the acquisition of y_d , is requested at this step, in

order to insure the correlation between the original observation and the one obtained following the prediction process y_p (i.e., the same linear combination given by Φ_d is applied to x_d and respectively, x_p^{init} , for obtaining y_d and respectively, y_p).

In the second phase, the reconstructed residual \hat{r} produces the reconstruction, i.e.:

$$\hat{x}_d = \hat{r} + x_p^{init}. \tag{9}$$

The prediction process then repeats, but this time DE/DC is used instead of interpolation (in the sequel, we denote by x_p the result of the DE/DC prediction). Specifically, DV_{d-1} and DV_{d+1} are the dense fields of left and right disparity vectors, respectively; these are obtained from DE applied to the current reconstruction, \hat{x}_d , of the current image and the left and right adjacent images. These disparity vectors subsequently drive the warping DC of the current image to produce the current prediction, x_p , and its corresponding residual, r , which are further used for obtaining the reconstruction \hat{x}_d .

In DC-CS reconstruction, each iteration of the second phase provides an incremental increase in recovery quality of the current view, thereby providing a better target to which to match the neighboring frames during the DE/DC stage. Indeed, the predictor x_p at iteration k of this second phase will be obtained by DC between the enhanced reconstruction \hat{x}_d (obtained after the $k - 1$ iteration) and its neighbors (i.e., the reference views, for which the quality does not change from one iteration to another, and for which the reconstruction is obtained by direct CS-recovery); the improvement of reconstruction quality at iteration k is due to the refinement of the disparity vectors, leading to a more sparse and smoother residual r , which is thus better reconstructed (leading to an enhanced view reconstruction, \hat{x}_d , at the end of the k -th iteration). This process could conceivably be repeated until \hat{x}_d converges to a final solution.

As such, the second phase of the algorithm may be repeated as long as the difference, D_r , between two successive residual energies is higher than threshold ϵ ,

$$D_r^k = E_r^{k-1} - E_r^k > \epsilon, \tag{10}$$

where $E_r^k = \|r^k\|_2^2$ is the energy of the prediction residual at iteration k . In order to speed-up the recovery process, one can limit the number of iterations; we have

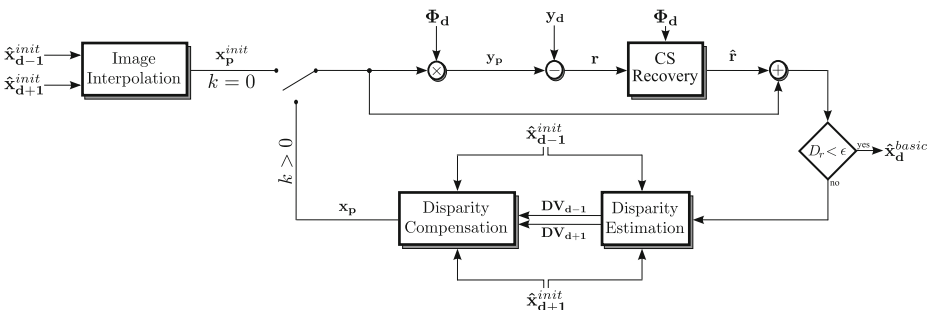


Fig. 2 The DC-CS basic stage reconstruction procedure for a single view

found that using $K = 3$ iterations provides quite adequate convergence (i.e., we have observed that D_r^k decreases very quickly, being less than 0.05 after the first three iterations, for most of the multiview image sets). The complete algorithm is described in Fig. 2.

We note that there exists a variety of DE/DC methods of varying sophistication, some producing high-quality predictions driven by depth or parallax information between views. Any of these DE/DC strategies could be used in DC-CS by simply placing them in the DE and DC blocks in Fig. 2. In [13, 31–33], a simple block-based DE/DC procedure (similar to traditional ME/MC) was used due to its decreased computational burden and complexity of implementation. In the present work, we employ a dense DE/DC method, namely the optical-flow algorithm proposed in [23], which provides much more accurate view predictions at each stage of recovery, therefore improving the final accuracy of multiview images recovered using the DC-CS framework as compared to a simple block-based compensation. For example, in Fig. 3 we show a comparison between optical flow and block-matching DE/DC (as used originally in [13, 31–33]) in our proposed image-recovery framework. These results demonstrate that recovery performance can indeed be aided by the use of more sophisticated DE/DC strategies.

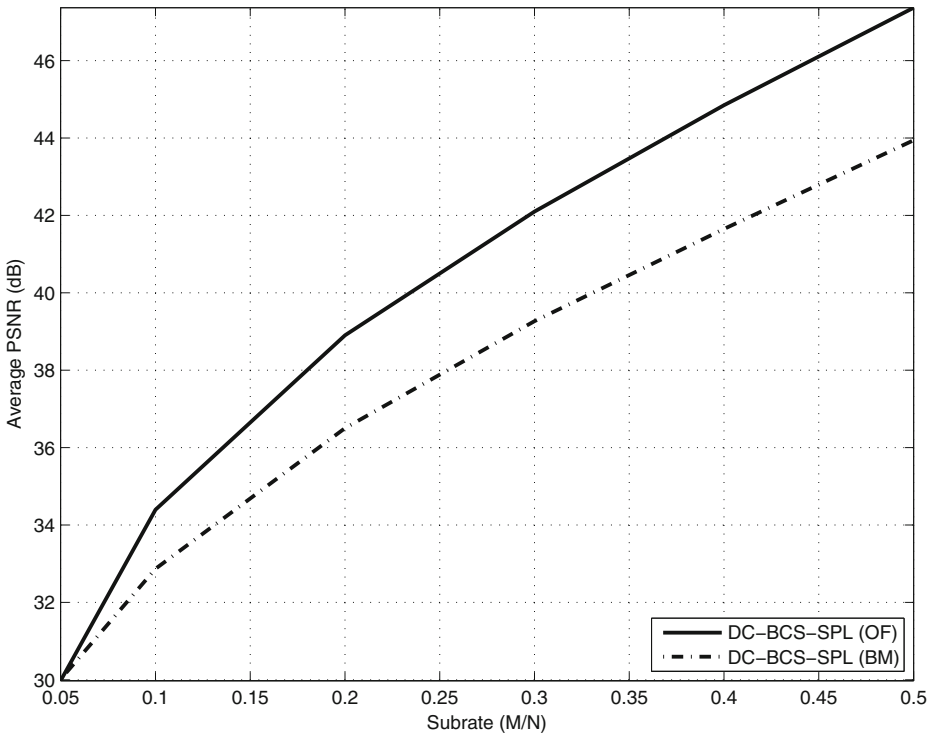


Fig. 3 Comparison between DC-BCS-SPL reconstruction using optical flow (OF) and block matching (BM) to accomplish DE/DC for the “Baby” multiview-image dataset

In some contexts, when global sampling is available to the acquisition device, TV-based image recovery techniques can provide very accurate recovery performance. In our prior work, [31–33], we focused exclusively on the application of BCS to multiview image acquisition due to its practical implementation advantages in terms of cost to the acquisition device, and also for its reduced computational reconstruction costs [13]. However, advances in the design and implementation of functionally derived projection matrices, as is done with structured random matrices (SRM) [16, 29], can allow for efficient implementation of global CS image-sampling strategies. Additionally, algorithmic advances in the calculation of TV minimization have decreased the computational cost for recovering high-dimensional signals by orders of magnitude [21]. These developments allow one to avoid the accuracy-cost trade-off inherent in the BCS framework, providing significant gains in reconstruction accuracy. We propose the use of TV-based view reconstruction, along with BCS-SPL, to show that the multistage multiview-recovery framework we propose can effectively exploit inter-view correlation regardless of the CS image-acquisition and recovery framework used.

4.2 Multistage reconstruction of multiview images

In the previous section, we described the CS recovery of a single frame within a multiview dataset, given a disparity-compensated prediction made from adjacent views. It was implicitly assumed that these left and right views themselves had already been reconstructed via some process. However, as these left and right views are also recovered from CS measurements, the performance of the recovery of any given view is dependent upon the quality of the views used as references. The higher the distortion present in the reference frames, the higher the distortion will be in the recovery of the given view of current interest.

To reconstruct the entire multiview dataset from individual CS measurements of each of the constituent frames, we propose a multistage recovery process. In the first, or initial, stage, each image in the multiview set is reconstructed individually from the received set of measurements using a CS recovery method. In the second stage (the “basic” stage), each image is reconstructed using the DC-CS procedure of Fig. 2 with the left and right reference views as obtained from the preceding initial stage.

Subsequently, one or more (e.g., R) refinement stages are performed. A refinement stage of the algorithm is simply the repetition of the second phase of the basic stage (i.e., the phase wherein DE/DC-based prediction is used for triggering the residual), in which the references used in the current DE/DC are given by the recoveries obtained at the previous stage (i.e., basic stage, for the first refinement or the last refinement level, for subsequent refinement stages). The stages could conceivably be repeated until there is no significant difference between consecutive passes; however, in our experimental framework described in Section 6, we consider up to $R = 4$ refinement stages in order to minimize the overall computational complexity of the reconstruction.

The pseudo-code for the basic and refinement stages is described in Algorithms 2 and 3 (Algorithm 1 describes only the prediction/residual based reconstruction part, used in both basic and refinement stages). Additionally, Algorithm 4 describes how each stage is used together to form the full DC-CS multiview image recovery system.

In edge cases where bidirectional reference views are not available for DE/DC or interpolation, unidirectional references may be substituted. In Algorithms 1–4, the notation $\{\cdot\}$ when used in conjunction with index d refers to a set of values over the index d . For example, $\{y_d\}$ refers to the set of measurements at each view, $\{y_1, y_2, \dots, y_{\text{NumViews}}\}$.

Algorithm 1 Prediction-based View Recovery

Input: $\hat{x}_d^0, y_d, \Phi_d, \hat{x}_{d-1}, \hat{x}_{d+1}, K, \epsilon$
 $\hat{r}_d^0 = 0$
for all $k \in \{1, 2, \dots, K\}$ **do**
 $DV_{d-1}^k = \text{DisparityEstimation}(\hat{x}_d^{k-1}, \hat{x}_{d-1})$
 $DV_{d+1}^k = \text{DisparityEstimation}(\hat{x}_d^{k-1}, \hat{x}_{d+1})$
 $x_p^k = \text{DisparityCompensation}(\hat{x}_{d-1}, \hat{x}_{d+1}, DV_{d-1}^k, DV_{d+1}^k)$
 $\hat{r}_d^k = \text{CSRecovery}(y_d - \Phi_d x_p^k, \Phi_d)$
 $\hat{x}_d^k = \hat{x}_d^{k-1} + \hat{r}_d^k$
 if $\|\hat{r}_d^k\|_2^2 - \|\hat{r}_d^{k-1}\|_2^2 < \epsilon$ **then**
 return \hat{x}_d^k
 end if
end for
return \hat{x}_d^K

Algorithm 2 Basic Stage Recovery

Input: $\{\hat{x}_d^{\text{init}}\}, \{y_d\}, \{\Phi_d\}, K, \epsilon$
for all $d \in \{1, 2, \dots, \text{NumViews}\}$ **do**
 $x_{p,d}^{\text{init}} = \text{ImageInterpolation}(\hat{x}_{d-1}^{\text{init}}, \hat{x}_{d+1}^{\text{init}})$
 $\hat{r}_d = \text{CSRecovery}(y_d - \Phi_d x_{p,d}^{\text{init}}, \Phi_d)$
 $\hat{x}_d = \hat{x}_{p,d}^{\text{init}} + \hat{r}_d$
 $\hat{x}_d^{\text{basic}} = \text{PredictionViewRecovery}(\hat{x}_d, y_d, \Phi_d, \hat{x}_{d-1}^{\text{init}}, \hat{x}_{d+1}^{\text{init}}, K, \epsilon)$
end for
return $\{\hat{x}_d^{\text{basic}}\}$

Algorithm 3 Refinement Stage Recovery

Input: $\{\hat{x}_d^{\text{basic}}\}, \{y_d\}, \{\Phi_d\}, R, K, \epsilon$
 $\hat{x}_d^0 = \hat{x}_d^{\text{basic}}, \quad \forall d$
for all $i \in \{1, 2, \dots, R\}$ **do**
 for all $d \in \{1, 2, \dots, \text{NumViews}\}$ **do**
 $\hat{x}_d^i = \text{PredictionViewRecovery}(\hat{x}_d^{i-1}, y_d, \Phi_d, \hat{x}_{d-1}^{i-1}, \hat{x}_{d+1}^{i-1}, K, \epsilon)$
 end for
end for
return $\{\hat{x}_d^R\}$

Algorithm 4 Full DC-CS Recovery

Input: $\{y_d\}, \{\Phi_d\}, R, K, \epsilon$
 $\hat{x}_d^{\text{init}} = \text{CSRecovery}(y_d, \Phi_d), \quad \forall d$
 $\{\hat{x}_d^{\text{basic}}\} = \text{BasicStage}(\{\hat{x}_d^{\text{init}}\}, \{y_d\}, \{\Phi_d\}, K, \epsilon)$
 $\{\hat{x}_d^{\text{refine}}\} = \text{RefinementStage}(\{\hat{x}_d^{\text{basic}}\}, \{y_d\}, \{\Phi_d\}, R, K, \epsilon)$
return $\{\hat{x}_d^{\text{refine}}\}$

We note that, for each view, a different random measurement matrix Φ_d is used, and the information retained in the different projections has a high probability of being complementary. Knowing that each view is highly correlated, the performance gains from the refinement iterations are also due to complementary, highly correlated information along the disparity axis.

4.3 Reconstruction of multiview video sequences

The multistage DC-CS procedure described above reconstructs an entire set of multiview images; however, the algorithm can be easily extended for use with multiview video in which we have multiple time samples of each view. To do so, we perform predictions not only along the disparity (or view) axis, but also along the temporal axis via ME/MC, as illustrated in Fig. 4. The algorithm is partitioned into three phases, much like DC-CS for multiview images. In the initial stage, each frame in each view in the multiview video is reconstructed individually from the received set of measurements using a CS recovery method.

In the second stage, for each image x_d^t at time t and view d , a prediction is created by directionally interpolating the CS reconstructions of the closest frames in both temporal and disparity directions, using the procedure in [17] for this directional interpolation, applied twice (once on the temporal direction, once on the disparity axis). These four neighboring views/frames— $x_{d-1}^t, x_d^{t-1}, x_{d+1}^t$, and x_d^{t+1} —are first spatially lowpass filtered, then a classical forward block-matching DE or ME is performed between them, which is further refined in order to obtain a bidirectional view/temporal interpolation.

The initial predictor x_p^{init} used for the image compensation is obtained by averaging the interpolations on temporal and disparity axes; i.e.,

$$x_p^{\text{init}} = 0.5 \cdot \text{ImageInterpolation}(\hat{x}_d^{t-1}, \hat{x}_d^{t+1}) \\ + 0.5 \cdot \text{ImageInterpolation}(\hat{x}_{d-1}^t, \hat{x}_{d+1}^t). \quad (11)$$

Next, we compute the residual between the measurements and the projection of the predicted frame; i.e:

$$r = y_d^t - \Phi_d^t x_p^{\text{init}}. \quad (12)$$

This residual in the measurement domain is then reconstructed using CS (i.e., $\hat{r} = \text{CSRecovery}(r, \Phi_d^t)$) and added back to the prediction to generate a reconstruction, $\hat{x}_d^t = \hat{r} + x_p^{\text{init}}$.

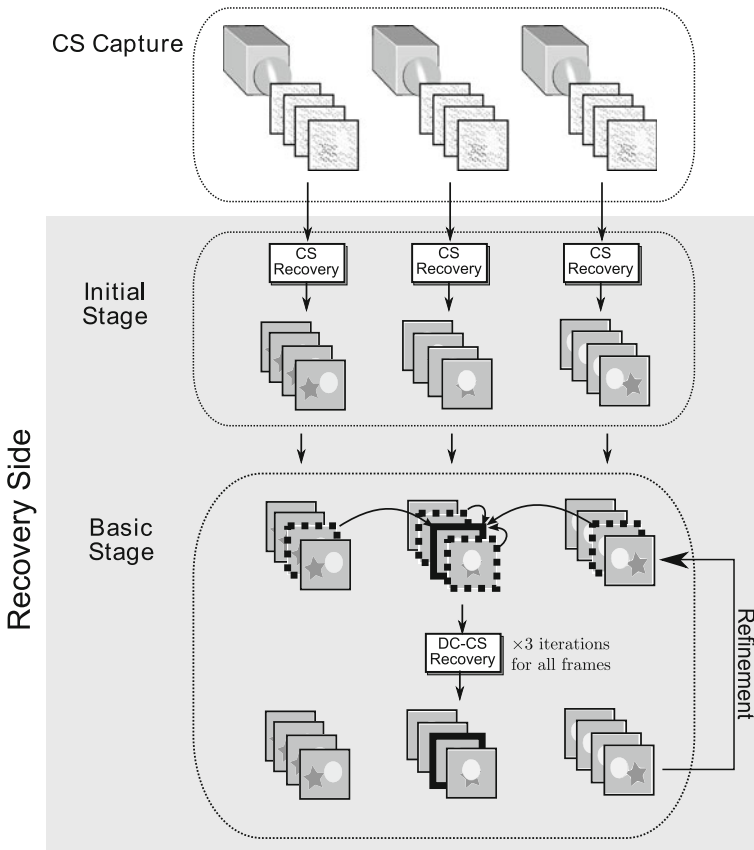


Fig. 4 The multistage DC-CS reconstruction framework using ME/MC and DE/DC for multiview video

\hat{x}_d^t is further refined in the basic stage by calculating fields of disparity vectors, (DV_{d-1}^t, DV_{d+1}^t) , and temporal motion vectors, (MV_d^{t-1}, MV_d^{t+1}) . These vectors then drive the compensation to form both view and temporal predictions of the current frame from the neighboring frames. The final prediction x_p (i.e., using DE/DC and ME/MC) is obtained by averaging these four predictions (as in (13)), and the procedure is repeated.

$$x_p = \frac{1}{4} [MC(\hat{x}_d^{t-1}, MV_d^{t-1}) + MC(\hat{x}_d^{t+1}, MV_d^{t+1}) + DC(\hat{x}_d^t, DV_{d-1}^t) + DC(\hat{x}_d^t, DV_{d+1}^t)]. \tag{13}$$

Similarly to the DC-CS phase used for the reconstruction of the multiview images, in the “basic” stage the quality of \hat{x}_d^t is successively improved at each iteration by refining both the motion and disparity vectors at each step, producing better predictions and therefore more compressible residuals which are more accurately recovered. For reducing the complexity associated to this iterative recovery process,

we iterate the x_d^t reconstruction (keeping the references unchanged) three times in our implementation framework. Note however that the residual energy difference criterion in (10) can be used for stopping the iteration process.

Subsequently, one or more refinement stages are performed. A refinement stage of the algorithm is simply the repetition of the basic stage as described above with the results from the second stage substituted for the references used to drive the compensated-CS reconstruction. The stages could conceivably be repeated until there is no significant difference between consecutive passes; however, as for the multiview image sets, we use only four refinement stages.

5 Other approaches to CS reconstruction of multiview imagery

To the best of our knowledge, there have been only several approaches proposed specifically in prior literature for the CS reconstruction of multiview images, and none for multiview video other than our own preliminary work in [13]. The most common approach is to assume that each image of the multiview set has been sampled via a CS measurement process independently of the other images (as we have done in our DC-CS framework); the reconstruction process then attempts to recover all the images of the multiview set jointly, exploiting the sparsity common to the disparate views.

For example, [22] reconstructs the multiview image set jointly, enforcing sparsity not only in each image separately but also in the view-to-view difference images between neighboring views. Chen and Frossard [9] adopts a somewhat similar strategy of joint reconstruction, except that the correlation between neighboring views is captured in a model more sophisticated than a simple sparse difference image. Specifically, [9] represents neighboring-view correlation with a local geometric transformation over an overcomplete structured dictionary. Joint reconstruction is also central to [26, 38] wherein the view-to-view correlation is modeled by requiring the reconstructed views to lie along a low-dimensional manifold.

The joint reconstruction proposed in [9, 22, 26, 38] is problematic as the computation burdens are likely to be significant, particularly so as the number of views increases. Indeed, [9, 22] consider the reconstruction of multiview datasets with only one to three different views. Wakin [38] and Park and Wakin [26] also suffer from the burden of having to formulate an explicit model for the low-dimensional manifold describing the multiview set. While [26, 38] primarily focus on the relative simple “far-field” problem of overlapping fields of view of a single large image, the more general multiview scenario involving parallax and occlusion is significantly more difficult to handle in practice. While [26] suggests that a “plenoptic manifold” may accommodate such general “near-field” problems, it is far from clear that such an approach is feasible in practice.

As the number of prior methods designed specifically for multiview imagery is somewhat limited, the similar problem of CS reconstruction of video has received more attention in recent literature. A number of algorithms for video CS reconstruction were developed for the particular case of dynamic magnetic resonance imagery (MRI). This type of image sequence tends to have less motion, and the motion tends to be less of a strictly translational nature, than does video acquired from natural photographic scenes. However, dynamic-MRI algorithms may be better suited to the

multiview scenario in which consecutive views differ primarily in disparity due to a changing viewpoint. Initial work adopted the volumetric reconstruction employed originally as in [39, 40]—for example, [14] reconstructs a dynamic MRI volume using a temporal Fourier transform coupled optionally with a spatial wavelet transform as a 3D sparsity basis.

However, given the computational issues with reconstructing volumes (similar to those surrounding joint multiview reconstruction), most CS reconstructions for video have focused on frame-based recovery that exploits the fact that successive frames are strongly correlated. Various strategies have been adopted to handle frame-to-frame correlation. For example, Vaswani [35, 36], Vaswani and Lu [37], Lu and Vaswani [24], Qiu et al. [28] have proposed a variety of related approaches for the CS reconstruction of dynamic MRI data. Fundamental to several of these techniques [24, 28, 36] is the general strategy of residual reconstruction from a prediction of the current frame as in (5); the key difference from the work proposed here is that, rather than using a ME/MC- or DE/DC-based prediction, Vaswani et al. employ a least-squares [36] or Kalman-filtered [28] prediction. These predictions are driven by an explicit sparsity pattern for the current frame; the techniques attempt to track this sparsity pattern as it evolves from frame to frame. It is assumed that the sparsity pattern evolves slowly over time, an assumption that may not hold in general video with arbitrary object motion. However, the “Modified-CS-Residual” algorithm of [24] is a prominent benchmark in the literature for gauging CS-reconstruction performance for not only dynamic MRI but also video as well.

Another reconstruction algorithm driven by prediction residuals was considered in [18, 19]. This algorithm, called k-t FOCUSS in [18], assumes that there exist one or two key frames obtained through some separate means, and then CS reconstruction is driven by residuals between each intervening non-key frame and a block-based bidirectional motion-compensated prediction from each of the key frames (or a single unidirectional prediction in the event that only one key frame is available).

Although there exist a number of other CS reconstruction algorithms for video in the literature, none of these other than k-t FOCUSS and Modified-CS-Residual have, to our knowledge, implementations readily available at the time of this writing; the same can be said for all the existing algorithms designed specifically for multiview imagery. As a consequence, in the experimental results of the next section, we focus on comparing DC-BCS-SPL to k-t FOCUSS and Modified-CS-Residual as well as several straightforward “intraframe” strategies that reconstruct each view independently.

6 Experimental results

In the following, we evaluate the performance of the proposed prediction-based reconstructions for both multiview image sets and multiview video sequences.

6.1 Multiview images

In order to observe the effectiveness of DC-CS recovery, we first evaluate its performance at each stage of reconstruction—i.e., at the initial stage, at the basic stage, and at the refinement stage, as defined in Section 4.2.

Table 1 PSNR performance (in dB) at each CS reconstruction stage of the proposed method and for two multiview image datasets

Algorithm	Subtrate					
	0.05	0.1	0.2	0.3	0.4	0.5
Bowling						
DC-TV						
7th Refinement	38.177	41.968	47.208	50.148	52.398	53.863
6th Refinement	38.160	41.970	47.200	50.140	52.391	53.860
5th Refinement	38.180	41.977	47.180	50.113	52.315	53.862
4th Refinement	38.135	41.916	47.034	50.092	52.224	53.864
3rd Refinement	38.086	41.717	46.764	49.909	52.133	53.803
2nd Refinement	37.824	41.447	46.331	49.534	51.879	53.617
1st Refinement	37.343	40.892	45.694	48.997	51.466	53.377
Basic	36.084	39.480	44.037	47.323	49.986	52.137
Initial	34.489	37.477	41.449	44.557	47.111	49.517
DC-BCS-SPL						
7th Refinement	32.341	35.963	40.975	44.700	47.683	50.036
6th Refinement	32.281	35.855	40.919	44.695	47.681	50.038
5th Refinement	32.196	35.627	40.831	44.680	47.679	50.040
4th Refinement	32.003	35.240	40.687	44.605	47.677	50.039
3rd Refinement	31.738	34.988	40.392	44.508	47.654	50.064
2nd Refinement	31.352	34.581	39.997	44.204	47.409	49.887
1st Refinement	30.695	33.958	38.689	43.092	46.575	49.301
Basic	29.912	33.307	37.662	41.239	44.171	46.722
Initial	29.094	32.862	36.304	38.916	41.020	43.129
Baby						
DC-TV						
7th Refinement	34.416	37.974	43.263	46.763	49.882	51.934
6th Refinement	34.344	37.981	43.228	46.758	49.838	51.870
5th Refinement	34.270	37.907	43.153	46.711	49.719	51.708
4th Refinement	34.102	37.745	42.933	46.589	49.430	51.547
3rd Refinement	33.658	37.305	42.389	46.197	49.072	51.391
2nd Refinement	33.230	36.678	41.699	45.431	48.492	50.977
1st Refinement	32.553	35.833	40.599	44.376	47.494	50.187
Basic	31.376	34.273	38.543	41.984	45.033	47.838
Initial	29.970	32.237	35.609	38.374	40.934	43.474
DC-BCS-SPL						
7th Refinement	30.013	34.729	38.953	42.193	44.927	47.622
6th Refinement	29.998	34.731	38.964	42.192	44.930	47.609
5th Refinement	30.001	34.672	38.958	42.158	44.905	47.563
4th Refinement	29.979	34.398	38.901	42.099	44.849	47.368
3rd Refinement	29.850	34.199	38.723	41.968	44.734	47.289
2nd Refinement	29.427	33.824	38.335	41.591	44.371	46.952
1st Refinement	29.006	33.223	37.558	40.782	43.539	46.187
Basic	28.079	32.065	35.794	38.598	41.060	43.426
Initial	26.854	30.746	33.603	35.685	37.541	39.307

The bold values correspond to the maximal PSNR (dB) reached for a given subtrate and image

In our experiments, we test two versions of DC-CS recovery—DC-BCS-SPL and DC-TV—which make use of the BCS-SPL¹ and TV-AL3² implementations of image

¹<http://www.ece.msstate.edu/~fowler/BCSSPL>

²<http://www.caam.rice.edu/~optimization/L1/TVAL3/>

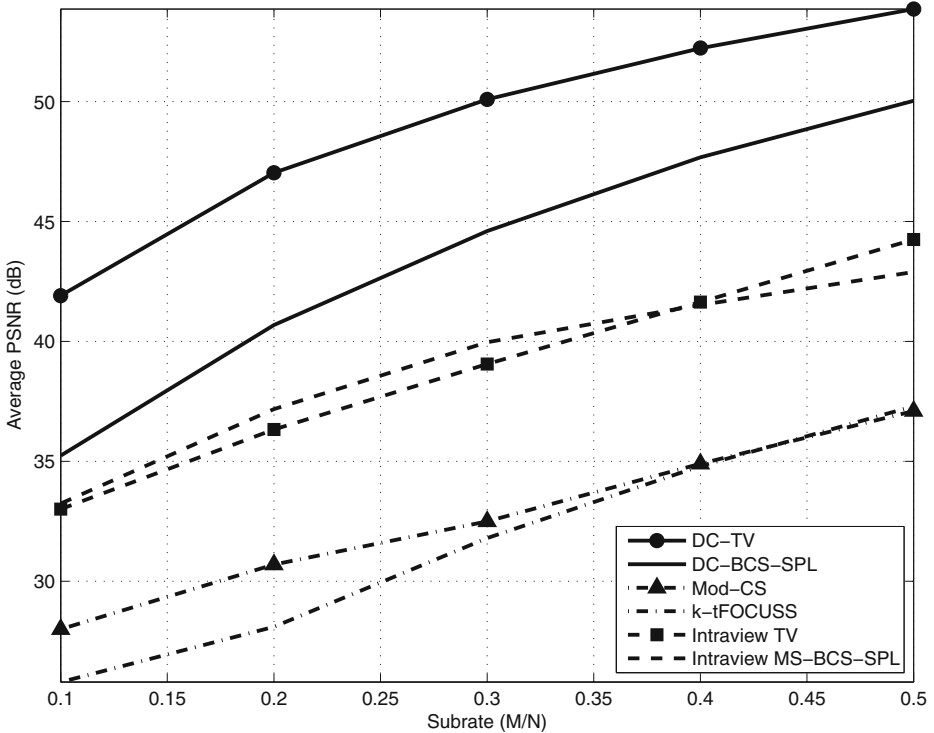


Fig. 5 Performance of various CS reconstruction algorithms for the “Bowling” multiview-image dataset

recovery, respectively. We note that BCS-SPL and TV-AL3 correspond to two different approaches to CS measurement—BCS-SPL relies on an explicitly defined block-based projection matrix, while TV-AL3 makes use of global, functionally generated SRMs [16, 29] for projection. For DC-BCS-SPL, we use a dual-tree discrete wavelet transform (DDWT) [20] with six levels of decomposition for the sparse representation basis, Ψ ; we note that the performance of the DDWT within the BCS-SPL framework was found to be among the best of the transforms investigated in [25]. For BCS sampling, a block size of 64×64 pixels is used. Different from [13] wherein a simple block-based approach is used to calculate the DE/DC view predictions, here we use the optical-flow implementation³ of [23]. It should be noted that, due to the variation in reconstruction quality that results from the random nature of CS measurement, all results are averaged over five independent trials. For multiview image data, we use the sample multiview images from the Middlebury stereo-image database.⁴ We consider grayscale versions of the first five views of each multiview image set.⁵ As the BCS-SPL recovery is block-based, the image resolution

³<http://people.csail.mit.edu/ceili/OpticalFlow/>

⁴<http://cat.middlebury.edu/stereo/data.html>

⁵Five $555 \times 626 \times 3$ multiview image sets: Aloe, Baby3, Bowling1, Plastic, and Monopoly

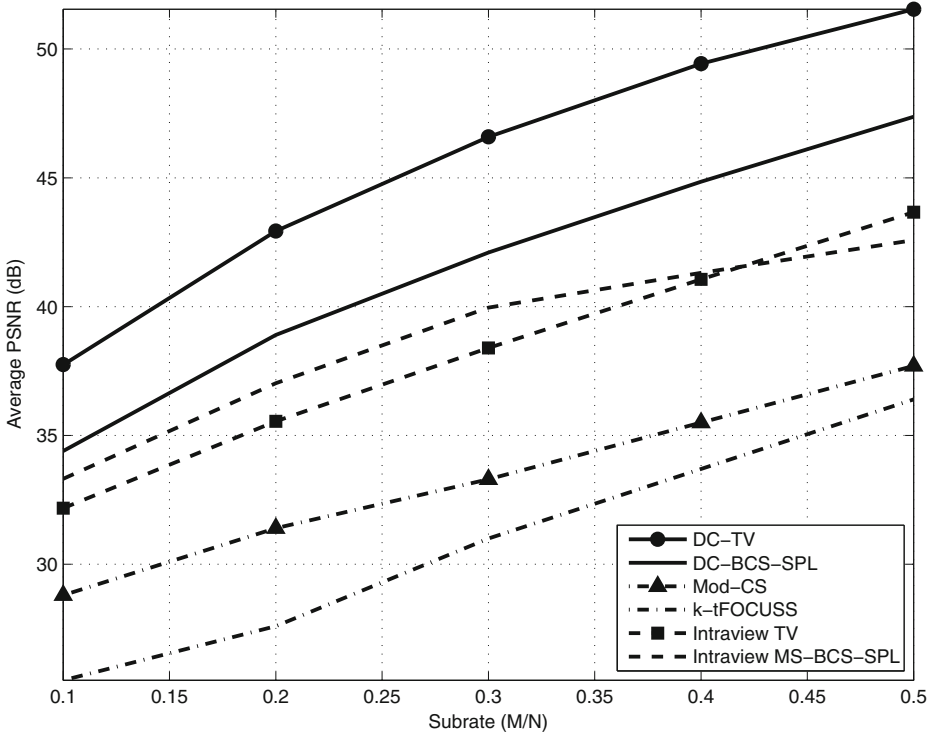


Fig. 6 Performance of various CS reconstruction algorithms for the “Baby” multiview-image dataset

should be a multiple of the BCS sampling block, therefore the views are further resized to 512×512 pixel resolution.

In this experimental framework, we evaluate reconstruction performance in terms of peak signal-to-noise ratio (PSNR) obtained for a range of subrates, M/N , with each view of the multiview dataset being acquired using the same subrate; we report the average PSNR obtained across all views for each multiview dataset. Detailed results are given for both the “Bowling” and “Baby” datasets in Table 1. In this table, we see that the incorporation of DE/DC-based prediction into the reconstruction process, as it occurs in the basic stage of Algorithm 2, provides a significant increase in reconstruction quality as opposed to the independent reconstruction of each view (the initial stage). Furthermore, each refinement stage (i.e., increasing R in Algorithm 3) further improves the reconstruction quality, with more pronounced gains for higher subrates. As mentioned in Section 4.2, in our framework, we have considered $R = 4$ refinement stages. It can be observed from Table 1 that, after the fourth refinement iteration, only small gains in performance are obtained. This empirical observation allows us to decrease reconstruction time by halting refinement at this point. Additionally, it can be seen that the DC-TV reconstruction offers the better reconstruction accuracy in comparison with DC-BCS-SPL at every recovery stage. These performance gaps can be attributed to the power of the TV prior for the recovery of natural images, which can provide a much more accurate

Table 2 PSNR performance (in dB) of various CS reconstruction algorithms for several multiview datasets

Algorithm	Substrate				
	0.1	0.2	0.3	0.4	0.5
Aloe					
Multistage DC-TV	29.0	33.0	36.8	40.4	44.1
Multistage DC-BCS-SPL	28.6	32.4	35.3	38.0	40.7
Modified-CS-Residual	25.3	27.3	29.1	30.9	32.8
k-t FOCUSS	22.3	24.5	27.7	29.8	32.1
Intraview MS-BCS-SPL	27.8	30.1	33.0	33.7	34.5
Intraview TV	25.7	27.6	29.2	30.7	32.4
Baby					
Multistage DC-TV	37.7	42.9	46.6	49.4	51.5
Multistage DC-BCS-SPL	34.4	38.9	42.1	44.8	47.4
Modified-CS-Residual	28.8	31.4	33.3	35.5	37.7
k-t FOCUSS	25.5	27.6	31.0	33.7	36.4
Intraview MS-BCS-SPL	33.3	37.0	40.0	41.3	42.6
Intraview TV	32.2	35.5	38.4	41.1	43.7
Bowling					
Multistage DC-TV	41.9	47.0	50.1	52.2	53.9
Multistage DC-BCS-SPL	35.2	40.7	44.6	47.7	50.0
Modified-CS-Residual	28.0	30.7	32.5	34.9	37.1
k-t FOCUSS	25.8	28.1	31.8	34.8	37.3
Intraview MS-BCS-SPL	33.2	37.2	40.0	41.5	42.9
Intraview TV	33.0	36.3	39.1	41.6	44.3
Monopoly					
Multistage DC-TV	35.2	41.7	46.3	49.4	51.8
Multistage DC-BCS-SPL	30.5	35.5	39.5	43.0	46.1
Modified-CS-Residual	25.9	27.9	29.6	31.7	33.8
k-t FOCUSS	24.4	26.3	29.4	31.6	34.5
Intraview MS-BCS-SPL	29.3	33.1	35.5	38.0	40.2
Intraview TV	29.8	35.0	39.5	43.7	47.6
Plastic					
Multistage DC-TV	46.6	51.5	53.7	55.2	56.7
Multistage DC-BCS-SPL	38.1	44.5	48.9	52.0	54.1
Modified-CS-Residual	29.8	32.9	34.9	37.6	40.0
k-t FOCUSS	28.2	30.3	34.1	36.9	39.8
Intraview MS-BCS-SPL	36.5	41.5	45.2	49.1	52.4
Intraview TV	43.1	48.8	52.7	55.6	57.9

The bold values correspond to the maximal PSNR (dB) reached for a given substrate and image

Table 3 Reconstruction time in seconds per view (spv)

Algorithm	Time (spv)
Intraview TV	3
Intraview MS-BCS-SPL	43
k-t FOCUSS	79
Multistage DC-BCS-SPL	190
Multistage DC-TV	207
Modified-CS-residual	1105

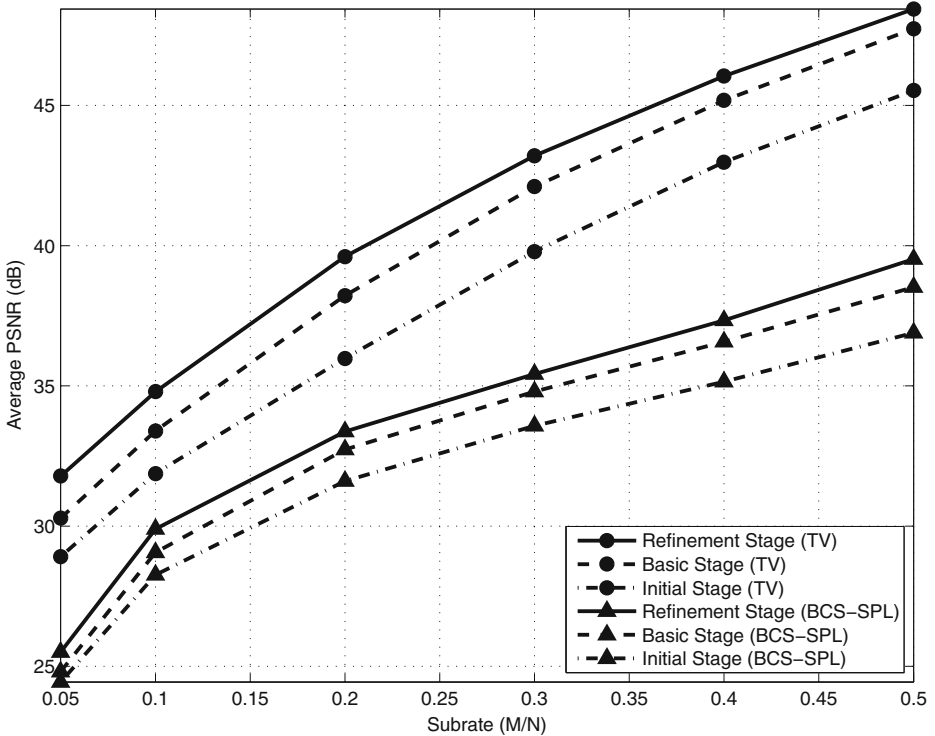


Fig. 7 Reconstruction performance for the “Ballet” multiview video

initial recovery. With such a starting point, increased performance is observed at each stage.

We now present a comprehensive comparison between several CS reconstruction algorithms for multiview images. We compare the multistage DC-CS reconstruction proposed in Section 4 to two prominent CS reconstruction algorithms, Modified-CS-Residual [37] and k-t FOCUSS [18, 19], both of which we have described previously in Section 5. As used with multiview images, k-t FOCUSS uses iterative recovery with DE/DC from non-key frames from the neighboring key frames. On the other hand, Modified-CS-Residual does not employ DE/DC but rather attempts to explicitly track the sparsity pattern frame to frame. We use the implementations of k-t FOCUSS⁶ and Modified-CS-Residual⁷ available from their respective authors. Although both k-t FOCUSS and Modified-CS-Residual were originally designed for the reconstruction of dynamic MRI data, they both constitute benchmark algorithms for the reconstruction of multiview imagery as well as. Both techniques, being oriented toward dynamic MRI, feature frame-by-frame CS measurement driven by a 2D full-frame Fourier transform applied identically to each frame with low-frequency coefficients benefiting from a higher subrate. The subrate for all frames

⁶http://bisp.kaist.ac.kr/research_02.htm

⁷<http://home.engineering.iastate.edu/~luwei/modcs/>

(key and non-key) for k-t FOCUSS is identical, contrary to its typical use with video in which key frames have increased subrate.

Finally, we compare to independent view-by-view, or “intraview”, reconstruction for each of the multiview datasets. We consider the multiscale (MS) variant of BCS-SPL originally proposed in [12]; in the results here, we refer to it as “Intraview MS-BCS-SPL.” We also consider the TV reconstruction of each view, referred to as “Intraview TV” in the results. The MS-BCS-SPL utilizes block-based CS measurement in the wavelet domain with blocks of size 16×16 , while intraview TV uses a full-frame block-Hadamard SRM [16].

Figures 5 and 6 illustrate the performance of the various reconstructions for different subrates; more complete results are found in Table 2. As it can be seen in the above mentioned figures and tables, DC-TV almost always outperforms the other techniques considered, sometimes by as much as 10–12 dB. The combination of a strong CS image-recovery technique, such as TV, and our iterative procedure of DE/DC prediction and residual recovery offers superior reconstruction accuracy.

Although none of the implementations have been particularly optimized for execution speed, we present reconstruction times for the algorithms in Table 3. Here, we measure the average length of time required to recover one frame out of the multiview dataset.

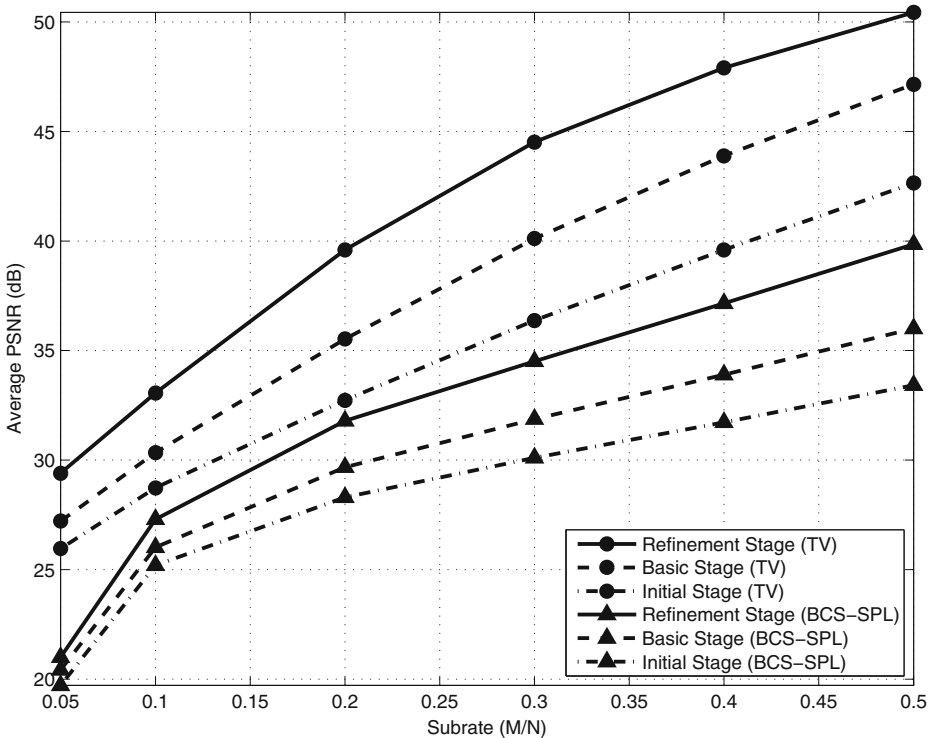


Fig. 8 Reconstruction performance for the “Book Arrival” multiview video

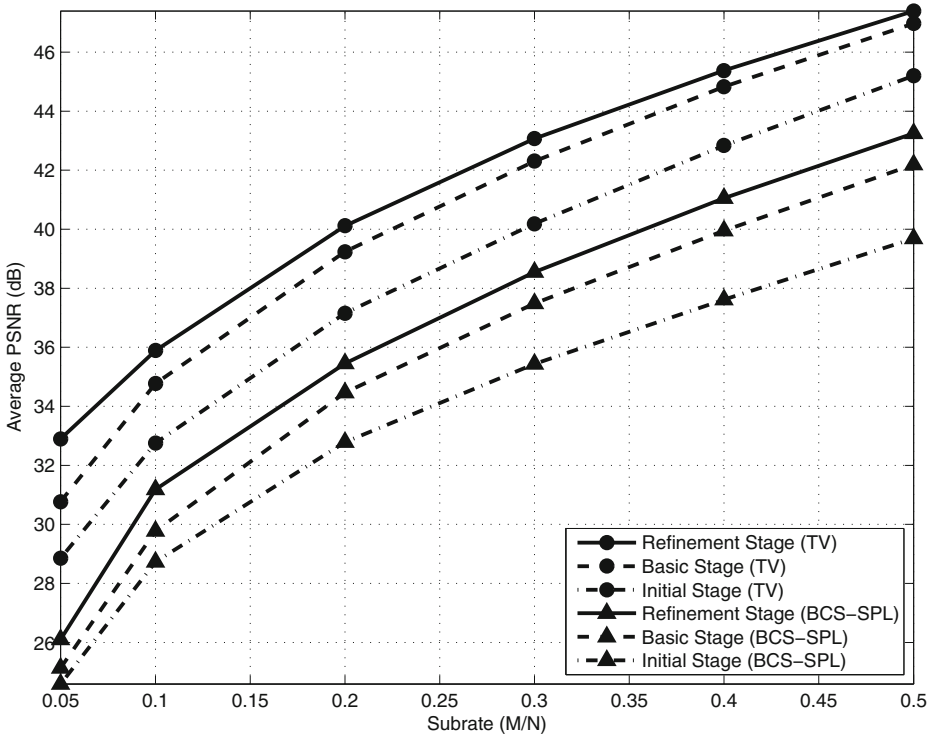


Fig. 9 Reconstruction performance for the “Break Dancer” multiview video

6.2 Multiview video

For multiview video, we consider the case in which each frame within each view of a multiview video sequence has same subrate. The DE/DC prediction across views within the multiview video sequence is identical to that used for multiview image recovery, i.e., the optical-flow implementation proposed in [23]. However, to handle large motion discrepancies between frames of the multiview sequence, block-based ME/MC using full-search ME with a block size of 16×16 and a search window of 32×32 is employed temporally.

Figures 7, 8 and 9 present the performance at each of the three stages of reconstruction over various subrates. Three 256×192 grayscale multiview video sequences are used, namely, “Book Arrival,”⁸ “Ballet,” and “Break Dancer.”⁹ The simulations are done, for each of these sequences, on the first five views and first five frames within each view, thus a total of 25 frames per multiview video sequence are used. As before, DE/DC coupled with ME/MC in the DC-CS reconstruction occurring in the basic stage improves reconstruction quality dramatically over the

⁸Provided courtesy of Fraunhofer HHI.

⁹The “Ballet” and “Break Dancer” multiview video sequences are available, courtesy of Microsoft Research, from <http://research.microsoft.com/en-us/um/people/sbkang/3dvideodownload/>.

Table 4 PSNR performance (in dB) at each CS reconstruction stage of the proposed method for several multiview video datasets

Algorithm	Subrate					
	0.05	0.1	0.2	0.3	0.4	0.5
Book arrival						
DC-TV						
4th refinement	29.406	33.066	39.597	44.517	47.903	50.438
3rd refinement	29.082	32.771	39.107	44.083	47.607	50.269
2nd refinement	28.650	32.293	38.405	43.399	47.083	49.932
1st refinement	28.070	31.546	37.345	42.255	46.057	49.146
Basic	27.215	30.341	35.526	40.117	43.892	47.154
Initial	25.956	28.739	32.735	36.379	39.594	42.642
DC-BCS-SPL						
4th refinement	21.004	27.301	31.791	34.509	37.158	39.855
3rd refinement	20.919	27.091	31.462	34.108	36.673	39.291
2nd refinement	20.813	26.828	31.037	33.583	36.030	38.536
1st refinement	20.663	26.488	30.473	32.880	35.158	37.502
Basic	20.422	26.017	29.676	31.868	33.899	36.005
Initial	19.721	25.196	28.301	30.106	31.723	33.416
Breakdancer						
DC-TV						
4th refinement	32.902	35.891	40.113	43.070	45.381	47.395
3rd refinement	32.858	35.989	40.128	43.057	45.350	47.395
2nd refinement	32.612	35.982	40.083	42.971	45.315	47.358
1st refinement	32.018	35.732	39.940	42.881	45.242	47.309
Basic	30.773	34.772	39.239	42.305	44.830	46.969
Initial	28.852	32.755	37.157	40.192	42.831	45.200
DC-BCS-SPL						
4th refinement	26.098	31.184	35.449	38.549	41.055	43.249
3rd refinement	25.945	31.046	35.517	38.600	41.076	43.254
2nd refinement	25.747	30.787	35.457	38.538	41.013	43.185
1st refinement	25.484	30.386	35.165	38.278	40.766	42.980
Basic	25.133	29.773	34.464	37.486	39.950	42.190
Initial	24.593	28.731	32.786	35.435	37.614	39.685
Ballet						
DC-TV						
4th refinement	31.784	34.790	39.613	43.206	46.050	48.441
3rd refinement	31.696	34.744	39.532	43.173	46.036	48.431
2nd refinement	31.478	34.574	39.400	43.048	45.939	48.396
1st refinement	31.096	34.230	39.104	42.841	45.753	48.244
Basic	30.285	33.394	38.221	42.107	45.178	47.739
Initial	28.907	31.880	35.984	39.797	42.980	45.536
DC-BCS-SPL						
4th refinement	25.507	29.899	33.372	35.425	37.335	39.523
3rd refinement	25.391	29.817	33.331	35.340	37.285	39.436
2nd refinement	25.247	29.668	33.306	35.324	37.266	39.366
1st refinement	25.062	29.441	33.145	35.205	37.091	39.145
Basic	24.811	29.060	32.731	34.795	36.576	38.525
Initial	24.442	28.260	31.607	33.574	35.151	36.893

The bold values correspond to the maximal PSNR (dB) reached for a given subrate and image

independent reconstruction in the initial stage, and the refinement stage produces even further quality improvement. As in the case of multiview image sequences, DC-TV provides superior PSNR as compared to DC-BCS-SPL for multiview video recovery. Full results for video recovery are given in Table 4.

7 Conclusions

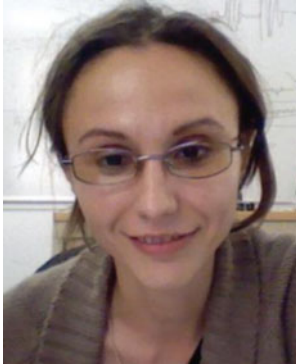
In this paper, we investigated the CS recovery of multiview images as well as multiview video sequences. Central to this reconstruction process was the creation of predictions of the current view from adjacent views via DE and DC, and, in the case of multiview video, predictions from temporally neighboring frames via ME and MC. These DE/DC- and ME/MC-based predictions were used in a CS reconstruction of a residual rather than the frame directly. Experimental results displayed a significant increase in performance when using signal predictions in comparison to recoveries which merely reconstruct each image independently from one another. Furthermore, a significant performance advantage was seen for the proposed techniques in comparison to several benchmark multiple-frame CS reconstruction techniques, thus demonstrating the effectiveness of the improvements we propose in this work: TV view, and residual, reconstruction combined with DE/DC based on optical flow.

References

1. Bioucas-Dias JM, Figueiredo MAT (2007) A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans Image Process* 16(12):2992–3004
2. Blumensath T, Davies ME (2009) Iterative hard thresholding for compressed sensing. *Appl Comput Harmon Anal* 27(3):265–274
3. Candès E, Tao T (2006) Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans Inf Theory* 52(12):5406–5425
4. Candès EJ (2006) Compressive sampling. In: *Proceedings of the International Congress of Mathematicians*, vol 3, pp 1433–1452. Madrid, Spain
5. Candès EJ, Wakin MB (2008) An introduction to compressive sampling. *IEEE Signal Process Mag* 25(2):21–30
6. Chambolle A, Lions PL (1997) Image recovery via total variation minimization and related problems. *Numer Math* 76(2):168–188
7. Chan TF, Esedoglu S, Park F, Yip A (2006) Total variation image reconstruction: overview and recent developments. In: Paragios N, Chen Y, Faugeras OD (eds) *Handbook of mathematical models in computer vision*, chap 2. Springer, New York
8. Chen SS, Donoho DL, Saunders MA (1998) Atomic decomposition by basis pursuit. *SIAM J Sci Comput* 20(1):33–61
9. Chen X, Frossard P (2009) Joint reconstruction of compressed multi-view images. In: *Proceedings of the international conference on acoustics, speech, and signal processing*, pp 1005–1008. Taipei, Taiwan
10. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, Baraniuk RG (2008) Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag* 25(2):83–91
11. Figueiredo MAT, Nowak RD, Wright SJ (2007) Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J Sel Areas Commun* 1(4):586–597
12. Fowler JE, Mun S, Tramel EW (2011) Multiscale block compressed sensing with smoother projected Landweber reconstruction. In: *Proceedings of the European signal processing conference*, pp 564–568. Barcelona, Spain

13. Fowler JE, Mun S, Tramel EW (2012) Block-based compressed sensing of images and video. *Foundations and Trends in Signal Processing* 4(4):297–416
14. Gamper U, Boesiger P, Kozerke S (2008) Compressed sensing in dynamic MRI. *Magn Reson Med* 59(2):365–373
15. Gan L (2007) Block compressed sensing of natural images. In: *Proceedings of the international conference on digital signal processing*, pp 403–406. Cardiff, UK
16. Gan L, Do TT, Tran TD (2008) Fast compressive imaging using scrambled block Hadamard ensemble. In: *Proceedings of the European signal processing conference*. Lausanne, Switzerland
17. Guillemot C, Pereira F, Torres L, Ebrahimi T, Leonardi R, Ostermann J (2007) Distributed monoview and multiview video coding. *IEEE Signal Process Mag* 24(5):67–76
18. Jung H, Sung K, Nayak KS, Kim EY, Ye JC (2009) k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magn Reson Med* 61(1):103–116
19. Jung H, Ye JC (2010) Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: what we can learn from video compression techniques. *Int J Imaging Syst Technol* 20(2):81–98
20. Kingsbury NG (2001) Complex wavelets for shift invariant analysis and filtering of signals. *Appl Comput Harmon Anal* 10:234–253
21. Li C (2009) An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing. Master's thesis, Rice University
22. Li X, Wei Z, Xiao L (2010) Compressed sensing joint reconstruction for multi-view images. *Electron Lett* 46(23):1548–1550
23. Liu C (2009) Beyond pixels: exploring new representations and applications for motion analysis. Ph.D. thesis, Massachusetts Institute of Technology
24. Lu W, Vaswani N (2009) Modified compressive sensing for real-time dynamic MR imaging. In: *Proceedings of the international conference on image processing*, pp 3045–3048. Cairo, Egypt
25. Mun S, Fowler JE (2009) Block compressed sensing of images using directional transforms. In: *Proceedings of the international conference on image processing*, pp 3021–3024. Cairo, Egypt
26. Park JY, Wakin MB (2012) A geometric approach to multi-view compressive imaging. *EURASIP J Appl Signal Process* 2012:37. doi:[10.1186/1687-6180-2012-37](https://doi.org/10.1186/1687-6180-2012-37)
27. Puri R, Majumdar A, Ishwar P, Ramchandran K (2006) Distributed video coding in wireless sensor networks. *IEEE Signal Process Mag* 23(4):94–106
28. Qiu C, Lu W, Vaswani N (2009) Real-time dynamic MR image reconstruction using Kalman filtered compressed sensing. In: *Proceedings of the international conference on acoustics, speech, and signal processing*, pp 393–396. Taipei, Taiwan
29. Rauhut H (2010) Compressive sensing and structured random matrices. In: Fornasier M (ed) *Theoretical foundations and numerical methods for sparse recovery*. Walter de Gruyter, Inc., Berlin
30. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D* 60(1–4):259–268
31. Trocan M, Maugey T, Fowler JE, Pesquet-Popescu B (2010) Disparity-compensated compressed-sensing reconstruction for multiview images. In: *Proceedings of the IEEE international conference on multimedia and expo*, pp 1225–1229. Singapore
32. Trocan M, Maugey T, Tramel EW, Fowler JE, Pesquet-Popescu B (2010) Compressed sensing of multiview images using disparity compensation. In: *Proceedings of the international conference on image processing*, pp 3345–3348. Hong Kong
33. Trocan M, Maugey T, Tramel EW, Fowler JE, Pesquet-Popescu B (2010) Multistage compressed-sensing reconstruction of multiview images. In: *Proceedings of the IEEE workshop on multimedia signal processing*, pp 111–115. Saint-Malo, France
34. Tropp J, Gilbert A (2007) Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans Inf Theory* 53(12):4655–4666
35. Vaswani N (2008) Kalman filtered compressed sensing. In: *Proceedings of the international conference on image processing*, pp 893–896. San Diego, CA
36. Vaswani N (2010) LS-CS-Residual (LS-CS): compressive sensing on least squares residual. *IEEE Trans Signal Process* 57(8):4108–4120
37. Vaswani N, Lu W (2010) Modified-CS: modifying compressive sensing for problems with partially known support. *IEEE Trans Signal Process* 58(9):4595–4607
38. Wakin MB (2009) A manifold lifting algorithm for multi-view compressive imaging. In: *Proceedings of the picture coding symposium*. Chicago, IL

39. Wakin MB, Laska JN, Duarte MF, Baron D, Sarvotham S, Takhar D, Kelly KF, Baraniuk RG (2006) An architecture for compressive imaging. In: Proceedings of the international conference on image processing, pp 1273–1276. Atlanta, GA
40. Wakin MB, Laska JN, Duarte MF, Baron D, Sarvotham S, Takhar D, Kelly KF, Baraniuk RG (2006) Compressive imaging for video representation and coding. In: Proceedings of the picture coding symposium. Beijing, China
41. Wang Y, Yang J, Yin W, Zhang Y (2008) A new alternating minimization algorithm for total variation image reconstruction. *SIAM J Imaging Sci* 1(3):248–272



Maria Trocan received her B.Eng. in Electrical Engineering and Computer Science from Politehnica University of Bucharest in 2004 and her Ph.D. from Telecom ParisTech (formerly Ecole Nationale Supérieure de Telecommunications) in 2007. She joined Joost - Netherlands in 2007, where she worked as research engineer involved in the design and development of video transcoding systems. Since May 2009 she is Associate Professor with the Signal, Image and Telecommunications Department at Institut Supérieur d'Electronique de Paris (ISEP). Her current research interests focus on image and video analysis and compression, sparse representations and wavelet-based processing techniques.



Eric W. Tramel received the B.S. and Ph.D. degrees in computer engineering in 2007 and 2012, respectively, both from Mississippi State University. In 2011, he was a Research Intern at Canon USA, Inc. from May to August. He also served as a Research Associate for the Geosystems Research Institute (GRI) at Mississippi State from 2009 to 2012. His research interests include compressed sensing, image and video coding, image and video multiview systems, data-compression, and pattern recognition.



James E. Fowler (S'91–M'96–SM'02) received the B.S. degree in computer and information science engineering and the M.S. and Ph.D. degrees in electrical engineering in 1990, 1992, and 1996, respectively, all from the Ohio State University. In 1995, Dr. Fowler was an intern researcher at AT&T Labs in Holmdel, NJ, and, in 1997, he held an NSF-sponsored postdoctoral assignment at the Université de Nice-Sophia Antipolis, France. In 2004, he was a visiting professor in the Département TSI at Ecole Nationale Supérieure des Telecommunications (ENST), Paris, France. He is currently Billie J. Ball Professor and Graduate Program Director of the Department of Electrical & Computer Engineering at Mississippi State University in Starkville, MS; he is also a researcher in the Geosystems Research Institute (GRI) at Mississippi State.



Beatrice Pesquet received the engineering degree in telecommunications from the Politechnica Institute in Bucharest in 1995 and the Ph.D. thesis from the Ecole Normale Supérieure de Cachan in 1998. In 1998 she was a Research and Teaching Assistant at Université Paris XI and in 1999 she joined Philips Research France, where she worked during two years as a research scientist, then project leader, in scalable video coding. Since Oct. 2000 she is with Télécom ParisTech (formerly, ENST), first as an Associate Professor, and since 2007 as a Professor, Head of the Multimedia Group. She is the Head of the UBIMEDIA common research laboratory between Alcatel-Lucent and Institut Télécom. Her current research interests are in source coding, scalable, robust and distributed video compression and sparse representations.