Video Compressed Sensing with Multihypothesis

Eric W. Tramel and James E. Fowler

Department of Electrical and Computer Engineering
Geosystems Research Institute (GRI)
Mississippi State University, Starkville, Mississippi, USA

Abstract

The compressed-sensing recovery of video sequences driven by multihypothesis predictions is considered. Specifically, multihypothesis predictions of the current frame are used to generate a residual in the domain of the compressed-sensing random projections. This residual being typically more compressible than the original frame leads to improved reconstruction quality. To appropriately weight the hypothesis predictions, a Tikhonov regularization to an ill-posed least-squares optimization is proposed. This method is shown to outperform both recovery of the frame independently of the others as well as recovery based on single-hypothesis prediction.

Introduction

Compressed sensing (CS) (e.g., [1, 2]) is a new signal-sampling and recovery model that has emerged in recent years, and there has been much work invoking CS in disparate applications such as natural-image acquisition, remote sensing, cognitive radio, and medical imaging, just to name a few. In these fields, the primary goal has largely been the blind recovery of CS-acquired signals. Much work has been done in this area with many different varieties of solvers being actively researched over the past several years in an effort to decrease recovery computation time without sacrificing distortion performance. For the most part, these reconstruction strategies are oblivious to the structure of the signal being recovered beyond a general assumption of sparsity, or compressibility, in some transform basis. Recently, however, several reconstruction approaches have focused on contexts in which side information about the signal content is available to aid signal recovery. Some proposed methods, such as Bayesian CS [3] as well as model CS [4, 5], exploit certain *a priori* knowledge of signal structure, or the probability thereof, to guide recovery. These methods, however, do not address the situation in which one or more predictions of the signal to be recovered are available to the CS reconstruction process. Video sequences are one form of data in particular in which predictions are commonly used in various forms of processing. Specifically, it is typical in video processing that one or more reference frames are used to make predictions of some current frame such that the resulting residual frame has dramatically lowered signal energy leading to more efficient representation and processing.

In this paper, we consider the CS recovery of video sequences in which frame-to-frame predictions are used to aid the CS recovery process. In effect, we perform CS

This material is based upon work supported by the National Science Foundation under Grant No. CCF-0915307.

In *Proceedings of the Data Compression Conference*, J. A. Storer and M.W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

recovery on the prediction residual which is, in most cases, significantly more compressible than the original frame, resulting in a higher-quality CS recovery. Key to our approach is the use of motion estimation (ME) and motion compensation (MC) such that the frame-to-frame predictions compensate for object motions between frames. Such use of ME/MC derives from traditional video-compression algorithms which make extensive use of sophisticated MC strategies.

One form of MC widely employed in traditional video compression is that of multihypothesis (MH) prediction in which multiple, distinct predictions are created and then combined to yield a composite prediction superior to any of the constituent single-hypothesis (SH) predictions [6]. As the primary contribution of this paper, we show how such MH prediction can be incorporated into the CS recovery of video so as to increase reconstruction quality over equivalent SH-driven recovery. Central to this discussion is a formulation of the MH prediction process in the domain of the random CS projections; as this formulation results in an ill-posed optimization, we resort to Tikhonov regularization [7] which is widely used to yield tractable solutions to such ill-posed problems. In experimental results, we compare our proposed Tikhonov-based multihypothesis regularization against both an equivalent SH-based reconstruction as well as the straightforward reconstruction in which each video frame is recovered independently from the others. We find that our proposed approach yields superior reconstruction across a broad range of subsampling rates.

Background

In essence, CS combines signal acquisition and dimensionality reduction by measuring a projection of the signal data, x, of dimensionality N using some basis, Φ, of dimensionality $M \times N$ where $M \ll N$; i.e.,

$$y = \Phi x, \tag{1}$$

where $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^M$. If x is sufficiently sparse in some transform basis Ψ, then x is recoverable from y by the optimization,

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|\Psi x\|_1, \quad \text{such that } y = \Phi x, \tag{2}$$

as long as Ψ and Φ are sufficiently incoherent, and M is sufficiently large. We define the subsampling rate, or *subrate*, imposed by (1) to be $S = M/N$. Usually, Φ is a random matrix such that it is incoherent with any chosen Ψ.

In practical applications, most natural signals are not truly sparse in any transform basis Ψ. Accordingly, a common variant of the recovery problem of (2) is to relax the equality for a bound; i.e.,

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|\Psi x\|_1, \quad \text{such that } \|y - \Phi x\|_2 \leq \epsilon. \tag{3}$$

Many different approaches to the relaxed recovery problem of (3) have been proposed over the years; central to all of these recovery methods, however, is a tradeoff between computational complexity and the quality of the recovery. Yet, regardless of how this

In *Proceedings of the Data Compression Conference*, J. A. Storer and M.W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

tradeoff is chosen, all of these methods are faced with the fact that CS recovery for large signals, such as a natural image or video of realistic dimensions, is going to be a time-consuming process.

Gan [8] suggests that, in the case of natural images, the computational complexity of CS reconstruction can be assuaged by breaking up the image into distinct blocks during acquisition. Block-based CS (BCS) removes the global sampling of x by a dense Φ and replaces it with a block-diagonal measurement matrix by which local sampling of x within distinct blocks of size $B \times B$ is accomplished. When the same Φ_B is used for every block, Φ takes on a block-diagonal form,

$$\Phi = \begin{bmatrix} \Phi_B & 0 & \cdots & 0 \\ 0 & \Phi_B & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \Phi_B \end{bmatrix}, \tag{4}$$

such that (1) can be effectuated in a block-by-block fashion; i.e.,

$$y_i = \Phi_B x_i, \tag{5}$$

where x_i is block i of the image. The size of Φ_B is $M_B \times B^2$ such that the subrate of BCS is $S = M_B/B^2$.

For recovery, [8] suggests a procedure that couples projected Landweber iterations with smoothing in the form of Wiener filtering. This smoothed projected Landweber (SPL) procedure thus combines a fast, iterative solution to (3) with the imposition of a smoothness constraint designed to eliminate blocking artifacts. In [9], the overall process of BCS sampling and SPL reconstruction was called BCS-SPL.

BCS-SPL was extended in [9] by use of bivariate shrinkage for thresholding and directional transforms such as a dual-tree discrete wavelet transform (DDWT) and a contourlet transform (CT). These modifications provide significant recovery-quality improvement while maintaining a reasonable reconstruction time. The results in [9] suggest that BCS-SPL augmented with such directional transforms is competitive with the state of the art for CS recovery of a single still image. As a consequence, we use BCS-SPL as the image-reconstruction procedure at the foundation of the video reconstruction we consider next.

Method

The straightforward application of CS to video would involve the vectorization of a 3D group of frames into a single 1D vector with sampling applied as in (1). However, the computation and memory issues associated with this approach are prohibitive; additionally, a global sampling simultaneously across the spatial and temporal extent of a group of frames is likely to be impractical to implement in a real sampling device [10]. Consequently, we focus on the situation in which video frames are sampled independently in a 2D fashion, e.g., by applying a suitable image-acquisition sampler in a frame-by-frame fashion as in [11]. To cut computation and memory, we focus on BCS image sampling as in (5) applied frame by frame.

In *Proceedings of the Data Compression Conference*, J. A. Storer and M.W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

Given a frame-by-frame acquisition, the most straightforward reconstruction would be to reconstruct the individual frames independently using the BCS-SPL procedure discussed above. However, such an independent BCS-SPL reconstruction ignores the fact that consecutive video frames are usually highly correlated. In traditional video compression, such frame-to-frame temporal correlation is exploited by using ME/MC to form a motion-compensated prediction of the current frame and then encoding the residual between the current frame and its ME/MC prediction. In [12], an approach to incorporating ME/MC prediction into the CS recovery of video was proposed such that CS reconstruction was applied to a projection-domain residual between the current frame and its ME/MC prediction. We overview this residual-reconstruction approach next.

Residual Reconstruction

Residual reconstruction seeks a sparser representation of a given signal by recovering the difference between the signal and some prediction. The philosophy is very similar to that of DPCM in traditional source coding—if a prediction is similar to the signal it is intended to approximate, then the value of the residual over most of the support should be insignificant in magnitude. In traditional video coding, this technique is used extensively to create highly compressible residual frames which are then compressed with a still-image coder.

Residual reconstruction can be easily integrated into the CS paradigm because it requires no change on the part of the signal acquisition and has a simple implementation on the reconstruction side. Suppose that we have a given signal, x, with a measurement basis, Φ, such that measurements, y are calculated via (1). If we are given some kind of prediction of x in the ambient domain of x—namely, $\tilde{x}$, which we hope satisfies $\tilde{x} \approx x$—then we can find the residual r between the two signals as $r = x - \tilde{x}$. Because y is acquired simply by taking the inner products of x with the rows of Φ, the projection of r into the measurement basis is

$$q = \Phi r = \Phi\left(x - \tilde{x}\right) = y - \Phi\tilde{x}. \tag{6}$$

Because of the linear nature of the signal-sampling process, a simple subtraction of a projection of $\tilde{x}$ provides us with a projected residual signal at the reconstruction side without changing our signal-acquisition procedure. This residual should be more amenable to CS recovery because it is expected to be much more compressible than x itself. The final reconstruction of y, $\hat{x}$, is calculated as

$$\hat{x} = \tilde{x} + \text{Reconstruct}(y - \Phi\tilde{x}, \Phi), \tag{7}$$

where Reconstruct$(\cdot)$ is some suitable CS recovery. The quality of $\hat{x}$ is directly tied to the ability of the reconstruction to recover r from q; i.e.,

$$\|x - \hat{x}\|_2 = \|x - (\tilde{x} + r + e_r)\|_2 = \|(x - \tilde{x}) - r - e_r\|_2 = \|e_r\|_2, \tag{8}$$

where e_r is the error resulting from a non-exact recovery of r.

In *Proceedings of the Data Compression Conference*, J. A. Storer and M.W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

The key to the successful use of residual reconstruction is thus to create a prediction $\tilde{x}$ that is as close as possible to x such that the resulting residual, r, is highly compressible. Thus, the goal is to carry out the optimization,

$$\tilde{x} = \arg\min_{\tilde{x}} \|x - \tilde{x}\|_2 . \tag{9}$$

However, the creation of the prediction $\tilde{x}$ occurs during CS reconstruction; as a consequence, x is unknown, and (9) cannot be implemented as written. There are two strategies to approximate (9) using only information known to the CS reconstruction. The first would be to approximate x with an initial CS recovery from y and use the resulting approximation to x to drive the prediction process; i.e.,

$$\tilde{x} = \arg\min_{\tilde{x}} \left\| \text{Reconstruct}(y, \Phi) - \tilde{x} \right\|_2, \tag{10}$$

where Reconstruct$(\cdot)$ is some suitable CS recovery. The resulting $\tilde{x}$ is then used in (7) to form the final reconstruction $\hat{x}$ using a CS reconstruction from the measurement-domain residual, $q = y - \Phi\tilde{x}$. This approach was used in the CS video reconstruction presented in [12].

An alternative strategy is to recast the optimization of (10) from the ambient signal domain of x into the measurement domain of y; specifically,

$$\tilde{x} = \arg\min_{\tilde{x}} \|\Phi x - \Phi\tilde{x}\|_2 = \arg\min_{\tilde{x}} \|y - \Phi\tilde{x}\|_2 . \tag{11}$$

Although (11) recasts the search for the prediction into the measurement domain, the Johnson-Lindenstrauss (JL) lemma [13–15] suggests that the solution of (11) will likely coincide with that of (9). In brief, the JL lemma holds that L points in $\mathbb{R}^N$ can be projected into a K-dimensional subspace while approximately maintaining pairwise distances as long as $K \geq O(\log L)$. As a consequence, the $\tilde{x}$ closest to x in (9) should map to the $\Phi\tilde{x}$ that is closest to y in (11), provided that the number of candidates searched in the minimizations is not too large.

Our experimental observations reveal that the measurement-domain prediction of (11) provides better predictions in general than the ambient-domain strategy represented by (10). This is due to the fact that (10) uses only a noisy approximation to x, whereas the JL lemma suggests that (11) should nearly duplicate the targeted procedure of (9). As a consequence, we focus on measurement-domain predictions in the form of (11) in the remainder of our development.

SH Frame Prediction for CS Reconstruction

In traditional video coding, frame predictions are calculated from temporally neighboring frames which are likely to have similar content to the target frame using ME/MC. Specifically, the frame at time t to be predicted, x_t, is split into blocks of size $B \times B$. The chosen reference frame or frames are then searched within a spatial region surrounding the location of the target block within x_t. The best-matching block, chosen according to some distortion measurement, in the reference frames then

forms the prediction of the target block. This is known as SH prediction in the video-coding community since a single, best-matching hypothesis prediction (a block in one of the reference frames, in this case) is chosen to represent the target block.

In the CS reconstruction of video wherein each frame has been sampled using BCS applied frame by frame, the ensemble of measurements for frame x_t is $y_{t,i} = \Phi x_{t,i}$, where i is a block index. In order to create a prediction of a given block, $x_{t,i}$, we recast (11) as

$$\tilde{x}_{t,i} = \arg \min_{x \in \mathcal{H}_{t,i}} \|y_{t,i} - \Phi x\|_2 , \tag{12}$$

where $\mathcal{H}_{t,i}$ is the set of blocks culled from the reference frame or frames within the search space given for block $x_{t,i}$ (typically a rectangular region about the spatial location of $x_{t,i}$ in the frame).

MH Frame Prediction for CS Reconstruction

Video coding has long exploited MH methods to improve video-coding quality [6]; common forms include subpixel-accurate MC [16], overlapped block MC [17], bidirectional MC (B-frames), and long-term-memory MC [18]. These techniques can be viewed as tradeoffs specific to a rate-limited environment; that is, these techniques impose specific structures on the hypotheses that form the ultimate prediction in order to limit the amount of additional motion-vector rate overhead entailed by multiple predictions of a single block. However, in the context of CS reconstruction, the MH predictions are all calculated at the reconstruction side of the system, there is no associated rate burden, and we are able to consider more intensive forms of MH prediction, essentially combining the best hypotheses available from the reference frames without the imposition of rate-limiting structure.

For a MH CS reconstruction, the goal is to reformulate (9) so that, instead of choosing a single hypothesis, we find an optimal linear combination of all hypotheses contained in the search set; i.e, (9) becomes

$$w_{t,i} = \arg \min_{w} \|x_{t,i} - H_{t,i} w\|_2 , \tag{13}$$

$$\tilde{x}_{t,i} = H_{t,i} w_{t,i}, \tag{14}$$

where we have also recast (9) for block-based prediction with i being the block index. Here, $H_{t,i}$ is a matrix of dimensionality $B^2 \times K$ whose columns are the rasterizations of the possible blocks within the search space of the reference frames, and $K = |\mathcal{H}_{t,i}|$. In this context, $w_{t,i}$ is a column vector which represents the optimal linear combination of the columns of $H_{t,i}$; the solution of this optimization can be calculated as a simple least-squares (LSQ) problem.

Of course, in the case of CS reconstruction, (14), like (9), cannot be implemented—we cannot calculate $w_{t,i}$ directly because we do not have access to $x_{t,i}$; we have only its measurement, $y_{t,i}$. We thus adopt the measurement-domain approach of (12), modifying it to the MH case. However, this makes the optimization a much more difficult, ill-posed problem, because we have to calculate the optimal linear combination within the projected space of Φ; i.e., combining (12) and (13) yields

$$\hat{w}_{t,i} = \arg \min_{w} \|y_{t,i} - \Phi H_{t,i} w\|_2 . \tag{15}$$

In general, $w_{t,i} \neq \hat{w}_{t,i}$ unless Φ is square, which is not the case for CS. The ill-posed nature of this problem necessitates some form of regularization of the LSQ optimization.

The most common approach to regularizing an LSQ problem is Tikhonov regularization [7] which imposes an ℓ_2 penalty on the norm of $\hat{w}_{t,i}$,

$$\hat{w}_{t,i} = \arg\min_{w} \|y_{t,i} - \Phi H_{t,i} w\|_2 + \lambda \|\Gamma w\|_2 , \tag{16}$$

where Γ is known as the Tikhonov matrix. The Γ term allows the imposition of prior knowledge on the solution; in some contexts, it may make sense to use a high-pass or difference operator for Γ to obtain a smooth result, or, in others, to set $\Gamma = I$ to impose an energy constraint on the solution. In our case, we take the approach that hypotheses which are the most dissimilar from the target block should be given less weight than hypotheses which are most similar. Specifically, we propose a diagonal Γ in the form of

$$\Gamma = \begin{bmatrix} \|y_{t,i} - \Phi h_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|y_{t,i} - \Phi h_K\|_2 \end{bmatrix}, \tag{17}$$

where $h_1, h_2, \ldots, h_K$ are the columns of $H_{t,i}$. With this structure, Γ penalizes weights of large magnitude assigned to hypotheses which have a significant distance from $y_{t,i}$ when projected into the measurement domain. For each block, then, $\hat{w}_{t,i}$ can be calculated directly by the usual Tikhonov solution,

$$\hat{w}_{t,i} = \left((\Phi H_{t,i})^T (\Phi H_{t,i}) + \lambda^2 \Gamma^T \Gamma \right)^{-1} (\Phi H_{t,i})^T y_{t,i}. \tag{18}$$

In this formulation, λ is a scale factor that controls the relative effect of the Tikhonov-regularization term in the optimization of (16). The choice of λ can have a large effect on the performance of the regularization, so it is important to find a value which imposes an adequate level of regularization without causing $\|y_{t,i} - \Phi H_{t,i} w\|_2$ to become too large. We found in practice that, over a large set of different frames, a value of $\lambda \in [0.1, 0.3]$ provided the best results; consequently, we use $\lambda = 0.25$ from this point on.

Experimental Results

We consider the first two consecutive frames, x_1 and x_2, of a given video sequence—the first frame, x_1, is used as a reference frame, while the second frame, x_2, is the "test frame" used to measure reconstruction performance. In all cases, the reference frame is BCS sampled with a relatively high subrate of $S_1 = 0.5$ and reconstructed using BCS-SPL. On the other hand, the test frame is BCS sampled using a range of subrates, $S_2 \leq S_1$. This disparity in subrates is intended to reflect the situation in which the video sequence is sampled with relatively high subrates for certain "key frames" which anchor the ME/MC-driven reconstruction process for one or more intervening "non-key frames" (e.g., [12]). Throughout, we use a block size of $B = 16$

In *Proceedings of the Data Compression Conference*, J. A. Storer and M.W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 193–202.

for BCS, and a discrete wavelet transform (DWT) with 4 levels of decomposition as the sparsity basis for BCS-SPL reconstruction.

The reconstructed reference frame is used to create a prediction of each block of the test frame; afterward, residual reconstruction (i.e., (6)–(7)) of the test frame is conducted. For comparison, we consider the performance of the SH prediction of (12), as well as the straightforward BCS-SPL reconstruction of the test frame independently of the reference frame, which we refer to as "independent reconstruction." In all cases, a spatial window size of ± 15 pixels about the current block is used as the search space for finding the hypotheses for both the SH and MH predictions.

The PSNR performance of the test-frame recovery as the subrate, S_2, for the test frame varies is presented in Fig. 1, while Fig. 2 presents visual results of reconstruction for $S_2 = 0.2$ for the *Football* sequence. As can be seen in Fig. 1, the proposed Tikhonov-regularized MH prediction provides superior recovery for x_2 across the range of tested subrates. Both MH and SH methods show significant performance gains over independent recovery, while the gap between MH and SH broadens as the subrate increases.

In terms of computation, SH prediction performs much more quickly than MH prediction, taking just 10 to 20 seconds in our MATLAB-based implementation. On the other hand, the Tikhonov regularization, which can take just a few minutes to calculate for an entire frame, appears to be a reasonable tradeoff between increased computation time and performance gain.

Conclusions

In this paper, we considered how the high degree of frame-to-frame temporal correlation in video signals can be exploited to enhance CS recovery by forming MH predictions using a distance-weighted Tikhonov regularization to find the best linear combination of hypotheses. The MH predictions were used to create a measurement-domain residual of the frame to be recovered—such a residual is typically much more compressible than the original frame making it much more amenable to CS recovery. This procedure for video recovery shows a significant distortion performance improvement over a straightforward recovery of the frames independently, as well as superior performance compared to recovery driven by SH prediction.

References

[1] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, Madrid, Spain, August 2006, pp. 1433–1452.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[3] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[4] M. F. Duarte, M. Wakin, and R. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden markov tree model," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, April 2008, pp. 5137–5140.

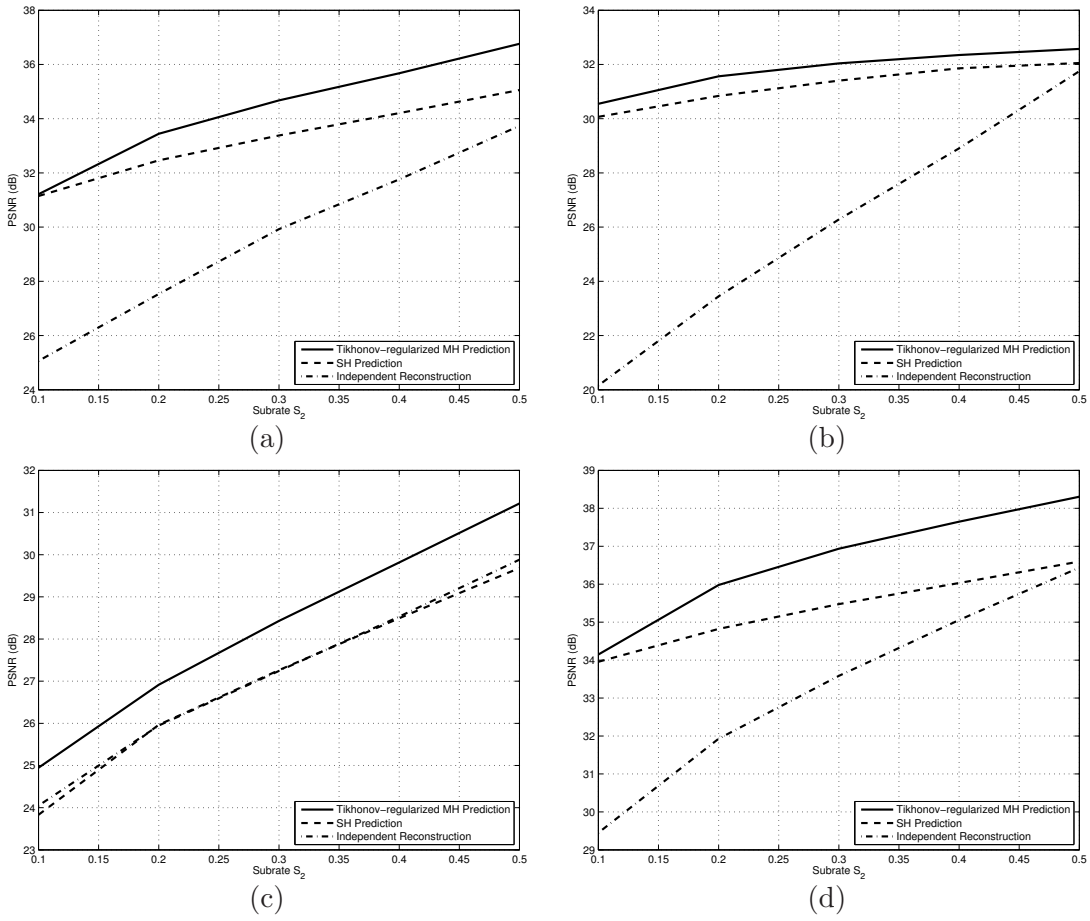

Figure 1: Recovery of frame x_2 of using frame x_1 as reference. (a) *Foreman*, (b) *News*, (c) *Football*, (d) *Susie*.

[5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[6] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, Minneapolis, MN, April 1993, pp. 437–440.

[7] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, D.C.: V. H. Winston & Sons, 1977.

[8] L. Gan, "Block compressed sensing of natural images," in *Proceedings of the International Conference on Digital Signal Processing*, Cardiff, UK, July 2007, pp. 403–406.

[9] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *Proceedings of the International Conference on Image Processing*, Cairo, Egypt, November 2009, pp. 3021–3024.

[10] M. F. Duarte, "Compressive sensing for signal ensembles," Ph.D. thesis, Rice University, 2009.

[11] 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 *Proceedings of the Picture Coding Symposium*, Beijing, China, April 2006.

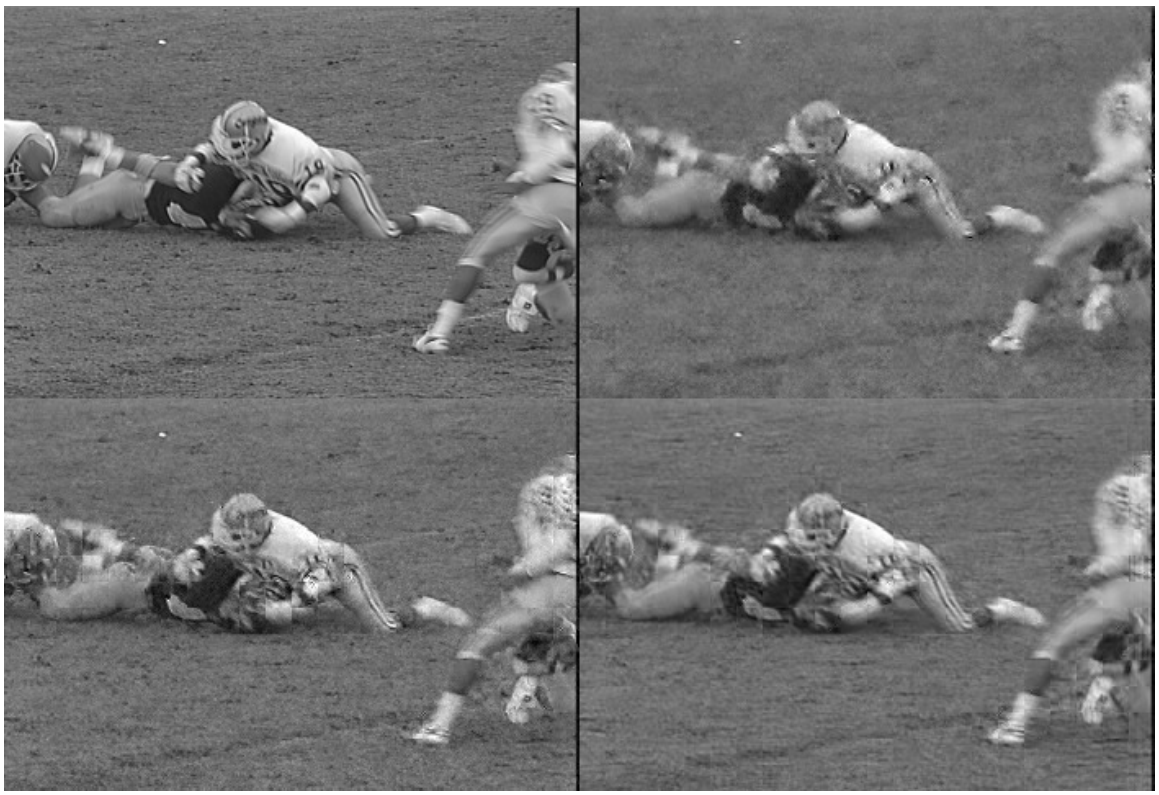

Figure 2: Recovery of frame x_2 of the *Football* sequence using frame x_1 as reference, $S_2 = 0.2$, $S_1 = 0.5$. Clockwise from upper left: original frame x_2; independent reconstruction (PSNR = 25.95 dB); recovery with Tikhonov-regularized MH prediction (PSNR = 26.91 dB); recovery with SH prediction (PSNR = 25.97 dB).

[12] S. Mun and J. E. Fowler, "Residual reconstruction for block-based compressed sensing of video," in *Proceedings of the IEEE Data Compression Conference*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, March 2011, pp. 183–192.

[13] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 1984.

[14] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60–65, January 2003.

[15] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and System Science*, vol. 66, no. 4, pp. 671–687, June 2003.

[16] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, April 1993.

[17] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 1, San Diego, CA, May 1992, pp. 184–187.

[18] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, February 1999.