DEMD-based Image Compression Scheme in a Compressive Sensing Framework

Mithilesh Kumar Jha   *jham73@gmail.com*
Department of Electrical Engineering, IIT, Delhi, INDIA

Brejesh Lall   *brejesh@ee.iitd.ac.in*
Department of Electrical Engineering, IIT, Delhi, INDIA

Sumantra Dutta Roy   *sumantra@ee.iitd.ac.in*
Department of Electrical Engineering, IIT, Delhi, INDIA

Abstract

Efficient representation of the background texture in video image frames, motivates compression strategies based on good perceptual reconstruction quality, instead of just bit-accurate reconstruction. This is especially true for video image frames in applications such as videos with structural patterns, and Bi-Directional Reflectance Distribution Function (BRDF) image frames of an object, where different images of an object in a single pose are taken in different illumination conditions. This paper investigates a new approach for an efficient representation of a class of images from textured videos and different BRDF images of an object, using sparse representation of the Directional Empirical Mode Decomposition (DEMD) residue of the frame. The efficient representation of the DEMD residue is achieved as a sparse coding solution based on a Discrete Wavelet Transform (DWT)-based sparsification. Experimental results demonstrate the effectiveness of the algorithm showing higher compression as compared to standard wavelet-based image compression schemes in a Compressive Sensing (CS) framework and JPEG2000, at similar perceptual reconstruction quality.

Keywords: Compressive Sensing (CS), Directional Empirical Mode Decomposition (DEMD), Discrete Wavelet Transform (DWT).

1. Introduction

1.1 Motivation

Textures are homogeneous patterns that contain spatial, temporal, statistical and perceptual redundancies, which intensity or colour alone cannot describe adequately because of its fine details and high frequency components. A variety of applications in computer vision, graphics, and image processing (such as robotics, defence, medicine, and geosciences) demand better compression with good perceptual reconstruction quality, instead of bit accurate (high PSNR) reconstruction (which existing image and video coding standards do not handle adequately). For homogeneous texture regions, maintaining the semantic meaning and coherence within the sequence is more important to the viewer than the specific actual pixel values. This is because the human brain is able to decipher important variations in data at scales smaller than those of the viewed object.

Conventional image compression schemes such as JPEG2000 [1] and video compression schemes such as MPEG2 [2] and H.264 [3] exploit statistical dependencies on the entire signal subject to a mean squared error (MSE) criterion. Efficient exploitation of statistical dependencies does not always correspond to the best psycho-visual result. Representing

textured video data using conventional approaches is not efficient primarily because of three reasons,

1. They are based on Shannon-Nyquist sampling [4] (and thus, do not fully exploit inherent sparsity present in the signal)

2. The acquisition and encoding are carried out on the entire signal, while most of the transformed data is discarded in the compression processes. (This wastes a significant amount of storage resources and has the overhead of a high computational cost. This is relevant for texture videos with fine detail and high frequency content.)

3. These schemes do not exploit the perceptual redundancy present in the texture data, thus limiting the amount of compression which can be achieved. This is an algorithmic limitation imposed by conventional schemes since for most application perceptual quality is more significant than bit accuracy.

The goal of a new generation lossy compression scheme should be to reduce the entropy, while preserving the perceptual quality of the frame using high-level computer vision tools and techniques. Most of these high-level processing techniques results in a small number of semantically relevant features, which can represent the salient features of the entire signal very accurately. Texture analysis and synthesis based compression scheme is one such method which addresses the issues with the texture data [5]. Texture analysis includes efficient representation and encoding of texture data, while the texture synthesis step reproduces a perceptually acceptable texture data at the decoder. Given this paradigm, a number of technical challenges arise, (i) To formulate an algorithm to efficiently represent and encode texture data, (ii) To create an approximation with subjective correspondence, and (iii) To efficiently synthesise a pixel at the decoder using reference information.

1.2 Contributions of this Paper, in a Nutshell

In this paper, we propose a novel approach for a more efficient representation and encoding of a class of video image frames from textured videos and different Bi-Directional Reflectance Distribution Function (BRDF) images of an object. The goal is to pre-process the input data in order to reduce the entropy of the frame and then use a texture model at the decoder to reconstruct these pixels with a satisfactory perceptual quality. The proposed approach uses a Compressive Sensing (CS)-based sparse representation of the Directional Empirical Mode Decomposition (DEMD) residue of the frame. The efficient representation problem of the DEMD residue of the frame, is modeled as a sparse coding solution based on a wavelet-based sparsification, whose solution is computed using convex optimisation. We do not encode original video image frame. Instead we encode the DEMD residue of all non-key frames in a Compressive Sensing (CS) framework, which due to the representation, is expected to have a significantly lower entropy, and can therefore be represented using a smaller number of bits. (Fig. 1 shows an example of the compression gain if we encode the DEMD residue frame instead of the original frame. The examples are for some representative sample texture video image frames used in this paper). The synthesis of non-key frames uses information from the key frames, and the corresponding DEMD residual image. The synthesis uses a correlation search and a patch-based multi-level IMF synthesis.

1.3 Related Work in the Area

Generally four approaches have been used to extract texture features: statistical-based methods, model-based methods, transform or spatial-frequency methods and structural
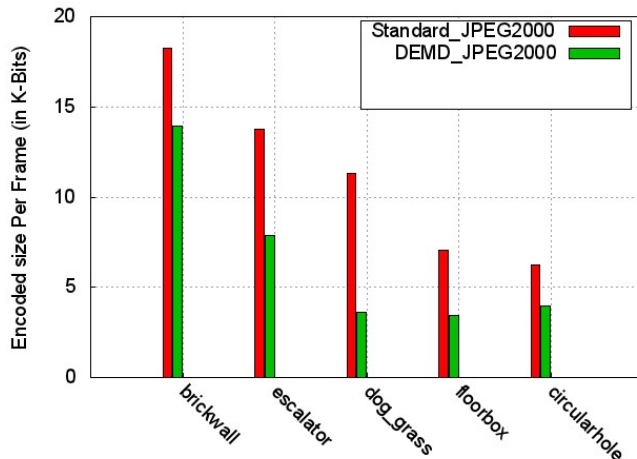
Fig. 1: DEMD residue encoding of a frame usually leads to better compression as compared to encoding the original video image frame: results for some standard test sequences. 'Standard_JPEG2000' indicates the video image frame encoded using the standard JPEG2000 encoder, and 'DEMD_JPEG2000' indicates the DEMD residue image encoded using JPEG2000.

methods. In a statistical-based methods, characteristics of homogeneous regions are chosen as the texture features such as the co-occurrence matrix or geometrical features such as edges [6]. Model-based methods assume that the texture is described by a stochastic model and use model parameters to segment the texture regions [7]. Wavelet-based Sub-band decomposition is often seen in spatial-frequency methods [8, 9]. Structural methods are based on the notion that textures are composed of primitives that are well-defined and spatially repetitive [10]. Bosch et al. [11] and Zhang and Bull [12, 13] have proposed spatial and spatio-temporal texture model-based analysis and synthesis schemes.

The texture synthesis process can be broadly classified as Parametric [14], or non-Parametric[15]. In the parametric approach, the image sequence is modeled by a number of parameters such as histograms, affine warpings, perspective warpings or correlation of pixel values. In the non-parametric approach, the synthesised texture is derived from an example texture as the reference that is known *a priori*. Non-parametric approaches are pixel-based or patch-based. Efros and Leung [15] propose pixel-based non-parametric sampling to synthesise texture. Wei and Levoy [16] further improve the above using a multi-resolution image pyramid based on a hierarchical statistical method. A limitation of the above pixel-based methods is an incorrect synthesis. This comes from incorrect matches, while searching for similar statistics. Patch-based methods overcome this limitation by considering features matching patch-boundaries with multiple pixel statistics. The popular choice for patch-based methods are based on graph-cut techniques [17, 18]. Non-parametric approaches can be applied to a wide variety of textures (with irregular texture patterns) and provide better perceptual results (higher Mean Opinion Score (MOS) values). However, these schemes are often computationally more complex.

Ndjiki-Nya et. al. [5] propose a texture analysis and synthesis framework using the frame-to-frame displacement and image warping techniques. This method is however not effective for non-rigid textured objects and the frame-by-frame synthesis can yield temporal inconsistencies. A similar texture synthesis technique based on texture analysis and segmentation using frame-to-frame mapping was proposed in [19]. Recent work in the parametric ap-

proach is typically based on Auto Regressive (AR) [7] or Auto Regressive Moving Average (ARMA)-based modeling [20, 21, 22]. AR- and ARMA-based models in texture synthesis enable blocks to be selectively removed at the encoding stage, and reconstructed at the decoder with acceptable perceptual quality. AR- and ARMA-based approach is suitable for the textures with stationary data, like steady water, grass and sky, however they are not suitable for structured texture with non-stationary data as blocks with non-stationary data are not amenable to AR modeling. Further, they are block-based approaches, and blocking artifacts can appear in the synthesised image. Region-based texture representation and synthesis algorithms [23, 24, 25] have been explored recently to address the limitations of the block based representation in handling the homogeneous texture and blocking artifacts. In a region-based approach, arbitrary shaped regions in image and video frames are segmented into homogeneous regions, each sharing similar texture properties. Byrne et al. [9, 18] has demonstrated region-based synthesis structure using a morphological and spectral image representation technique [26]. Region-based representations have also been explored using multi-resolution wavelet-based decomposition [27, 28, 29]. This work is limited to document class images.

The work of Zhang et al. [30] involves a DEMD-based synthesis approach with a small 64×64 representative texture patch. The authors start from a highest level (say, k) of the DEM decomposition of this patch. The k-th IMF corresponding to the full-resolution size is constructed by taking information from smaller 8×8 patches inside the sample, and placing them in the template for the full-resolution Intrinsic Mode Function (IMF) (with overlap, to enforce smoothness). For the next lower levels, the authors perform a correlation search for each 8×8 block in the full-resolution template, with the closest-matching 8×8 block in the $k - 1$th level decomposition of the 64×64 representative texture sample. The authors show good results for textured image synthesis with perceptually acceptable quality. However this approach leaves the following important questions unanswered: (a) The selection criterion for the small representative texture: its size (64×64) and location, (b) Selection criterion for the size of the smaller 8×8 patch, and the extent of the overlap (10-20% of the patch size in examples in the paper), (c) Objective assessment of the quality of the synthesised image, (d) The level of IMF decomposition required for optimal synthesis, (e)Inability to handle texture sequences with irregular patterns (since the empirically-assumed 64×64 sample patch may not represent the given texture completely, for an irregular texture, as the authors themselves mention in the paper [30]) and (f) Exploitation of the significant sparsity in the DEMD residue frames.

1.4 Our Work in Perspective

We propose a hybrid approach, combining the advantages of parametric and non-parametric methods. This enables us to handle a wider variety of textured videos and BRDF images than what can be handled by either method in isolation. DEMD can handle textures with stationary and non-stationary data. The proposed scheme is integrated into a compressive sensing framework to exploit the sparsity present in textures video image frames and BRDF images. We encode the DEMD residue of all non-key frames (equivalent to P and B in a MPEG/H264 terminology) in a compressive sensing (CS) framework. The synthesis of all non-key frames is achieved using correlation search and patch-based multi-level IMF synthesis from the DEMD representation of key frames and the DEMD residue of the frame being synthesised. Unlike [30], we provide both compression results, as well as an objective (PSNR) and perceptual (MOS-based) assessment of the synthesised frames. We use an energy-based heuristic to determine the number of decomposition levels in the DEMD.

4

This is estimated at the encoder and sent to the decoder as a control parameter. The decoder uses it for the decomposition and synthesis. Unlike the work of Zhang et al. [30], we perform the DEMD-based analysis of the full resolution frame. Hence, our approach does not suffer from the limitations of choosing a small representative patch size (which is not easy to identify in an irregular texture pattern).

To the best of our knowledge, no related work address these issues. A very preliminary work in this direction was our earlier work [31]. The current submission differs considerably from the above preliminary work in the following ways. The earlier work encoded the IMFs' maxima and minima for multiple IMF levels, instead of the more efficient residue encoding in the current work. Moreover, the DEMD residue frame encoding in a CS framework - is a major contribution of the current submission. A heuristic in the earlier work [31] was to drop the residue (assuming that the residue was not sufficiently informative), leading to comparatively worse reconstruction quality, as well as affecting compression (since it encoded the IMF extrema as well).

The rest of the paper is organised as follows. Sec. 2, provides an overview of DEMD and CS-based image encoding and decoding framework. Sec. 3 provides details of our test set-up and experimental results followed by the conclusions in Sec. 4.

2. DEMD and CS-based Image Coding

Huang et al. [32] proposes the 'Empirical Mode Decomposition(EMD)' of a signal $f(t)$ into a 'low frequency/residue' term $r_1(t)$ and a 'high frequency/detail' part $i_1(t)$. Decomposing each $i_k(t)$ gives a K-level decomposition according to Eq. 1.

$$f(t) = \sum_{k=1}^{K} i_k(t) + r_K(t) \tag{1}$$

Here, functions $i_k(t)$ are the Intrinsic Mode Functions (IMFs) [32, 33, 34]. Such a decomposition enables us to handle any class of signals, non-linear, linear, non-stationary, and stationary. The decomposition at each level is a simple numerical technique, and has an associated local scale (of an oscillation), and involves a perfect reconstruction. DEMD is the extension of the EMD concept to 2-D, in the work of Liu et al. [35, 36, 37]. DEMD considers the decomposition in a dominant image direction θ (which is natural for textures):

$$f(x, y) = \sum_{k=1}^{K} i_k^{\theta}(x, y) + r_K^{\theta}(x, y) \tag{2}$$

A Wold decomposition-based method [38] helps estimate θ, using the maximum of the Radon transform of the spectrum [30]. Using DEMD for image and video application demands efficient representation and coding of the 'low frequency/residue' term r_K in Eqs. 1, 2. In this paper, we model this as a sparse coding solution in a compressive sensing framework. Compressive Sensing (CS) [39, 40] enables the representation of a real valued P-sparse signal $\mathbf{x} \in \mathbf{R}^N$, using only P-basis vectors (where $P << N$) as Eq. 3. The DEMD residue fits naturally in a sparse representation (for a sparse signal of length N, we can represent it with $P << N$ nonzero coefficients).

In Compressive Sensing (CS), most sparse signals can be reconstructed from a small number of measurements (M), using algorithms like convex optimisation, greedy methods and iterative thresholding [41, 42]. Significant theoretical contributions have been published on the compressive sensing framework in recent years [39, 40, 41] targeting to image and video compression applications [43, 44, 45, 46, 47, 48]. Compressive Sensing framework mainly
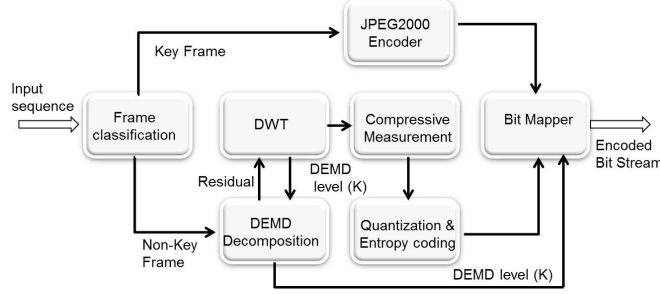
Fig. 2: The Encoder framework

consists of three stages i.e., sparsification by transformation, measurement (projection) and optimisation (reconstruction) and the CS recovery algorithms mainly depend on two fundamental principles i.e., Sparsity and Incoherence [39]. Designing a good measurement matrix with large compression effects and designing a good signal recovery algorithm, are the two major challenges of applying the CS technique in image frame compression.

$$\mathbf{x} = \sum_i^N s_i \psi_i \qquad or \qquad \mathbf{x} = \Psi\mathbf{S}$$
$$\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\mathbf{S} = \Theta\mathbf{S} \tag{3}$$

Here y is an $M \times 1$ measurement vector (where $M << N$). The Ψ is an $N \times N$ sparsifying basis matrix and Φ is an $M \times N$ sensing matrix. The matrix Φ represents a dimensionality reduction matrix i.e., it maps $\mathbf{R}^N$ to $\mathbf{R}^M$, where M is typically much smaller than N. The coherence measures the largest correlation between any two elements of the sensing basis (Φ) and the sparsifying basis (Ψ) and is defined as Eq. (4). Signal $\mathbf{x}$ can be reconstructed from measurement vector $\mathbf{y}$ (solving the convex optimisation problem), if Θ satisfies a Restricted Isometry Property (RIP) [40].

$$\mu(\Phi, \Psi) = \sqrt{N}.\max_{1 \leq P, j \leq N}\langle\langle\phi_P, \psi_j\rangle\rangle, \ 1 \leq \mu(\Phi, \Psi) \leq \sqrt{N} \tag{4}$$

The following sections, we discuss our encoder and decoder framework.

2.1 CS-based Encoder Framework

We consider a test sequence as being composed of a set of frames, where each set of frames has a *key* frame (equivalent to the *I* frame in MPEG/H264 terminology), followed by a set of *non-key* frames. We can have either a fixed number of such *non-key* frames, or have this varied adaptively according to the motion in adjacent frames, and the associated propagation error. Fig. 2 gives an overall block diagram of the proposed encoder.

A *key* frame is sent to a JPEG2000 encoder, which creates the encoded bit-stream. For a *non-key* frame to be encoded, the system creates the DEMD decomposition of the frame according to Eq. 2. We choose the computationally light Haar wavelets (O(N) for an N-point signal to decide on the number of decomposition levels K for the DEMD process. We send K to the decoder as a side information with the encoded bit stream header corresponding to non-key frames. We compute the ratio of the energies in the HH and LL sub-bands. We stop the DEM decomposition once this ratio becomes a small fraction of that for the original input frame. We selected this fraction empirically (0.01) based on monitoring the
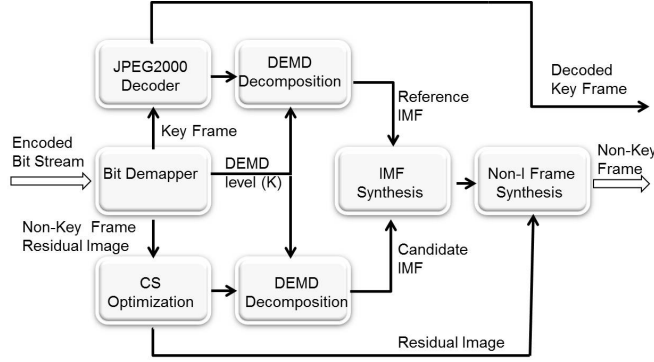
Fig. 3: The Decoder framework

synthesis quality (MOS as well as PSNR), on a wide range of representative textured video image frames. This ensures that most of the information in the signal has been captured in the K levels.

The encoder sends the penultimate-level residues $\{r_{K-1}^{\theta}(x, y)\}$ to a CS sensing block (*We send the $(K-1)$th level residue to enable the decoder to have some information content, to enable it to perform a search for the closest patch in the DEMD decomposition image at a particular level*). We use DWT for the sparsifying matrix (Ψ) and Noiselet [49] for sensing matrix (Φ) for CS measurements according to Eq. 3. We have chosen Noiselet for CS Measurements because they are highly incoherent with the considered sparse domain and the RIP tends to hold true for reasonable values of M. Moreover Noiselets comes with very fast algorithms and just like the Fourier transform, the Noiselet matrix does not need to be stored to be applied to a vector. This is very important for efficient numerical computations without which applying CS can be very complex. In addition, the CS measurement matrix created is orthogonal and self-adjoint, therefore it is easy to manipulate. Scalar quantization over the measurement vector is proposed to achieve better compression ratio. The motivation behind this is coming with the fact that there will be a lot of spatial redundancy across frames from the same class of video image frame and BRDF images. Thus, it is reasonable to assume spatial correlation of a kth level decomposition between a *key* and a *non-key* frame. A *key* frame is encoded using JPEG2000 to have maximum possible information in the encoded bit stream. For a *non-key* frame, we encode $(K-1)$ level residue in a CS framework, which is expected to contain far less information as compared to the original frame, to enable it to be more efficiently compressed as a bit stream.

2.2 CS-based Decoder Framework

The decoder receives the encoded bit stream, with each sequence composed of a *key* frame, the DEMD decomposition level (K) as a control parameter and a set of *non-key* $(K-1)$th level DEMD residue frames $\{r_{K-1}^{\theta}(x, y)\}$. Fig. 3 shows a block level view of the proposed decoder framework.

$$min\|\mathbf{x}\|_1 \quad subject \quad to \quad \Theta\mathbf{x} = s_a, \quad \|\mathbf{x}\|_1 = \sum_i |x_i| \tag{5}$$

The decoder first decomposes the decoded *key* frame into K DEMD components $i_k^{\theta_k, I}(x, y)$. The decoder also estimates θ exactly as the encoder does it. For every *non-key* frame, the decoder starts from the CS optimisation solution of the $(K-1)$th level residue frame. In all our experiments, we have used the solution of the convex optimisation problem according

to Eq. 5 (l_1-norm minimization with equality constraint) for good recovery. This matches with the l_0 norm as the RIP is ensured due to Noiselet measurements. This optimisation principle leads to sparser decompositions [41] and ensures fast and stable resolution better than other methods from the class of greedy algorithms, due to non-differentiability of the l_1-norm. We have chosen l_1 norm as it provides a computationally viable approach to sparse signal recovery.

The decoder performs an additional decomposition of $r_{K-1}^{\theta,non-key}(x,y)$ into $i_K^{\theta,non-key}(x,y)$ and $r_K^{\theta,non-key}(x,y)$. As mentioned in the previous section (Sec. 2.1), we assume that the decomposition proceeds till there is very little information content in the Kth level residue. The $(K-1)$th level residue is the last decomposition level with some useful information in the residue image $r_{K-1}^{\theta,non-key}(x,y)$. The decoder does a correlation-based search for each block in the target IMF image $i_{K-1}^{\theta,non-key}(x,y)$, with a block in the corresponding *key* frame IMF image $i_{K-1}^{\theta,key-frame}(x,y)$, with an overlap. Each pixel in the IMF image has a 3-tuple feature vector: the magnitude ('instantaneous bandwidth'), and the two angular directions (the 'instantaneous frequencies') [50, 30]. To get the best matching patch, we do a local search in two 8×8 blocks on each side of the candidate block in the upper level newly-synthesised IMF image (with overlap), and select the one with the least Euclidean distance from the corresponding block in the *key* frame IMF. We repeat the procedure for the other decomposition levels as well, using the same correspondence information for the blocks from the correspondence information at level $(K-1)$.

3. Experiments and Results

In this section, we present the experimental results for representative video frames having structural textures (*brickwall, escalator, floorbox, dog_grass and blackhole*) and BRDF images (*brdf*) which differ in terms of both sparsity and high energy (see Fig. 5 and 4). All the test sequences are 128x128, 8-bit, gray scale videos. The experiments are carried out in MATLAB and the l_1-$Magic$ [51]. The Wavelet bases are generated using the WAVE-LAB [52] package and the Noiselet bases using an algorithm as suggested in [53]. The rate in Fig. 6 is calculated as $(M/N)*H$, where H is the entropy (in bits per pixel, empirically chosen as 8 for all our experiments). PSNR and MOS have been used as the quality metrics. Mean Opinion Score (MOS) computation was done by collecting responses of various students and staff working in our lab and averaging them.

We present the synthesis results of our proposed scheme (DEMD-DWT-CS) and conventional DWT-based texture synthesis scheme in a CS framework (DWT-CS) [54, 44]. We have compared our simulation results with the conventional DWT-based image synthesis scheme in a CS framework as presented in [54, 44]. In addition, we have done a comparative study of our synthesis results with standard JPEG2000 (for both actual and DEMD residue image coding). The texture synthesis performance is measured using different CS measurement samples. We make the following observations:

- Fig. 6 demonstrates the rate-distortion curve for all test sequences used in our experiments. The proposed scheme (DEMD-DWT-CS) outperforms the conventional DWT-CS scheme and can provide significantly better PSNR for the same compressive measurements (bits/pixel value in Fig. 6) or better compression at the same PSNR. The CS recovery scheme that performs the l_1-$norm$ minimization on the actual image in wavelet domain is far less efficient than performing the minimization on the DEMD residue image. In addition, we can observe that the synthesis quality can be improved by increasing the number of CS measurement samples (bits/pixel value).

Table 1: Even a single level of IMF decomposition provides considerable compression gain over a standard DWT-CS scheme [54, 44], at acceptable quality levels (PSNR as well as MOS)

Sequence	%Gain over DWT-CS	Quality (PSNR, dB) DWT-CS	Quality (PSNR, dB) Our Scheme	Quality (MOS) DWT-CS	Quality (MOS) Our Scheme
brickwall	27.63	20.05	22.99	4.2	4.3
escalator	46.00	19.26	27.99	4.0	4.5
floorbox	33.03	30.66	32.3	4.5	4.5
dog_grass	62.94	28.24	31.75	4.5	4.5
blackhole	33.59	29.87	32.58	4.5	4.5
brdf	40.48	22.23	25.26	4.5	4.5

- When compared with JPEG2000, coding the DEMD residue image than the actual image using JPEG2000: provides much better PSNR at same rate or better compression at the same PSNR value for all class of images (see Fig. 6). It is important to note that low bit-rate encoding and recovery in a CS framework is generally less efficient as compared to standard JPEG2000. The reason for this is the presence of significant energy in the high frequency region for structural textures. The coding efficiency loss in a scene with very high frequency transition is due to the fact that compressive sensing recovery of sparse signals typically requires a number of measurements to be larger than the number of non-zero samples.

- Table 1 shows sample compression results with just a single level of IMF decomposition. Figs. 5 and 4 show the synthesis results of *non-key* frames using the proposed scheme. All images shown with a prominent boundary are synthesised images. The DEMD residue image corresponding to *non-key* frames are synthesised using convex optimisation in the compressive sensing framework, and shown in the top row in Figs. 5 and 4. The IMF images corresponding to *non-key* frames are synthesised (see Figs. 5 and 4) using the DEMD representation of *key* frames and the DEMD residual image of the frame being synthesised. As one can observe, our proposed scheme can synthesise *diverse* structural patterns: periodic (*brickwall and escalator* and statistical (*circularhole, floorbox and dog_grass*) with acceptable synthesis quality.

4. Conclusion

In this paper, we propose a novel approach for a more efficient representation and encoding of a class of video image frames from textured videos and different BRDF images of an object, using DEMD in a compressive sensing framework. We encode the DEMD residue of all non-key frames instead of encoding the original frame (as normally done in a standard coding scheme) in a compressive sensing framework. All *non-key* frames from textured videos and different BRDF images of an object are synthesised using information from the key frames (equivalent to an *I* frame in MPEG/H.264), and the corresponding DEMD residual image of the frame being synthesised. The synthesis uses a correlation search and a patch-based multi-level IMF synthesis. All the *Key* frames are coded using JPEG2000 to preserve the maximum entropy in the reference data. Experimental results show that the proposed scheme provides better compression results as compared to standard wavelet-based image coding schemes in a CS framework and JPEG2000, with similar reconstruction quality.

References

[1] I.-T. R. T.800, "Jpeg-2000:core coding system," *International Telecommunication Union, Tech. Rep.*, 2006.

[2] ITU-T Rec. H.262 and ISO/IEC 13818-2 MPEG-2, "Generic Coding of Moving Pictures and Associated Audio Information - Part 2 Video," 1994.

[3] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4/AVC), "Advanced video coding for generic audio visual services," *Standard version 7, ITU-T and ISO/IEC JTC 1*, 2007.

[4] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. AIEE*, pp. 617 – 644, 1928.

[5] P. Ndjiki-Nya, B. Makai, A. Smolic, H. Schwaz, and T. Wiegand, "Improved H.264/AVC Coding using Texture Analysis and Synthesis," *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 2, pp. 849 – 852, 2003.

[6] I. Elfadel and R. Picard, "Gibbs random fields, cooccurrences, and texture modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pp. 24 – 37, 1994.

[7] A. Khandelia, S. Gorecha, B. Lall, S. Chaudhury, and M. Mathur, "Parametric Video Compression Scheme using AR based Texture Synthesis," *Proc. IAPR- and ACM-sponsored Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, pp. 219 – 225, 2008.

[8] J. R. Smith and S.-F. Chang, "Quad-tree segmentation for texture-based image query," *Proceedings of the second ACM International Conference on Multimedia*, pp. 279 – 286, 1994.

[9] J. Byrne, S. Ierodiaconou, D. R. Bull, D. Redmill, and P. Hill, "Unsupervised image compression-by-synthesis within a jpeg framework," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 2892 – 2895, 2008.

[10] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, pp. 786 – 804, 1979.

[11] M. Bosch, F. Zhu, and E. J. Delp, "Video coding using motion classification," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 1588 – 1591, 2008.

[12] F. Zhang, D. R. Bull, and N. Canagarajah, "Region based texture modelling for next generation video codecs," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 2593 – 2596, 2010.

[13] F. Zhang and D. R. Bull, "Enhanced video compression with region based texture models," *Picture Coding Symposium(PCS)*, pp. 54 – 57, 2010.

[14] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," *in Proc. IEEE Workshop Statist. Comput. Theories Vision*, pp. 1 – 32, 1999.

[15] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," *International conference on computer vision*, pp. 1033 – 1038, 1999.

[16] L. Y. Wei and M. Levoy, "Fast Texture Synthesis using tree-structured vector quantization," *in Proc. SIGGRAPH, New Orleans, Louisiana, USA*, pp. 479 – 488, 2000.

[17] V. Kwatra, A. Schodl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures : image and video synthesis using Graph Cuts," *in Proc. SIGGRAPH, San Diego, CA, USA*, pp. 277 – 286, 2003.

[18] S. Ierodiaconou, J. Byrne, D. R. Bull, D. Redmill, and P. Hill, "Unsupervised image compression using graphcut texture synthesis," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 2289 – 2292, 2009.

[19] M. Bosch, F. Zhu, and E. J. Delp, "Spatial texture models for video compression," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 93 – 96, 2007.

[20] A. Stojanovic, M. Wien, and J. R. Ohm, "Dynamic texture synthesis for H.264/AVC inter coding," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 1608 – 1612, 2008.

[21] A. Stojanovic, M. Wien, and T. K. Tan, "Synthesis-in-the-loop for video texture coding," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 2293 – 2296, 2009.

[22] H. Chen, R. Hu, D. Mao, R. Thong, and Z. Wang, "Video coding using dynamic texture synthesis," *IEEE International Conference on Multimedia and Expo*, pp. 203 – 208, 2010.

[23] P. Hill, "Wavelet based texture analysis and segmentation for imgae retrieval and fusion," *PhD,University of Bristol*, 2002.

[24] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Transaction on Image Processing*, pp. 625 – 638, 1994.

[25] A. Dumitras and B. G. Haskell, "An Encoder-Decoder Texture Replacement Method with Application to Content-Based Movie Coding," *IEEE Transactions on Circuits and Sytems for Video Technology*, vol. 14, pp. 825 – 840, 2004.

[26] R. O'Callaghan and D. Bull, "Combined morphological-spectral unsupervised image segmentation," *IEEE Transaction on Image Processing*, pp. 49 – 62, 2005.

[27] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet domain hidden markov model," *IEEE Transaction on Image Processing*, pp. 1309 – 1321, 2001.

[28] J. Li and R. M. Gray, "Context-based multiscale classification of document images using wavelet coefficient distribution," *IEEE Transaction on Image Processing*, pp. 1604 – 1616, 2000.

[29] M. Acharya and M. K. Kundu, "An adaptive approach to unsupervised texture segmentation using m-band wavelet," *Signal Processing*, pp. 1337 – 1356, 2001.

[30] Y. Zhang, Z. Sun, and W. Li, "Texture Synthesis based on Directional Empirical Mode Decomposition," *Computer and Graphics*, pp. 175 – 186, 2008.

[31] M. K. Jha, B. Lall, and S. Dutta Roy, "Video Compression Scheme Using DEMD Based Texture Synthesis," in *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 90 – 93, 2011.

[32] N. E. Huang, Z. Shen, and S. R. Long, "The Empirical Mode Decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of Royal Society of London (Series A)*, vol. 1, pp. 903 – 995, 1998.

[33] A. Linderhed, *Empirical Mode Decomposition as Data-Driven Wavelet-Like Expansions*. PhD thesis, Dissertation No. 909, Linkoping University, 2004.

[34] P. Flandrin and P. Goncalves, "Adaptive Image Compression with Wavelet Packets and Empirical Mode Decomposition," *International Journal of Wavelets, Multiresolution and Information Processing*, pp. 1 – 20, 2004.

[35] Z. Liu, H. Wang, and S. Peng, "Texture classification through Directional Empirical Mode Decomposition," *Proc. International Conference on Pattern Recognition (ICPR)*, vol. 4, pp. 803 – 806, 2004.

[36] Z. Liu, H. Wang, and S. Peng, "Texture Segmentation using Directional Empirical Mode Decomposition," *Proc. IEEE International Conference on Image Processing (ICIP)*, vol. 1, pp. 279 – 282, 2004.

[37] Z. Liu and S. Peng, "Directional Empirical Mode Decomposition and its application to Texture Segmentation," *Science in China Series E: Information Sciences*, vol. 2, pp. 113 – 123, 2005.

[38] J. Francos, A. Meiri, and B. Porat, "A unified texture model based on a 2-d wold-like decomposition," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2665 – 2678, 1993.

[39] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transaction on information Theory*, pp. 489 – 509, 2006.

[40] D. Donoho, "Compressed sensing," *IEEE Transaction on information Theory*, pp. 1289 – 1306, 2006.

[41] R. G. Baraniuk, "Compressed sensing[lecture notes]," *IEEE Signal Processing Magazine*, pp. 118 – 121, 2007.

[42] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Scientific Computing Vol. 20*, pp. 33 – 61, 1998.

[43] Y. Zhang, S. Mei, Q. Chen, and Z. Chen, "A novel image/video coding method based on compressive sensing theory," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1361 – 1364, 2008.

[44] D. Venkatraman and A. Makur, "A compressive sensing approach to object-based surveillance video coding," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3513 – 3516, 2009.

[45] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," *Picture Coding Symposium (PCS)*, pp. 1 – 4, 2009.

[46] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressed imaging for video representation and coding," *Picture Coding Symposium*, pp. 1289 – 1306, 2006.

[47] Y. Yang, O. C. Au, L. Fang, X. Wen, and W. Tang, "Perceptual compressive sensing for image signals," *IEEE International Conference on Multimedia and Expo*, pp. 89 – 92, 2009.

[48] C. Li, H.Jiang, P. Wilford, Y. Zhang, and M. Scheutzow, "A new compressive video sensing framework for mobile broadcast," *IEEE Transcations on Broadcasting*, pp. 197 – 205, 2013.

[49] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comp. Harman. Anal. vol. 10, no. 1*, 2001.

[50] T. Bulow and G. Sommer, "Hypercomplex signals- a novel extension of the analytic signal to the multidimensional case," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2844 – 2852, 2001.

[51] E. Candes and J. Romberg, "L1-magic," *www.l1-magic.org.*

[52] D. Donoho, A. Maleki, and M. Shahram, "Wavelab," *http://www-stat.stanford.edy/wavelab.*

[53] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, pp. 14 – 20, 2008.

[54] A. Schulz, L. Velho, Eduardo, and A. B. da Silva, "On the empirical rate-distortion performance of compressive sensing," *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 3049 – 3052, 2009.
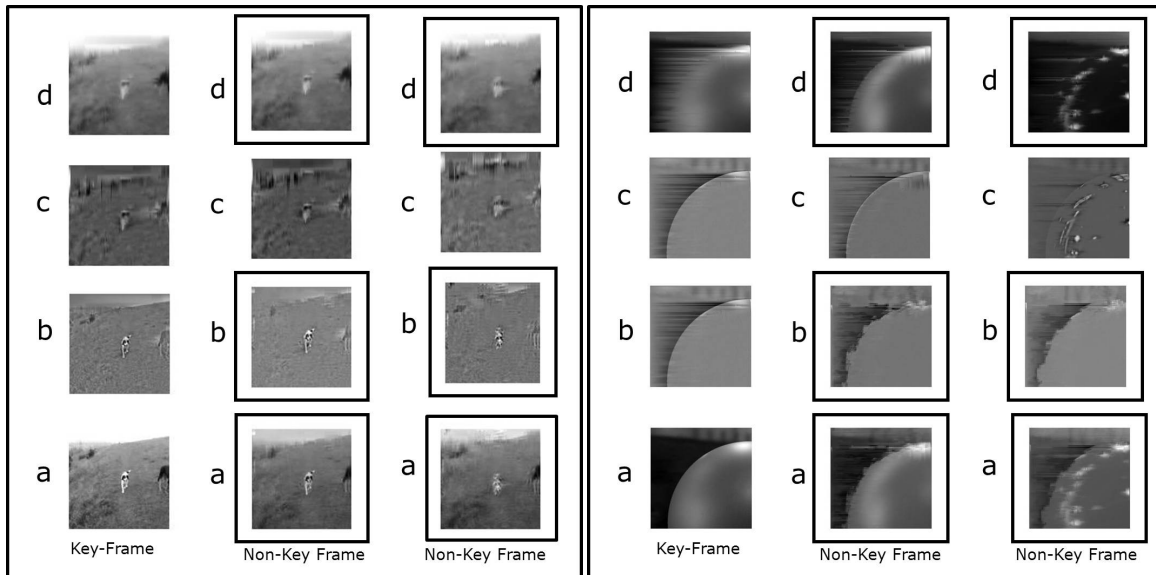

Fig. 4: Synthesis results for sequences *dog_grass*, *brdf*: a. Input *key*-frame and synthesised *non-key* frames, b. 1st level IMF image for *key*-frame and 1st level synthesised IMF image for *non-key* frames, c. 2nd level IMF image, and d. DEMD residue image for *key*-frame and CS synthesised DEMD residue image for *non-key* frames
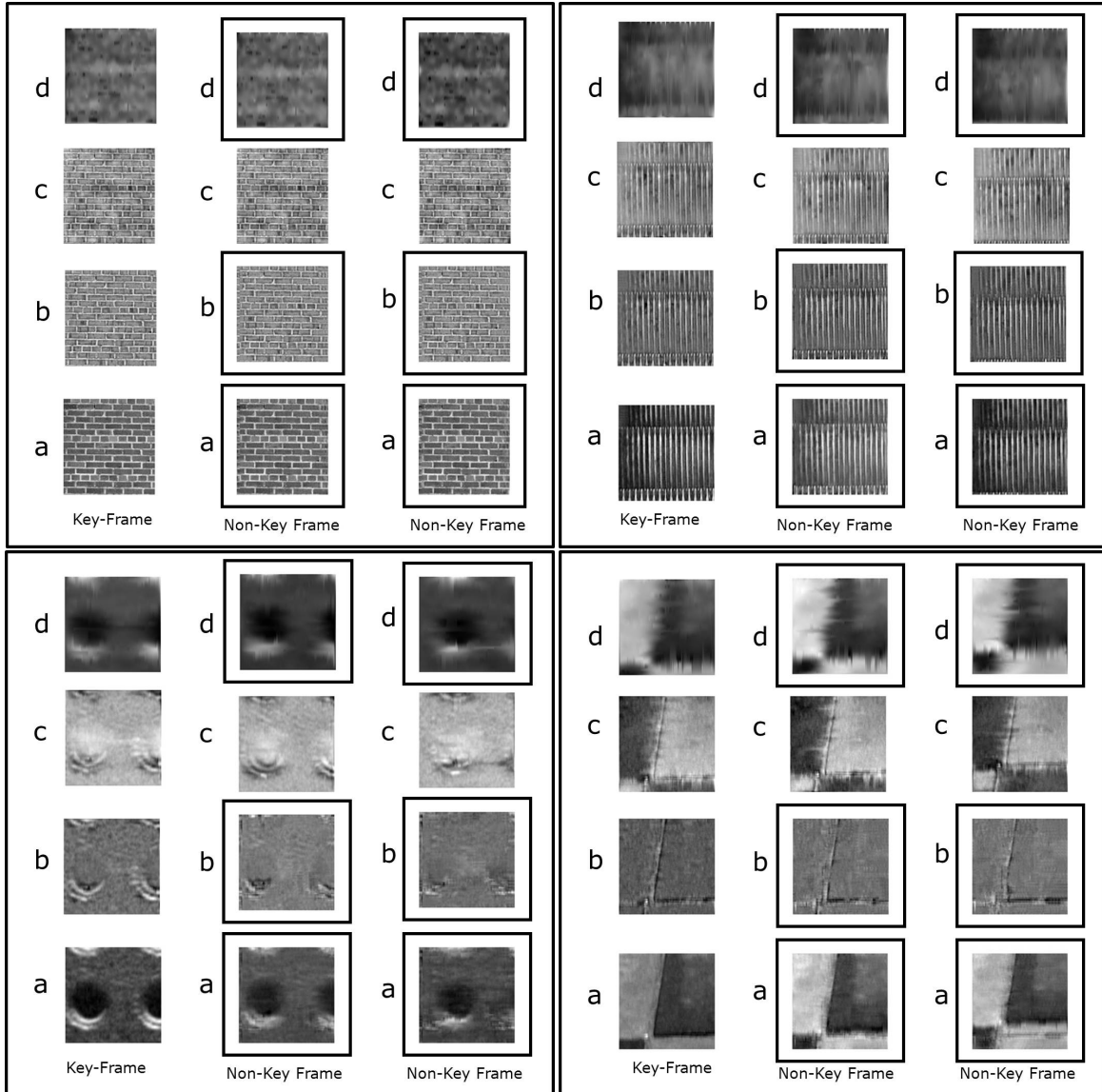
Fig. 5: Synthesis results for sequences *brickwall, escalator, blackhole* and *floorbox*: a. Input *key*-frame and synthesised non-*key* frames, b. 1st level IMF image for *key*-frame and 1st level synthesised IMF image for non-*key* frames, c. 2nd level IMF image, and d. DEMD residue image for *key*-frame and CS synthesised DEMD residue image for non-*key* frames
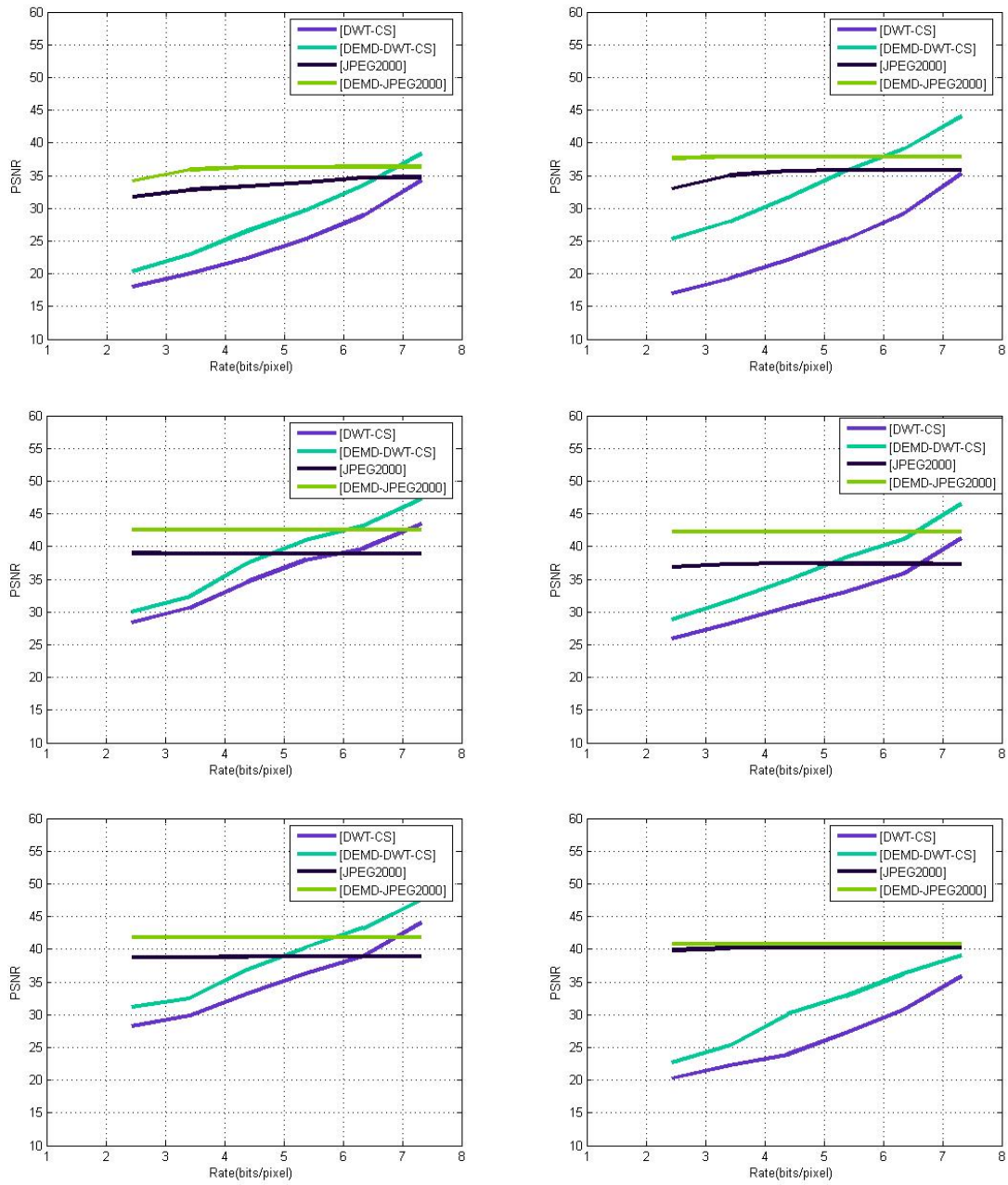
Fig. 6: Rate Distortion results for test sequences *brickwall*, *escalator*, *floorbox*, *dog_grass*, *circularhole* and *brdf_image* in raster-scan order: 'DWT-CS' indicates the coding of actual frame and 'DEMD-DWT-CS' indicates the coding of DEMD residue frame in a Compressive Sensing (CS) framework. 'JPEG2000' indicates the JPEG2000 coding of the actual frame, and 'DEMD-JPEG2000' indicates the coding of DEMD residue frame in the JPEG2000 framework. CS recovery that performs l_1-*norm* minimization on the DEMD image is far more efficient than the CS recovery of the actual image in the wavelet domain. Similarly, the JPEG2000 coding of a DEMD residue image is more efficient than the coding of the actual image.