Multiple Description Image Coding with Local Random Measurements

Xianming Liu[1,2], Xiaolin Wu[2,3], Debin Zhao[1]
[1]School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, China, 150001
[2]Department of Electrical & Computer Engineering,
McMaster University, Hamilton, Canada, L8S4K1
[3]Institute of Image Communication and Information Processing,
Shanghai Jiao Tong University, Shanghai, 200240, China
{xmliu,dbzhao}@jdl.ac.cn xwu@ece.mcmaster.ca

Abstract: In this paper, an effective multiple description image coding technique is developed to achieve competitive coding efficiency at low encoder complexity, while being standard compliant. The new technique is particularly suitable for visual communication over packet-switched networks and with resource-deficient wireless devices. To keep the encoder simple and standard compliant, multiple descriptions are produced by quincunx spatial multiplexing. Each side description is a polyphase downsampled version of the input image, but the conventional low-pass filter prior to downsampling is replaced by a local random binary convolution kernel. The pixels of each resulting side description are local random measurements and placed in the original spatial configuration. The advantages of local random measurements are two folds: 1) preservation of high-frequency image features that are otherwise discarded by low-pass filtering; 2) each side description remains a conventional image and can therefore be coded by any standardized codec to remove statistical redundancy of larger scales. The decoder performs joint upsampling of received description(s) and recovers the image from local random measurements in a framework of compressive sensing. Experimental results demonstrate that the proposed multiple description image codec is competitive in rate-distortion performance compared with existing methods, with a unique strength of recovering fine details and sharp edges at low bit rates.

Key words: Multiple description coding, random sampling, compressive sensing, sparse representation.

1 Introduction

Multiple description coding (MDC) has been proven to be a promising technique to cope with transmission errors when compressed media contents are delivered via error-prone channels [1][2]. In the MDC approach, the source data are coded into two or more code streams, called side descriptions, which are transmitted via lossy diversity channels/network. There is no guarantee that the decoder will receive all side descriptions. The MDC design objective is to maximize the quality of the reconstructed signal using whatever descriptions received; the more side descriptions received, the higher the reconstruction fidelity. Unlike

1068-0314/14 $31.00 © 2014 IEEE
DOI 10.1109/DCC.2014.64

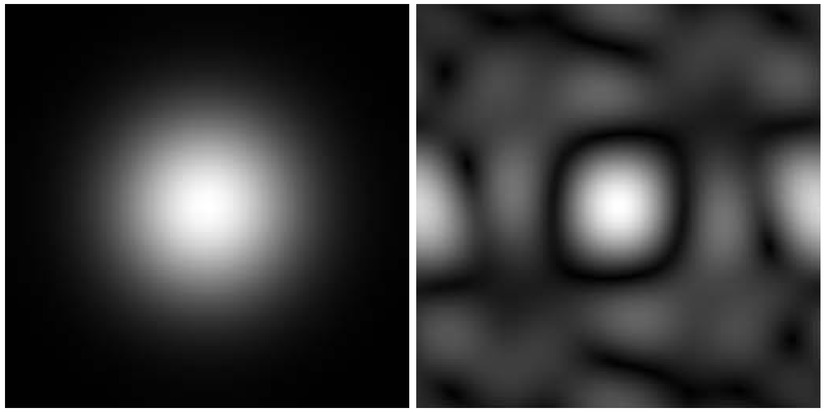

Figure 1: Fourier spectrums of Gaussian kernel (left) and local random convolution kernel (right).

the layered or progressive transmission schemes [3][4], where code layers can only be decoded in a particular sequence, MDC does not have algorithmic interdependence of side descriptions; any subset of side descriptions can be decoded. This property lends MDC greater flexibility and robustness to survive complex adverse network conditions.

In the literature, there are four types of MDC techniques: 1) polyphase downsampling-based multiplexing in spatial (PDSM) or transform domain (PDTM)[13][14], 2) multiple description quantization (MDQ) [5][6][7], 3) correlation transform (CT) [8]-[11], and 4) uneven erasure protected packetization (UEP) [12]. All these MDC techniques, except the polyphase downsampled based methods, are not compatible with existing image compression standards such as JPEG and JPEG 2000; they adopt completely different coding methodologies from existing image/video compression standards and cannot be used without significantly changing the standardized code stream syntax.

In the PDSM approach to MDC [13], multiple side descriptions are different polyphase downsampled versions of the input image, each of which remains a conventional image of pixels on rectangular lattice; thus each side description can be coded by any compression standard as is. The PDSM-based MDC approach is completely standard compliant, because the MD encoder acts merely as a preprocessor (downsampling) and the MD decoder as a postprocessor (upsampling) to the standardized codec. The conventional preprocessing before downsampling is low-pass filtering for the purpose of removing aliasing at the MD reconstruction stage. However, this also prevents the recovery of sharp spatial details beyond the cut-off frequency of the low-pass filter. Considering the importance of edges and other structured high-frequency features to perceptual quality, we propose to replace the low-pass downsampling filter, such as Gaussian and box filters, by a $w \times w$ binary convolution kernel of random coefficients. As the latter is a broad-band filter (see Fig. 1 to compare the spectrum responses of random convolution and Gaussian kernels), it retains high-frequency information in the down-sampled image and hence leaves the MD decoder the possibility of reconstructing sharper and clearer images than the low-pass prefilters.

To summarize we depict the architecture of the proposed local random filtering based multiple description image coding system in Fig. 2. To avoid unnecessary clutters, the sys-

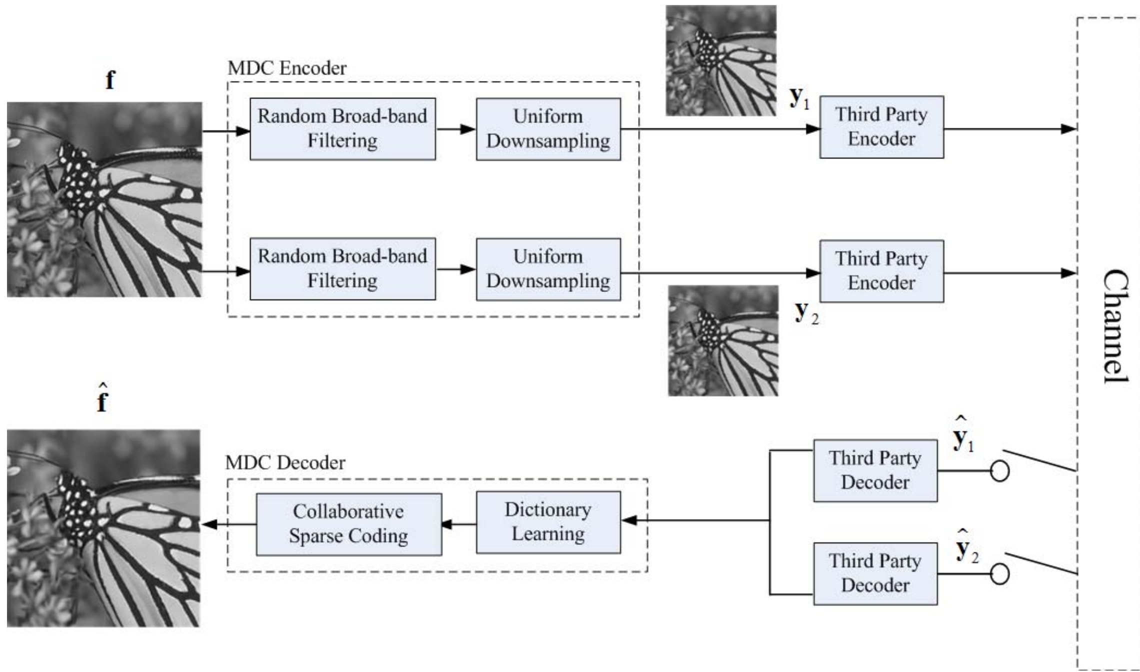

Figure 2: Block diagram of the proposed multiple description image coding system

tem is illustrated for two descriptions. Note that the value of a pixel in a downsampled image (side description), after the random broad-band filtering and polyphase downsampling, can be viewed as a random measurement of a $w \times w$ image window. Although pixel values are local random measurements, the downsampled images can and should be compressed by a third party or standard image codec, because natural images comprise large smooth regions and thus statistical redundancies between the pixels still exist. As shown in Fig. 2, the system decoder side is a cascade of the straightforward hard decoding of $k \geq 1$ received description(s) and a sparsity-based reconstruction of the transmitted image using all the kWH local random measurements of the image, where $W \times H$ is the resolution of the downsampled images. The thrust of this research, which is also our main contribution, is to develop sparsity-based soft-decoding technique for compressive sensing image recovery. Specifically, in addition to relying on constraints provided by local random measurements, the soft MD decoding incorporates a sparse image representation learnt from the pixel patches of hard decoded side description(s). The said sparse representation is in the form of PCA dictionary in which pixel patches are of sparse coding in the PCA bases. By combining the constraints of the local random measurements and the learnt sparse representation, the soft MD decoder exploits both the high-frequency information retained by the broad-band filter of random convolution and the self-similarities of pixel patches at larger scales. This novel image restoration strategy gives the proposed soft MD decoder its superior rate-distortion performance to other standard-compliant image MDC methods.

The rest of the paper is organized as follows. In Section II the above sketched sparsity-based MD decoding technique is detailed. Section III presents experimental results and discussions about the behavior and properties of the new image MDC technique in comparison with others. Section IV concludes the paper.

2 Sparsity-based MD Decoding

As shown in Fig. 2, the proposed PDSM-based image MDC system has a very simple, fast and standard-compliant encoder. By design the MDC system shifts the burden of achieving high coding performance to the decoder, very much like in distributed source coding. At the decoder side, the MD reconstruction of the original image signal $\mathbf{f}$ starts with separate hard decoding(s) of $k \geq 1$ received side description(s). This hard decoding process prepares a set of kWH local random measurements of $\mathbf{f}$ and interfaces with a soft MD decoder that restores $\mathbf{f}$ from these local random measurements. The soft MD decoding is carried out in the well known framework of compressive image recovery; the recovery algorithm works the same way independent of the number of available local random measurements. At this point, a unique advantage of the propose image MDC system is noteworthy: the unification of the center decoder and all side decoders. In conventional MDC systems of K descriptions, in addition to the center decoder there can be as many as $2^K - 1$ side decoders, one for each possible subset of the K descriptions. In contrast, the new image MDC system uses only one MD decoder that suits all possible channel deliveries of the K descriptions. Let $\mathbf{y}$ be a decompressed downsampled image at the decoder,

$$\mathbf{y} = \mathbf{\Phi}\mathbf{f} + \boldsymbol{\upsilon}, \tag{1}$$

where $\mathbf{\Phi} \in \Re^{\frac{1}{4}WH \times WH}$ is the local random measurement matrix with downsampling integrated in, $\boldsymbol{\upsilon}$ is the compression noise generated by the hard decoder.

2.1 Sparsity Model and Dictionary Learning

The reconstruction of $\mathbf{f}$ from the set of received and hard decoded side descriptions $\mathbf{y}$ is an ill-posed inverse problem. The performance of the soft MD decoding algorithm primarily depends on how well it can explore and use priors in regulating the solution to the inverse problem. One popular technique to incorporating the prior knowledge about images is via a so-called sparsity model, in which an image is approximated by a sparse linear combination of elements in an appropriately chosen dictionary $\mathbf{\Psi}$: $\mathbf{f} = \mathbf{\Psi}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is the approximation error. In order to jointly exploit the priors of the random convolution kernels $\mathbf{\Phi}$ that produces $\mathbf{y}$ and dictionary $\mathbf{\Psi}$ to sparsely express the input image $\mathbf{f}$, the soft MD decoding problem is formulated as

$$\arg\min_{\{\boldsymbol{\alpha},\mathbf{\Psi}\}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\alpha}\|^2 + \lambda \|\boldsymbol{\alpha}\|_1 \tag{2}$$

where λ is a Lagrangian multiplier. The above formulation leads to a nonlinear optimization problem. It can be made linear and solved by alternatingly fixing one of $\boldsymbol{\alpha}$ and $\mathbf{\Psi}$ and optimizing over the other.

In initialization we need to build a dictionary $\mathbf{\Psi}$ for sparse representations of pixel patches in image $\mathbf{f}$. To this end, we obtain the first estimate $\widehat{\mathbf{f}}_0$ of the input image $\mathbf{f}$ by solving the standard compressive sensing recovery problem:

$$\widehat{\mathbf{f}}_0 = \arg\min_{\mathbf{f}} \left\|{\mathbf{\Psi}_0}^T\mathbf{f}\right\|_1 + \lambda \|\mathbf{y} - \mathbf{\Phi}\mathbf{f}\|_2^2, \tag{3}$$

where $\mathbf{\Psi}_0$ is a sparsity space (e.g., TV, DCT or wavelet space). Using $\widehat{\mathbf{f}}_0$ we build a set of M dictionaries $\{\mathbf{\Psi}_m\}_{1\leq m\leq M}$ in a learning process, each $\mathbf{\Psi}_m$ for a pattern class of $d\times d$ pixel patches. Locally adaptive dictionaries are needed because natural images typically exhibit non-stationary statistics, consisting of many heterogeneous regions of significantly different geometric structures or statistical characteristics. To learn dictionaries $\{\mathbf{\Psi}_m\}_{1\leq m\leq M}$, we classify $d\times d$ pixel patches of $\widehat{\mathbf{f}}_0$ into M groups S_m, $1\leq m\leq M$, by clustering. Considering each group S_m as a set of samples generated by $\mathbf{\Psi}_m$, we perform the principle component analysis (PCA) on S_m and let the resulting PCA bases be the words of dictionary $\mathbf{\Psi}_m$.

2.2 *Refinement by Collaborative Sparse Coding*

In the proposed soft MD decoding method, similar pixel patches in each cluster S_m are not only used to learn an adaptive dictionary, but also make the sparse decomposition more robust. Common sparse coding methods for compressive sensing recovery are patch based, and each patch is estimated independently [16][17]; they may make very different estimates for similar patches due to numerical instability of sparse decompositions. In light of this one can improve the performance of current sparse coding techniques by forcing similar pixel patches to have close estimates. Let $\mathbf{b}_\imath^f$ be a pixel patch in $\widehat{\mathbf{f}}_0$ centered at pixel location i, and S_i be the cluster such that $\mathbf{b}_\imath^f\in S_i$, we jointly optimize the sparse codes of all member patches in S_i to obtain more accurate estimates:

$$\arg\min_{\{\boldsymbol{\alpha}_{ii}\}}\left\{\sum_{\mathbf{b}_\imath^f\in S_i}\|\mathbf{b}_\imath^y-\mathbf{\Phi}_\imath\mathbf{\Psi}_i\boldsymbol{\alpha}_{ii}\|^2+\lambda\sum_{\mathbf{b}_\imath^f\in S_i}\|\boldsymbol{\alpha}_{ii}\|_1+\gamma\sum_{\mathbf{b}_\imath^f,\mathbf{b}_\jmath^f\in S_i}\|\boldsymbol{\alpha}_{ii}-\boldsymbol{\alpha}_{i\jmath}\|^2\mathbf{W}_{\imath\jmath}\right\}\quad(4)$$

where $\mathbf{b}_\imath^y$ is the down-sampled pixel patch corresponding to $\mathbf{b}_\imath^f$ in the hard decoded image $\mathbf{y}$; $\mathbf{\Phi}_\imath$ is the reorganized measurement matrix from $\mathbf{\Phi}$, which contains the local random convolution kernels of pixels in $\mathbf{b}_\imath^y$; $\mathbf{W}_{\imath\jmath}$ measures the similarity between a pair of patches $(\mathbf{b}_\imath^f,\mathbf{b}_\jmath^f)$, which is defined as:

$$\mathbf{W}_{\imath\jmath}=\exp\left\{-\frac{\left\|\mathbf{b}_\imath^f-\mathbf{b}_\jmath^f\right\|^2}{\sigma^2}\right\},\sigma>0.\quad(5)$$

The additional regularization is in the graph-Laplacian form [15], which stipulates that similar patches in S_i have similar sparse representation coefficients. Incorporating the similarity preserving term into the objective function of collaborative sparse coding can improve the robustness and accuracy of the solution of Eq. 4.

The above minimization problem can be solved by the feature-sign search algorithm. The detail of optimization can be found in [15]. Upon solving (4) and obtaining the optimal sparse coding vectors $\mathbf{A}_\imath^*=\{\boldsymbol{\alpha}_{ii}^*\}$, we restore the current patch $\mathbf{b}_\imath^f$ to be the weighted average of all reconstructed patches in cluster S_i:

$$\widehat{\mathbf{b}}_\imath^f=\frac{1}{|S_i|}\sum_{\mathbf{b}_\imath^f\in S_i}\mathbf{W}_{ii}\mathbf{\Psi}_i\boldsymbol{\alpha}_{ii}^*=\mathbf{\Psi}_i\mathbf{A}_\imath^*\mathbf{W}_\imath,\quad(6)$$

where $|S_i|$ is the cardinality of S_i, $\mathbf{W}_i = \left[\mathbf{W}_{i1}, \cdots, \mathbf{W}_{i|S_i|}\right]^T$ is the weight vector.

In practical implementation, the image patches are overlapped sampled to better suppress artifacts: $\mathbf{b}_i^f = \mathbf{R}_i\widehat{\mathbf{f}}_0$, where $\mathbf{R}_i$ is the matrix extracting patch $\mathbf{b}_i^f$ from $\widehat{\mathbf{f}}_0$ at location i. The whole image $\mathbf{f}$ can be reconstructed by averaging all the reconstructed patches [16]:

$$\widehat{\mathbf{f}} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{R}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{\Psi}_i \mathbf{A}_i^* \mathbf{W}_i\right). \tag{7}$$

where N is the number of all sampled patches. The above procedure of updating $\mathbf{\Psi}$ and $\widehat{\mathbf{f}}$ can be iterated to achieve a successive refinement.

3 Experimental Results and Analysis

In this section, experimental results are presented to demonstrate the performance of the proposed image MDC technique, in comparison with other two standard-compliant image MDC techniques: spatial multiplexing MD (SMMD) using low-pass prefiltering [13], and polyphase downsampling transform multiplexing (PDTM) [14]. These image MDC methods are tested for two balanced descriptions, each of which is is coded by JPEG 2000 and transmitted over a lossy network independently of the other. The description erasure probability p is the same for the two descriptions.

Fig. 3 presents the fidelity-rate (PSNR vs. central rate) curves of side and central decoders for the three tested methods. The results on four commonly used test images: *Bike*, *Lena*, *Monarch* and *Leaves* are reported. In all test cases, both the side and central decoders of the new image MDC method outperform those of the other two competing methods. For the side decoder, the new method outperforms the SMMD and PDTM methods by up to 1dB and 1.5dB, respectively; for the central decoder, the improvements over the two competing methods are up to 1.2dB and 2dB, respectively.

When motivating this research in the introduction, we set the goal of preserving high-frequency image features, because this is crucial to perceptual image quality of MDC. Indeed, as we expected, the advantage of broad-band prefiltering over low-pass prefiltering in recovering sharp edges and fine details is quite significant. In Fig. 4 and Fig. 5 the three image MDC methods are compared in perceptual image quality; the proposed method appears to have superior visual quality to its competitors in both side and central descriptions. When one of the two descriptions is lost, the proposed side decoder still preserves edges well and reconstructs the original image with good visual quality, whereas the PDTM method produces objectionable visual artifacts (e.g., jaggies, ringings and aliasing) in areas of edges and textures. The SMMD method ranks in the middle in visual quality, and it also produces some ghosting and aliasing artifacts along edges. In contrast, the images produced by the new method are much cleaner, with the structures and sharpness of edges and textures are well preserved. Common compression artifacts shown in SMMD and PDTM, such as jaggies along edges, are greatly reduced by the proposed technique. The advantage of local random sampling in the recovery of high frequency features is convincingly demonstrated.

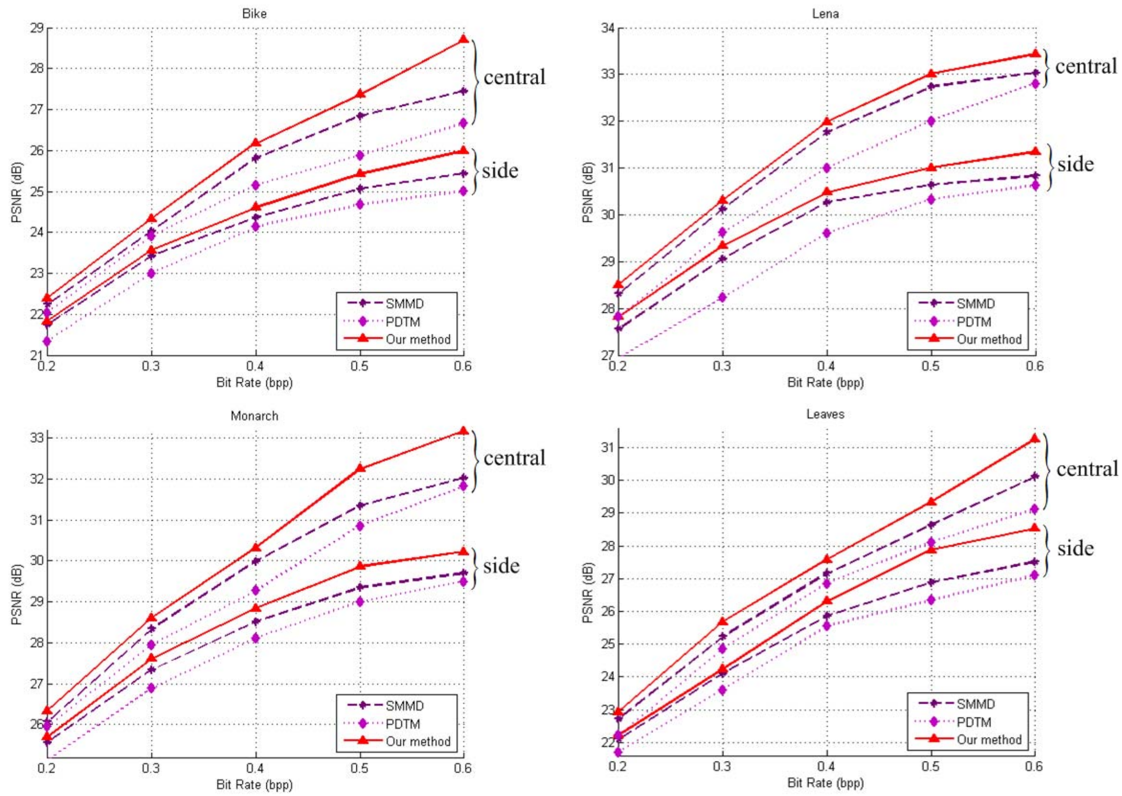

Figure 3: Comparison of three image MDC methods in PSNR values of side and central decoders versus the central rate.

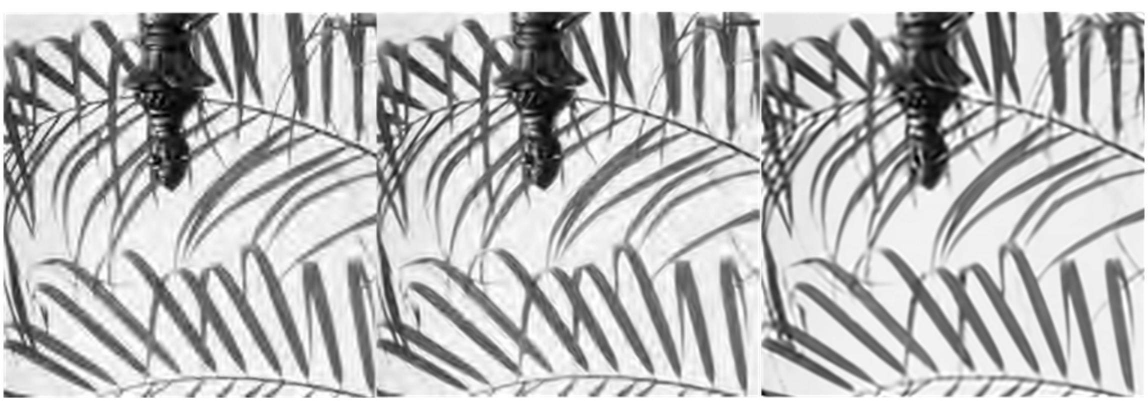

Figure 4: Comparison of decoded *Leaves* images when only one description is received at rate 0.2bpp. Left: PDTM; Middle: SMMD; Right: the proposed method.

Also, we compare the three standard-compliant image MDC methods in terms of average distortion versus the description erasure probability p:

$$\bar{D}(p) = (1-p)^2 D_c + 2p(1-p)D_s \tag{8}$$

where D_c and D_s are the central and side distortions, respectively. In Fig. 6, the average distortion $\bar{D}(p)$ is converted to average PSNR and plotted for different total rate. Three

Figure 5: Comparison of decoded *Monarch* images when both descriptions are received at total rate 0.6bpp. Left: PDTM; Middle: SMMD; Right: the proposed method.

families of PSNR curves are presented for the three different image MDC methods; each family of curves are parameterized by p. As shown by the figure, the proposed method enjoys fairly large gains in average PSNR over the other two methods. The margin of improvement in average PSNR performance increases as the channel loss probability p decreases.

4 Conclusions

We developed a standard-compliant multiple description image coding technique of low encoder complexity and a unified central and side decoder. In order to recover high-frequency image structures even at low bit rates, the encoder generates side descriptions by polyphase downsampling after broad-band prefiltering of random convolution kernel. The MD decoder is an image restoration process in a framework of compressive sensing recovery. Experimental results demonstrate that the new image MDC method achieves competitive rate-distortion performance.

5 Acknowledgement

This work was supported by National Science Foundation of China under Grants 61300110, 61272386 and 61331014.

6 References

[1] M. Alasti, K. Sayrafian-Pour, A. Ephremides, and N. Farvardin, “Multiple description coding in networks with congestion problem,” IEEE Trans. Inf. Theory, vol. 47, no. 3, pp. 891-902, Mar. 2001.

[2] V.K. Goyal, “Multiple description coding: Compressionmeets the network,” IEEE Signal Process.Mag., vol. 18, no. 5, pp. 74-93, Sep. 2001.

[3] T. Sikora, “Trends and perspectives in image and video coding,” Proc. IEEE, vol. 93, no. 1, pp. 6-17, Jan. 2005.

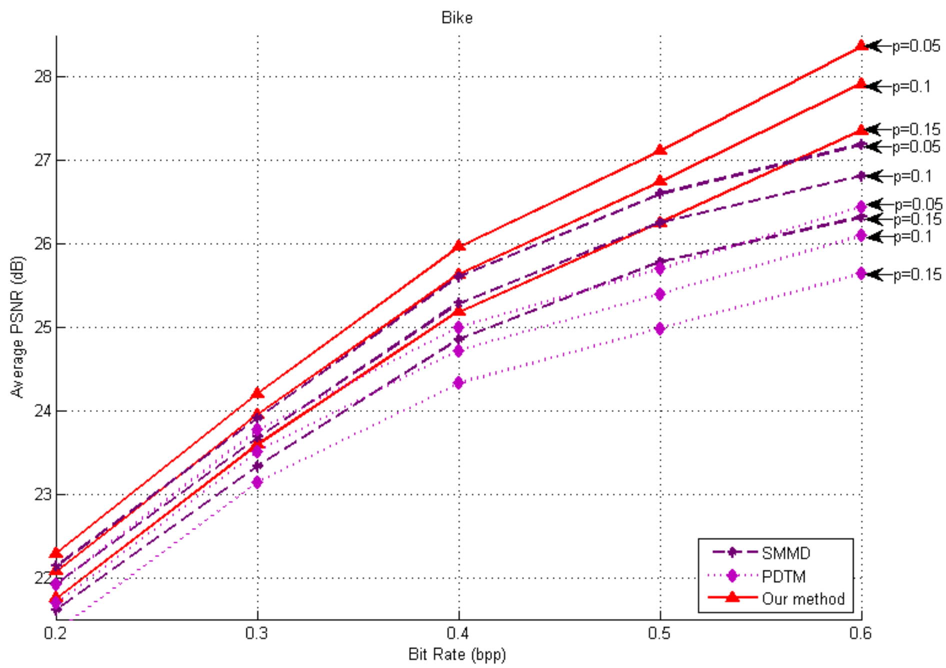

Figure 6: Comparison of three tested methods in average PSNR (dB) versus the central rate (bpp), when p = 0.05, 0.1, 0.15.

[4] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Robust layered multiple description coding of scalable media data for multicast," IEEE Signal Process. Lett., vol. 12, no. 2, pp. 154-157, Feb. 2005.

[5] V. Vaishampayan, "Design of multiple description scalar quantizers," IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 821-834, May 1993.

[6] J. Hafarkhani and V. Tarokh, "Multiple description trellis-coded quantization," IEEE Trans. Commun., vol. 47, no. 6, pp. 799-803, Jun. 1999.

[7] V. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple description vector quantization with lattice codebooks: Design and anaysis," IEEE Trans. Inf. Theory, vol. 47, no. 5, pp. 1718-1734, Jul. 2001.

[8] Y. Wang, M. T. Orchard, V. A. Vaishampayan,et al., "Multiple description coding using pairwise correlating transforms, IEEE Trans. On image Processing, vol. 10, no. 3, pp. 351-366, 2001.

[9] J. C. Batllo and V. A. Vaishampayan, "Asymptotic performance of multiple description transform codes," IEEE Trans. Inform. Theory, vol. 43, no. 2, pp. 703-707, Mar. 1997

[10] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian vectors," in Proc. DCC, Apr. 1998, pp. 388-397.

[11] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," IEEE Trans. Inform. Theory, vol. 47, no. 6, pp. 2199-2224, Sep. 2001.

[12] S. Dumitrescu, X. Wu, and Z. Wang, "Globally optimal uneven error protected packetization of scalable code streams, IEEE Trans. on Multimedia, vol. 6, no. 2, pp. 230-239, Apr. 2004.

[13] X. Zhang, X. Wu, “Standard-compliant multiple description image coding by spatial multiplexing and constrained least-squares restoration,” MMSP 2008: 349-354

[14] T. Tillo and G. Olmo, “Data-dependent pre- and postprocessing multiple description coding of images,” IEEE Trans. Image Process., vol. 16, no. 5, pp. 1269-1280, May 2007.

[15] S. Gao, I. Tsang and L. Chia, “Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 35:(1):92-104, Jan 2013.

[16] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Trans. Image Process., 15(12):3736-3745, Dec. 2006.

[17] W. Yin, S. Osher, D. Goldfarb, J. Darbon, “Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing,” SIAM J. Imaging Sciences 1(1): 143-168 (2008)