

Implementation of Dictionary Learning-based Image Compression using Local Variance

Sumangala Gangadharamath Shekhariah¹, Sreedharamurthy Sameraya Kanthiah²

¹Department of Electronics & Communication, Faculty of Engineering, Government Engineering College Haveri, India

²Department of Electronics & Communication, Faculty of Engineering, UBDT College of Engineering, Davanagere, India

DOI: <https://doie.org/10.1213/Jbse.2024442109>

Article History: Received : 12 Jul 2024

Revised: 18 Nov 2024

Accepted: 10 Dec 2024

ABSTRACT: Compared to traditional transformed-based strategies, compression of pictures utilizing learned dictionaries employing the K-SVD algorithm has demonstrated greater results. However, applying K-SVD to image patches without taking into account the local characteristics of the image can cause rate-distortion performance to suffer. Image features vary within an image and such variations can be effectively used in adaptive image processing techniques. An important image characteristic such as the local variance of image intensity is one such feature, which can be utilized in adaptive image coding technique. In this work, we propose dictionary learning-based adaptive coding based on the local variance of the image region to improve the R-D performance of the compression scheme. The technique is applied to typical test images. PSNR and SSIM are used to assess the performance of the suggested approach. According to the simulation results, the suggested technique shows improved performance at specified bit rates.

Keywords: Image, Compression, Dictionary, Learning, Sparse representation

1. INTRODUCTION

In digital communication systems, compression of image files is of great concern due to storage capacities and transmission rate constraints. The ever-increasing demand for digital images and videos seeks stronger and robust compression techniques. As a result, compressing the image by retaining only the basic information required to reconstruct it is beneficial in many applications. The transform-based compression techniques such as JPEG[1] and JPEG2000 [2] have been dominant for many decades. The core idea was to transform an image into a few transform coefficients. Then the coefficients were encoded to generate the compressed bit stream. But the non- adaptability of these fixed basis techniques to a wide range of signal phenomena results in limited expressiveness which lead to a conceptual change of sparse representation of signals using overcomplete dictionaries [3], [4]. The scheme behind this method is to interpret a signal as a sum of few vectors multiplied by some weights and code a huge quantity of visual information with only a few samples. Here the vectors belong to a matrix of basis functions known as a dictionary. Initially the technique was applied to compression of facial images [5], [6]. Skretting et al.[7] investigated an application-specific learned dictionary and applied it to picture compression to remove blocking artefacts. Authors in [8] have designed an input image adaptive dictionary which is specifically trained for the given input image. The work in[9] discusses the feasibility of using sparse representation methods in compressing fingerprints. The compression performance of fingerprints using these dictionaries reveals that a proper training method and method of constructing the dictionary will significantly contribute to better performance. The natural images are composed of smooth regions, edges, and diverse textures. Patel et al. [10] in their work classified the blocks of image into three different classes using equal mean normalized standard deviation computation and they were independently coded by using dictionaries, which are learned specific to each class. Pati et al.[11] constructed the trained dictionary using combined DCT and Gabor basis. The Gabor functions consider the directional features, which describe the vertical, horizontal, and directional components in the given image. Huang et al. [12] developed novel geometric dictionaries, which could efficiently encode the textural characteristics of the image. The directional features and various patterns are captured by several geometric dictionaries, designed for both smooth and heterogeneous regions of the image. Akbari et al.[13] in their work proposed a method for improving the rate-distortion performance by investigating the fact that the subjective significance of every block varies with its spatial position in the human visual system. An entropy-based selection of atoms is proposed by Hao et al. [14] to reduce the coding cost by limiting the selection of atoms. A universal dictionary is developed which could do better with both quantization and coding. In the works proposed by Akbari et al. [15] a sampling rate alteration is adopted to reduce the bit rate. At the encoder side, a decimator to eliminate the high-frequency information down samples the original image and bit allocation is enhanced only for low-frequency components. The coding is actually performed on a low-resolution copy of the original image and a suitable up-sampling is done at the decoder to recover the original image in a compressed form. Kalluri et al.[16], [17] used rate-distortion (R-D) optimal selection of sparse

coefficients. The method explores mixed sparse/DCT coding schemes on natural images. The literature reveals many studies where there were attempts to improve R-D performance by learning big-size and more complex dictionaries incorporating K-SVD[5] and OMP algorithms[18]. Examining the contribution of local features of natural images to improve the rate-distortion performance with a simple dictionary at low bit rates is less explored. This paper presents an effective sparse coding of images based on the local variance characteristics of the input image. The objectives of this paper are i) Implementation of the proposed learned dictionary-based image compression on natural images utilizing local variance of ii) Performance evaluation of the proposed scheme at various bit rates. The remaining paper is structured as follows: section 2 focuses on overview of sparse representation. Section 3 discusses the proposed methodology and algorithm. Section 4 provides the experimental results and discussion and section 5 concludes the paper.

2. SPARSE REPRESENTATION

A sparse representation is a numerical technique in which a signal is interpreted as a linear combination of a small number of items chosen from a big matrix of basis functions. Sparse representation of the input image data is obtained by a learned over-complete dictionary. The scheme behind this method is to interpret a signal as a sum of a few vectors multiplied by some weights. Here the vectors belong to a matrix of basis functions known as a dictionary. A linear combination of a basis matrix $D \in R^{d \times N}$ to express a signal $B \in R^{d \times 1}$ is defined as $B = DS$. This equivalence defines a linear system where $D = [d_1, d_2, d_3 \dots d_N]$ is a matrix (dictionary) of dimension $d \times N$ and $B \in R^{d \times 1}$ is a signal vector and it is supposed to be represented as a linear combination of some columns of D . Each column of matrix D is called an atom. It has N basis vectors (atoms). The product of dictionary elements and sparse vector $S \in R^{N \times 1}$ with k nonzero entries results in a linear combination of k atoms with different weights producing the signal. The system $B = DS$ is an under-determined system when $d < N$. The result is not unique since the dictionary has more than d non-zero atoms. Out of a plethora of options, one must be chosen that is the most minimal. The sparsity of S is indicated by l_0 -norm $\|s\|_0$, which reflects the count of nonzero values of S . The answer is found by tackling the optimization criterion $min\|s\|_0$ satisfying the condition, $DS = B$. The sparsity is achieved when $\|s\|_0 \ll N$. It is known that images possess redundant information. The adjacent pixels are highly correlated. When it comes to tasks like compression, the representation that is used should be efficient in capturing a significant portion of the information contained in the signal with a very small number of coefficients. As a result, selecting or obtaining the basis matrix is crucial. K-Singular Value Decomposition (KSVD) is a popular dictionary learning algorithm that has been shown to efficiently compress a variety of images. Sparse coding is the process of finding the representation coefficients based on the given signal and the dictionary. In this stage, the input data vector is approximated by the linear combination of a few dictionary elements and sparse representation coefficients. This operation is named atom decomposition, which is accomplished by iterative algorithms.

3. PROPOSED METHOD

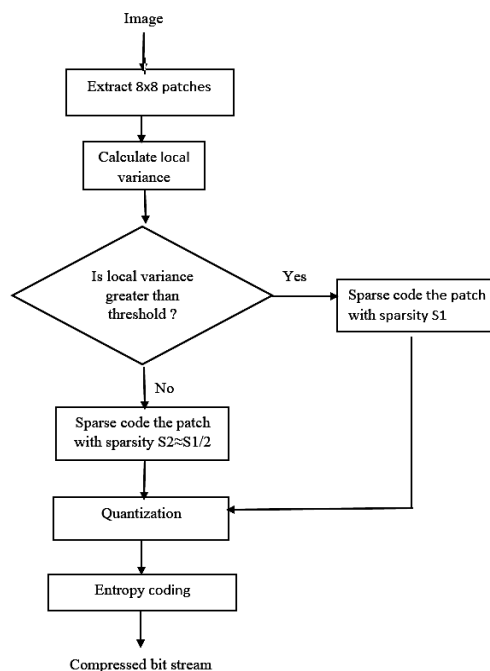


Figure 1. Flow diagram of the proposed scheme

The proposed scheme consists of computation of local variance of each non-overlapping image block of size 8x8, offline dictionary learning, sparse representation of image blocks, quantization of sparse coefficients, and entropy coding. The proposed scheme of coding is shown in Figure 1.

Algorithm of the Proposed work:

Input:
 Trained dictionary $D \in R^{d \times N}$, Input image $I \in R^{M \times N}$
 Extract image patches $b_i = [b_1, b_2, b_3 \dots b_p]$; P is the number of patches
 for $i=1:P$
do
 1. Compute local variance for each patch

$$\sigma^2 = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{m,n} - \mu)^2$$
 where, $\mu = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{m,n}$
 2. Set threshold local variance value t_i .
 3. **if** computed $\sigma^2 > t_i$, find sparse coefficients S_1 such that

$$\min_{D,B} \{\|B - DS\|^2\}$$
 subject to $\forall \|s_i\|_0 \leq S_1$, where S_1 is fixed number of nonzero entries.
 -Quantize and entropy code the sparse coefficients S_1
else if computed $\sigma^2 < t_i$, find sparse coefficients S_2 such that

$$\min_{D,B} \{\|B - DS\|^2\}$$
 subject to $\forall \|s_i\|_0 \leq S_2$, where S_2 is fixed number of nonzero entries, and

$$S_2 \approx \frac{S_1}{2}$$

 -Quantize and entropy code the sparse coefficients S_2
end for
Output: Compressed bitstream

3.1. Local variance-based image patch classification

Image features such as brightness, contrast, and texture vary from image to image and from region to region within the image. Such a variety of characteristics puts a high demand on the adaptability and robustness of algorithms for image processing, posing a major challenge. A potential solution to the challenge is to adapt image processing and analysis technologies to the characteristics of the local image. This solution, referred to as adaptive image processing and analysis, has been used successfully in various applications and has gained significant interest. In adaptive image processing and analysis, the image statistics used to test local image features play a very significant role. The local variation of image intensity is one of the statistics that effectively characterizes the local features of the image and is widely used as a means of measuring smoothness. It has been used in image processing to study spatial image structure and texture. The spread of regional variation reveals image structural information. Local variance has been utilized in many areas of adaptive image processing and analysis [19], [20], [21]. For an image patch of size MxN its local variance of intensity σ is calculated using the relation

$$\sigma^2 = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{m,n} - \mu)^2 \tag{1}$$

Where $I_{m,n}$ represents the image patch and μ represents the mean of pixel intensities of a patch and it is computed as,

$$\mu = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{m,n} \tag{2}$$

An image region or patch which has the least variation is identified as a homogeneous region whereas a region that has complex textures such as edges exhibits higher variance [22]. Thus, the structure of the image is revealed by the distribution of local variance and can be used as an effective tool in the context of adaptive image coding based on a learned dictionary.

3.2. Dictionary learning and sparse coding

The required dictionary is learned in offline mode. For this, nonoverlapping patches are extracted from the set of training images. Column vectors are created from the patches or image blocks and used in an overcomplete dictionary learning. The dictionary is trained using the K-SVD[5] algorithm. The trained dictionary is used in both the encoder and decoder.

Sparse coding is finding the representation coefficients based on the given signal and the dictionary. In this stage, the input data vector is approximated by the linear combination of a few dictionary elements and sparse representation coefficients. This operation is named atom decomposition, which is accomplished by iterative algorithms. Before sparsely coding the image blocks, the local variance of each block is computed. The patches are identified as homogeneous and complex patches by comparing their local variance values against a preset threshold. Accordingly, the image patches are sparsely coded using the OMP algorithm with a prefixed sparsity level. The sparsity level or number of non-zero spare coefficients for each

patch is considered as a stopping criterion for OMP algorithm iterations, which is in turn fixed based on the patch's local variance value.

3.3. Quantization and entropy coding

Quantization of sparse coefficients is a common step in data compression applications. The goal of quantization is to represent the sparse coefficients with a limited number of bits or levels, reducing data storage requirements. During quantization, some coefficients may become zero, which lowers the number of bits required to represent both the coefficients and their dictionary indices. This causes the bit rate to be reduced even more. Before entropy coding, uniform quantization is used to quantize the nonzero coefficients. Algorithm for quantization of sparse coefficients:

1. Input:
 - Sparse coefficients vector S , obtained from sparse coding.
 - Number of quantization levels L .
2. Determine the range:
 - Calculate the range (maximum and minimum values) of the coefficients in S . This helps determine the quantization step size.
3. Quantization step size:
 - Calculate the quantization step size $step_size$ based on the range and the number of quantization levels L :

$$step_size = (max_val - min_val) / (L - 1)$$
4. Quantization:
 - Quantize each coefficient in S individually:

$$quantized_X = round((X - min_val) / step_size) * step_size + min_val$$
5. Output:
 - The $quantized_S$ vector contains the quantized versions of the sparse coefficients.

Huffman coding with `huffmanenco` function is used for encoding step.

4. RESULTS AND DISCUSSION

Experiments are carried out in MATLAB 2015a on Intel Core i5 processor. The complete process of image coding consists of training a dictionary, obtaining sparse coefficients, quantizing and entropy coding the sparse coefficients. In this experiment, the dictionary is learned offline on eight images selected from SIPI database as shown in Figure 2 using the K-SVD algorithm. All the images used for training are grayscale 8-bit images of dimension 512x512.



Figure 2. Training image set

The dictionaries trained to obtain sparse representation are as shown in Figure 3. The overcomplete DCT dictionary, shown in Figure 3(a) of size 64x441 is used to initiate the training. Each 8x8 image patch collected from all the training set images is converted into a 64x1 learning vector. A total of 2970 such vectors are used for training out of 29696 patches extracted from training images. 100 iterations are applied to train a dictionary of size 64x441 as shown in Figure 3(b).

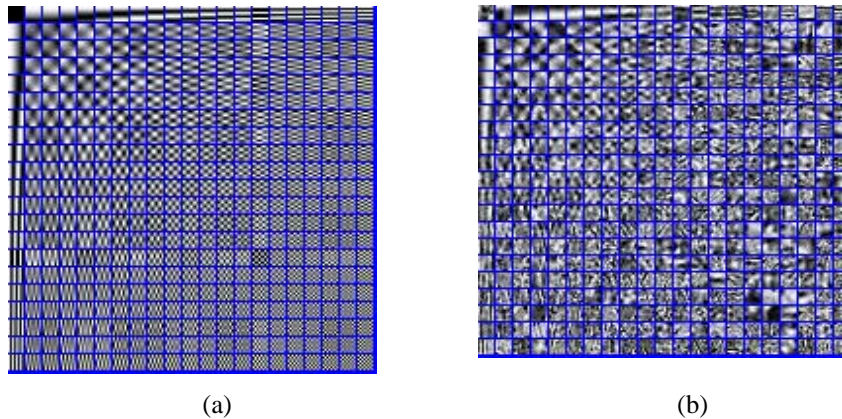


Figure 3. (a) Overcomplete 64x441 DCT dictionary (b) Overcomplete 64x441 dictionary

Standard test images like Lena, boat, Barbara, couple, and Baboon images are used for testing the proposed method. The local variance of test images is computed by splitting the whole image into blocks or patches of size 8x8. The local variance distribution patterns of test images are shown in Figure 4. The patterns reveal the texture complexities of the images patchwise. Patch local variance values are more spread for textured images as shown in Figure 4(a), Figure 4(b) and Figure 4(d). Whereas, they are closely distributed for smooth images as shown in Figure 4(c). Complex textured patches or image blocks have a higher value of local variance than smooth textured patches. For each test image, the mean value of local variances of all the patches is computed and set as a threshold. The number of non-zero sparse coefficients for patches having local variance less than the threshold is specified to be half of the number of non-zero sparse coefficients for patches having local variance greater than the threshold to code only complex textured blocks with more sparse coefficients.

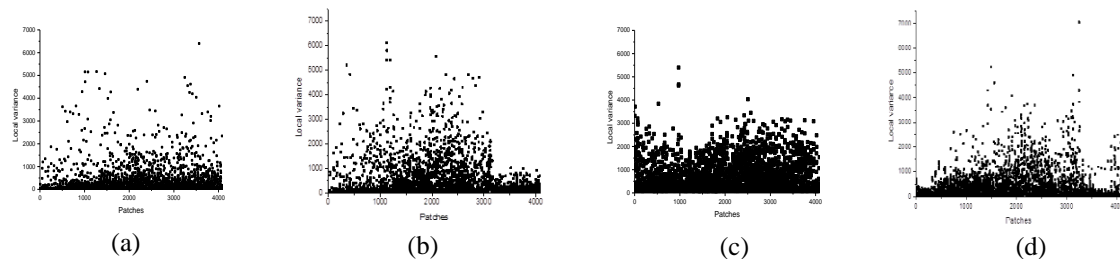


Figure 4 . Local variance distribution of (a) Lena (b) boat (c) Barabara (d) couple images

The performance evaluation of the proposed local variance based method is carried out by comparing the results with state-of-art JPEG technique and conventional method of sparse coding with K-SVD . PSNR and SSIM are used as quality metrics for performance evaluation. Figure 5 shows the subjective assessment for Baboon image at 0.2 bpp. The comparative results are as shown in Table 1. On an average, the proposed method enhances the PSNR gain by 5 dB at 0.2 bpp and 3.5 dB at 0.5 bpp compared to JPEG. Also, the proposed method yields an average PSNR gain around 0.8 dB at 0.2 bpp and 0.5 bpp. The comparison of SSIM values for test images at 0.2 bpp are presented in Table 2. It shows that the proposed method achieves better SSIM values at lower bit rates compared to the baseline JPEG and K-SVD methods.

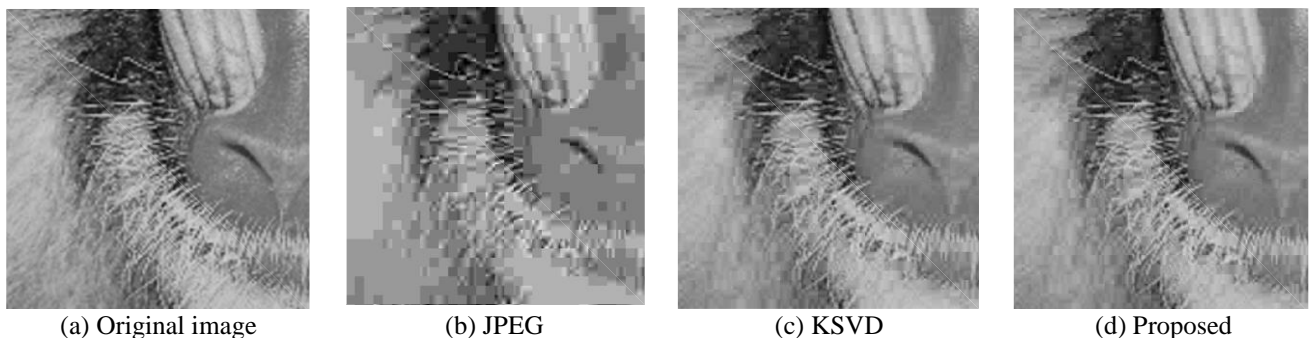


Figure 5. Subjective evaluation of test image

Table 1. Comparison of proposed method with JPEG and conventional K-SVD method

Image	PSNR (dB) at 0.2 bpp			PSNR (dB) at 0.5 bpp		
	JPEG	KSVD	Proposed method	JPEG	KSVD	Proposed method
Lena	29.95	34.39	35.41	35.87	37.08	37.30
Boat	26.25	30.73	31.57	31.23	33.48	34.52
Barbara	23.29	28.77	29.39	27.82	32.64	33.47
Couple	25.42	30.34	31.13	30.52	33.40	34.41
Baboon	22.80	28.50	28.88	28.09	33.13	33.88

Table 2. Comparison of SSIM values at 0.2 bpp

Image	JPEG	KSVD	Proposed method
Lena	0.87	0.96	0.96
Boat	0.81	0.95	0.96
Barbara	0.74	0.94	0.95
Couple	0.78	0.95	0.96
Baboon	0.75	0.89	0.90

5. CONCLUSION

This paper presents a sparse representation-based natural image compression technique based on image statistics. Specifically, in the proposed approach, each image patch can be sparsely represented by a few coefficients which are chosen as per the local variance of each patch. As demonstrated by the experimental results, the proposed approach significantly consumes less number of coefficients and exhibits improved R-D performance when compared to sparse coding of patches without considering the local variance effect. Tests conducted on various images are compared. The results show that proposed method outperforms baseline JPEG and K-SVD methods.

REFERENCES

- [1] G. K. Wallace, "The JPEG still picture compression standard", IEEE Transactions on Consumer Electronics, 1992.
- [2] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000, doi: 10.1109/30.920468.
- [3] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010, doi: 10.1109/JPROC.2010.2040551.
- [4] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010, doi: 10.1109/JPROC.2009.2037655.
- [5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for Sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006, doi: 10.1109/TSP.2006.881199.
- [6] O. Bryt and M. Elad, "Compression of facial images using the K-SVD algorithm," *J. Vis. Commun. Image Represent.*, vol. 19, no. 4, pp. 270–282, May 2008, doi: 10.1016/j.jvcir.2008.03.001.
- [7] K. Skretting and K. Engan, "Image compression using learned dictionaries by RLS-DLA and compared with K-SVD," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic: IEEE, May 2011, pp. 1517–1520. doi: 10.1109/ICASSP.2011.5946782.
- [8] I. Horev, O. Bryt, and R. Rubinstein, "Adaptive image compression using sparse dictionaries", *International Conference on Systems, Signals and Image Processing*, April 2012.
- [9] Guangqi Shao, Yanping Wu, Yong A, Xiao Liu, and Tiande Guo, "Fingerprint compression based on sparse representation," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 489–501, Feb. 2014, doi: 10.1109/TIP.2013.2287996.
- [10] J. N. Pati, J. Jose, and S. Patnaik, "Application of content specific dictionaries in still image coding," *IEICE Trans. Inf. Syst.*, vol. E98.D, no. 2, pp. 394–403, 2015, doi: 10.1587/transinf.2014EDP7135.
- [11] N. Pati, A. Pradhan, L. K. Kanoje, and T. K. Das, "An approach to image compression by using sparse approximation technique," *Procedia Comput. Sci.*, vol. 48, pp. 769–775, 2015, doi: 10.1016/j.procs.2015.04.213.
- [12] D. Huang, X. Tao, M. Xu, S. Gao, and J. Lu, "Image compression via multiple learned geometric dictionaries," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington DC, DC, USA: IEEE, Dec. 2016, pp. 1373–1377. doi: 10.1109/GlobalSIP.2016.7906066.
- [13] A. Akbari, M. Trocan, and B. Granado, "Image compression using adaptive sparse representations over trained dictionaries," in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, Montreal, QC, Canada: IEEE, Sep. 2016, pp. 1–6. doi: 10.1109/MMSP.2016.7813346.
- [14] H. Wang, Y. Xia, and Z. Wang, "Dictionary learning-based image compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing: IEEE, Sep. 2017, pp. 3235–3239. doi: 10.1109/ICIP.2017.8296880.
- [15] A. Akbari and M. Trocan, "Downsampling based image coding using dual dictionary learning and sparse representations," in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Vancouver, BC: IEEE, Aug. 2018, pp. 1–5. doi: 10.1109/MMSP.2018.8547104.
- [16] M. Jiang, M. Kalluri, N. Ling, J. Zheng, and P. Zhang, "An approach to image compression using R-D optimal OMP selection," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, Montréal, QC, Canada: IEEE, May 2016, pp. 2230–2233. doi: 10.1109/ISCAS.2016.7539026.
- [17] M. Kalluri, M. Jiang, N. Ling, J. Zheng, and P. Zhang, "Adaptive RD optimal sparse coding with quantization for image compression," *IEEE Trans. Multimed.*, vol. 21, no. 1, pp. 39–50, Jan. 2019, doi: 10.1109/TMM.2018.2847228.
- [18] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA: IEEE Comput. Soc. Press, 1993, pp. 40–44. doi: 10.1109/ACSSC.1993.342465.
- [19] S. Aja-Fernandez, R. S. J. Estepar, C. Alberola-López, and C.-F. Westin, "Image quality assessment based on local variance" in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 4815–4818, doi: 10.1109/IEMBS10037.2006.

- [20] R. Aparna and M. J. Josemartin, "Application of image intensity local variance measure for analysis of distorted images," in *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, Thiruvanthapuram, India: IEEE, Jul. 2017, pp. 382–386. doi: 10.1109/NETACT.2017.8076800.
- [21] J. Yang, G. Zhu, and Y.-Q. Shi, "Analyzing the effect of JPEG compression on local variance of image intensity," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2647–2656, Jun. 2016, doi: 10.1109/TIP.2016.2553521.
- [22] Y. Shi and Y. Ding, "On the role of local variance in image fidelity assessment," in *2010 2nd International Conference on Signal Processing Systems*, Dalian, China: IEEE, Jul. 2010, pp. V1-403-V1-406. doi: 10.1109/ICSPS.2010.5555521.