Perceptually Optimized Sparse Coding for HDR Images via Divisive Normalization

Lijuan Xie*, Xiang Zhang*, Shiqi Wang*, Shanshe Wang*, Xinfeng Zhang*, Siwei Ma*†

*Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University, Beijing, China

†Peking University Shenzhen Graduate School, Shenzhen, China

Email: {xielijuan, x_zhang, sqwang, sswang, swma}@pku.edu.cn, xfzhang@ntu.edu.sg

***Abstract*—High dynamic range (HDR) imaging techniques have been widely advocated that could shape next generation of digital photography. However, the popularity of HDR contents is hindered by the lack of displaying devices for rendering HDR images which could be very expensive. To tackle this, extensive tone-mapping operators (TMOs) have been proposed in order for transforming HDR images to viewable low dynamic range (LDR), and also applied in the backward-compatibility based HDR compression. However, how to efficiently improve the compression performance based on the perceptual evaluation is seldom addressed. In this work, we first propose a quality evaluation index for measuring the quality of the LDR image with the access of pristine HDR image. Then a sparse coding framework for efficiently compressing the LDR image, which is generated from its HDR version using TMO, is presented. Finally the compression efficiency could be improved by jointly optimize the sparse coding process in terms of the proposed quality metric based on the divisive normalization mechanism. Extensive experiments have shown that the proposed scheme can improve the perceptual quality of the compressed LDR image.**

Keywords*—*High dynamic range, sparse representation, perceptual optimization, divisive normalization.

I. Introduction

Compared with traditional low dynamic range (LDR), high dynamic range (HDR) format provides a larger range between brightest and darkest luminance captured in an image or video. With the potential for performing the real-world scene and imitating the human visual system (HVS), HDR technology attracts extensive attentions from both academe and industry in recent years. However, the specialized display devices for HDR images or videos are extremely expensive at current stage, severely hindering its development and popularity. Meanwhile, the traditional devices not supporting HDR format are daily used. To address this issue, comprehensive tone-mapping operators (TMOs) have been proposed to balance the demand of HVS and display devices. As for image coding part, existing compression algorithms can be classified into two categories: backward-compatibility and perceptual-based compression methods [1].

Tone-mapping based backward-compatible methods maintain compatibility with conventional 8-bit codecs and legacy displays which can sustain tone-mapped HDR contents. Fig. 1 illustrates the general flow chart of backward-compatible HDR compression methods. At first, tone-mapping method is employed to generate the corresponding LDR content from input HDR image. Then, the content is compressed by 8-bit encoder and decoder to obtain LDR output stream accommodating the legacy displays. Simultaneously, for ultimate display, HDR-LDR residual is compressed by independent coding. Finally, reconstructed HDR content can be obtained by the residual and LDR content recovered by real-time inversing tone-mapping.

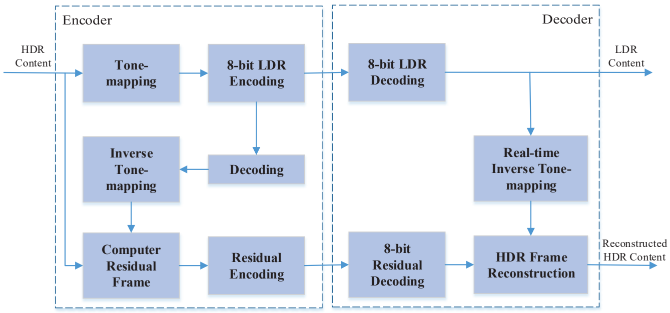


Fig. 1. The framework of backward-compatible HDR image compression.

On the other hand, perception-based methods take the masking effects of HVS into account, focusing on utilizing perception models to imitate the real-world scene. There exists many ways to achieve the perception transformation. In [2], the divisive normalization transform (DNT) is employed for optimizing the visual quality during video coding, where the DNT has been proven to be powerful in reducing higher order correlations of natural images.

It is generally believed that a successful image quality assessment (IQA) metric not only plays the role of evaluating image quality, but also react upon the procedure of image processing [3]–[5]. In our previous work [6], we proposed an objective IQA metric for tone-mapped images base on sparse representation called SMTI, which is composed of two indices: sparse-domain similarity and natural behaviors measurement. The performance on the public database shows that SMTI leads to accurate predictions on tone-mapped images.

In this work, we aim to optimize the perceptual quality during HDR image compression according to the proposed SMTI index, where the LDR image converted from its HDR version is compressed using sparse coding. The reason for using sparse coding is twofold. On one hand, sparse coding has shown its power in dealing with rich, varied and directional information contained in natural scenes [7] and extensive successive applications based on sparse model have been observed in image restoration [8], visual information estimation [9]–[11] and image compression [12]. On the other hand, the SMTI is also based on sparse coding theory, which brings convenience in optimization. The contribution of this work lies in the

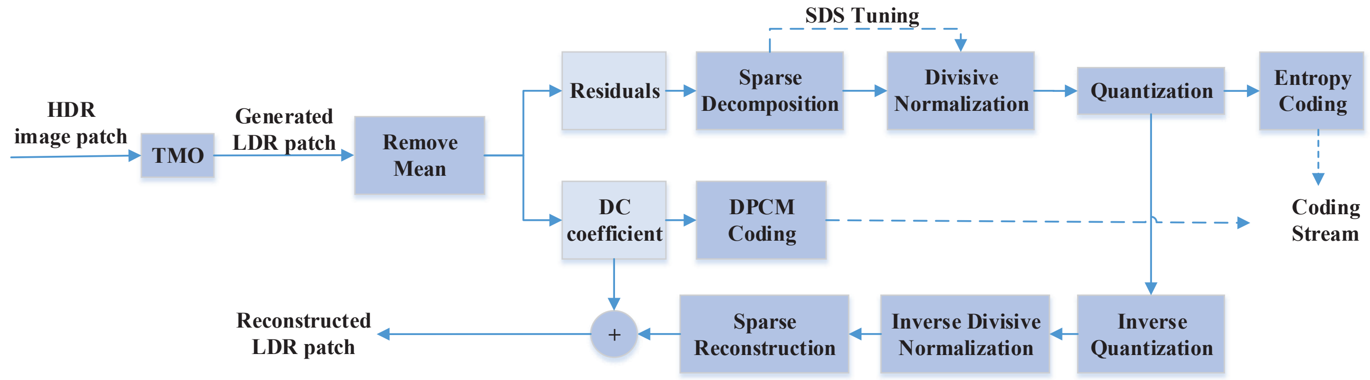


Fig. 2. The framework of the proposed HDR image coding based on sparse representation and SDS-inspired divisive normalization.

following aspects. Firstly, we build an efficient HDR image compression framework based on the sparse representation. Secondly, the DNT is utilized to optimize the SMTI index during the sparse coding procedure, yielding more pleasing perceived visual quality.

The rest of this paper is organized as follows. Section II overviews our previous work on image quality assessment for HDR images. In Section III, the proposed perceptually optimized compression framework based on divisive normalization is described. Section IV shows the experimental results of the proposed scheme in HDR image compression. Finally, Section V concludes this paper.

II. Quality Assessment for HDR Images

In this section, we briefly review our previous work on quality assessment for HDR images [6], where a local feature measurement called sparse-domain similarity is proposed for measuring the structural fidelity between an HDR image and its LDR version.

A. Sparse Representation

In sparse coding, input image is partitioned into non-overlapped patches $\boldsymbol{x}_i \in \mathbb{R}^d, i = 1, 2, \cdots, N$. Then, each image patch $\boldsymbol{x}_i$ can be precisely recovered by combining a few primitives in an over-complete dictionary $\boldsymbol{D}$ ($\boldsymbol{D} \in \mathbb{R}^{d \times k}$). Specifically, the function of sparse representation is formulated as follows,

$$(\boldsymbol{D}, \{\boldsymbol{\alpha}_i\}) = \underset{\boldsymbol{D},\{\boldsymbol{\alpha}_i\}}{\arg\min} \sum_k \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_2^2, \text{s.t.} \|\boldsymbol{\alpha}_i\|_0 < L, \quad (1)$$

where $\boldsymbol{\alpha}_i \in \mathbb{R}^k$ is the sparse coefficient vector representing the pristine image signals, the notation $\|\bullet\|_2$ represents the l_2 norm and L denotes the sparse level constraining l_0 norm of $\boldsymbol{\alpha}_i$. In this work, the typical KSVD algorithm [13] is employed to train the content-adaptive dictionary $\boldsymbol{D}$. It is worth noting that LDR image patches are extracted in advance, followed by applying the same dictionary into HDR patches.

Subsequently, an NP hard problem appears to calculate the appropriate sparse coefficients, which is solved by the orthogonal matching pursuit (OMP) algorithm [14]. For each LDR patch $\boldsymbol{x}_i^l$, OMP algorithm can figure out the suitable primitives $\boldsymbol{\Psi}_l$ and the corresponding sparse coefficients $\boldsymbol{\alpha}_i^l$,

$$\boldsymbol{\alpha}_i^l = \underset{\boldsymbol{\alpha}_i^l}{\arg\min} \sum_k \|\boldsymbol{x}_i^l - \boldsymbol{D}\boldsymbol{\alpha}_i^l\|_2^2, \text{s.t.} \|\boldsymbol{\alpha}_i^l\|_0 < L. \quad (2)$$

For the HDR patch $\boldsymbol{x}_i^h$, the sparse coefficient is defined as follows,

$$\boldsymbol{\alpha}_i^h = ((\boldsymbol{\Psi}_l^T \boldsymbol{\Psi}_l)^{-1} (\boldsymbol{\Psi}_l)^T) \boldsymbol{x}_i^h \quad (3)$$

As such, we acquire the sparse coefficients for LDR $\boldsymbol{\alpha}_l$ and HDR $\boldsymbol{\alpha}_h$, respectively. And the coefficients are projected into a unified subspace learnt from LDR ones in order to achieve better reference to the generated tone-mapped images in compression procedure.

B. Sparse-domain Similarity Index

As mentioned in SMTI model, sparse coefficients are pre-filtered by a local standard deviation function and normalized with respect to its energy in order to transform them into a uniform space. Considering the identity and stability of coefficients, we convert HDR into LDR's range as,

$$\boldsymbol{\beta}_h = \frac{\boldsymbol{\alpha}_h * f_\sigma}{\|\boldsymbol{\alpha}_h * f_\sigma\|_2} \cdot \|\boldsymbol{\alpha}_l * f_\sigma\|_2, \quad (4)$$

where $\boldsymbol{\beta}_h$ is the normalized coefficient vector of HDR patches, $*$ denotes the 2-D convolution operator and f_σ is a 3×3 std kernel function. Hence, the sparse-domain similarity (SDS) between the original HDR image $\boldsymbol{\beta}_h$ and its LDR version $\boldsymbol{\alpha}_{lr}$ can be defined as follows,

$$\begin{aligned} SDS &= \frac{\sum_{k=1}^{L} \{2\beta_h(k)\alpha_{lr}(k) + C_1\}}{\sum_{k=1}^{L} \{\beta_h(k)^2 + \alpha_{lr}(k)^2 + C_1\}} \\ &= 1 - \frac{\sum_{k=1}^{L} \{\beta_h(k) - \alpha_{lr}(k)\}^2}{\sum_{k=1}^{L} \{\beta_h(k)^2 + \alpha_{lr}(k)^2 + C_1\}}, \end{aligned} \quad (5)$$

where k indicates the sparse level and C_1 is a stabilizing constant.

III. Divisively Normalized Optimization for HDR Image Coding

In this section, we aim to optimize the HDR image coding efficiency in terms of the proposed SMTI index, where the divisive normalization transform (DNT) technique is employed for optimization.

A. Divisively Normalized Optimization

After linear sparse decomposition, divisive normalization transform (DNT) is performed as a nonlinear operation in accounting for perceptual responses in biological visual system [15], which aims to transform all the sparse coefficients into a perceptual uniform space [2].

The proposed HDR image coding framework is illustrated in Fig. 2. First, a tone-mapping operator is applied for converting HDR image into LDR. Then, the DC value is extracted from the LDR image and lossless encoded by DPCM coding. The residual is then successively transformed by sparse coding and divisive normalization for optimizing the SMTI index. Finally, normal quantization and entropy coding procedures are applied to compress the residual coefficients.

Specifically, the sparse coefficients via divisive normalization can be written as $\alpha'(k) = \alpha(k)/f$, where $\alpha'(k)$ denotes the normalized coefficients and f is the normalization factor. Considering the DNT operation, the SDS index can be rewritten as follows,

$$\begin{aligned} SDS &= 1 - \frac{\sum_{k=1}^{L}\{\beta_h'(k)f - \alpha_{lq}(k)f\}^2}{\sum_{k=1}^{L}\{\beta_h(k)^2 + \alpha_{lr}(k)^2 + C_2\}} \\ &\approx 1 - \frac{\frac{\sum_{k=1}^{L}\{\beta_h(k) - \alpha_{lr}(k)\}^2}{L}}{\mathbb{E}\left\{\sqrt{\frac{\sum_{k=1}^{L}\beta_h(k)^2 + \alpha_{lr}(k)^2}{L}} + C_2\right\}^2}, \end{aligned} \tag{6}$$

where $\alpha_{lq}(k)$ indicates the quantized LDR coefficient, $\mathbb{E}(\bullet)$ is the mathematical expectation operator and C_2 is a stabilizing constant avoiding denominator to be zero. Accordingly, the divisive normalization factor f is defined as,

$$f = \frac{\sqrt{\frac{\sum_{k=1}^{L}\beta_h(k)^2 + \alpha_{lr}(k)^2}{L} + C_2}}{\mathbb{E}\left\{\sqrt{\frac{\sum_{k=1}^{L}\beta_h(k)^2 + \alpha_{lr}(k)^2}{L}} + C_2\right\}}, \tag{7}$$

by which sparse coefficients in an image can be transformed into a perceptual uniform space for the purpose of perceptual optimization.

B. Implementation Issues

Unfortunately, problem occurs when implementing the framework. The sparse coefficients for the original HDR image $\boldsymbol{\beta}_h$ is unavailable in decoder side and encoder side can not access the reconstructed LDR coefficients $\boldsymbol{\alpha}_{lr}$. However, both of them are necessary in (7) for calculating f.

Considering that the generated LDR coefficients $\boldsymbol{\alpha}_l$ can be directly obtained in encoder side, the normalization factor is further rewritten by,

$$f' = \frac{\sqrt{\frac{corr*\sum_{k=1}^{L}\alpha_l(k)^2}{L} + C_2}}{\mathbb{E}\{\sqrt{\frac{corr*\sum_{k=1}^{L}\alpha_l(k)^2}{L}} + C_2\}}, \tag{8}$$

which is based on the hypothesis $corr * \boldsymbol{\lambda} = \boldsymbol{\gamma}$, where $\boldsymbol{\lambda} = \boldsymbol{\alpha}_l^2$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}_h^2 + \boldsymbol{\alpha}_{lr}^2)$. Then the delta quantization parameter (DQP) is transmitted in the stream,

$$DQP = QP(Q \cdot f') - QP(Q), \tag{9}$$

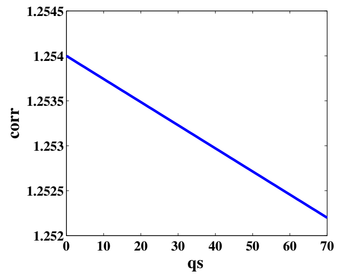


Fig. 3. The curve plot of quantization step qs and the correlation $corr$ between the generated LDR coefficients and the sum of original HDR coefficients and reconstructed LDR coefficients.

where the mapping function is $QP(q) = 6log_2(q) + 4$.

Furthermore, we exploit the relationship between the quantization step qs and $corr$, as shown in Fig. 3. The fitting curve can be formulated as a simple linear function,

$$corr = p_1 * qs + p_2, \tag{10}$$

where p_1 and p_2 are model parameters.

IV. Experimental Results

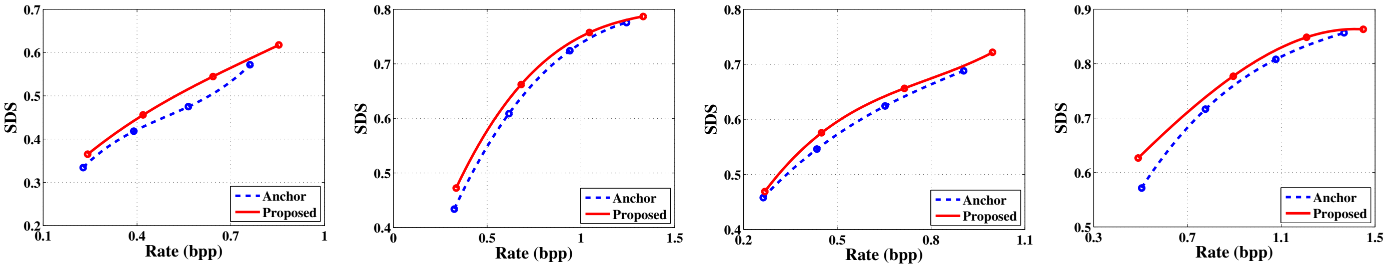
To evaluate the performance of the proposed optimization method, we compare our proposed scheme with the anchor, where the DNT optimization is not performed in the anchor. Specifically, both input HDR images and the corresponding LDR images are partitioned into 8×8 non-overlapped patches, after being transformed from RGB color space to YCbCr color space. Only Y component is compressed. Different quantization parameters are tested from 22, 28, 34 and 40.

Fig. 4 illustrated the RD performance comparison on the public dataset [16], where the red solid and blue dashed lines indicates the proposed method and anchor, respectively. It shows that our proposed optimization algorithm can achieve obvious coding gains according to the proposed SDS index.

Furthermore, the subjective comparison between the proposed method and anchor under similar bitrate is demonstrated in Fig. 5. One can observe that the proposed method can achieve more pleasing perceptual quality as well as higher SDS values, where the structure information can be largely reserved in (c) and (f) while the textures in (b) and (e) are over smoothed.

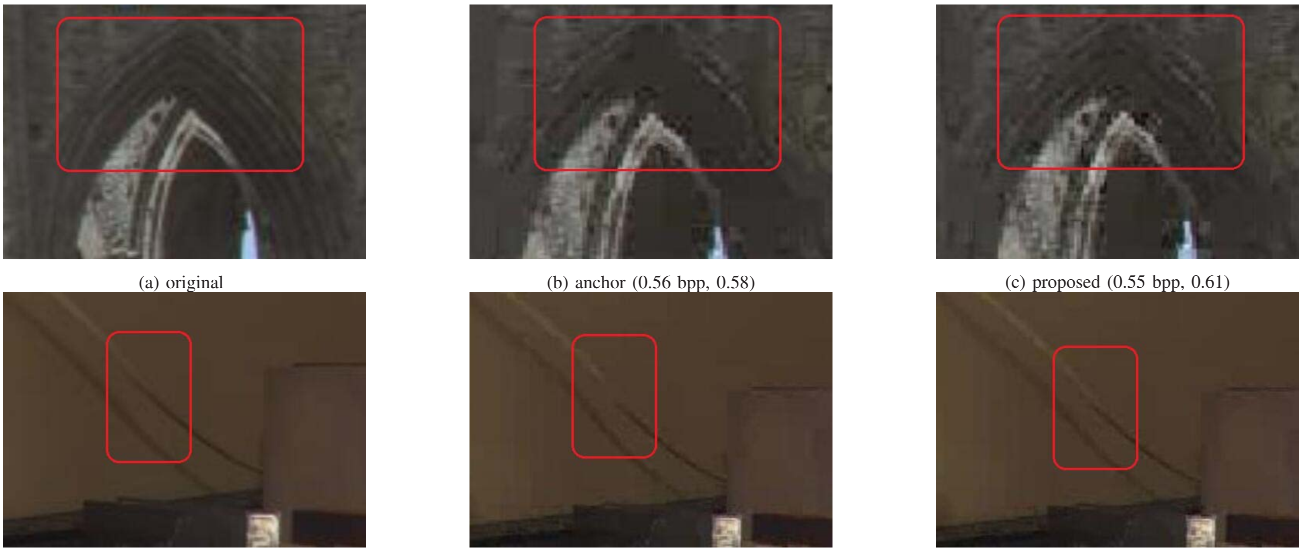
V. Conclusion

In this paper, a perceptually optimized compression framework for HDR images is proposed to optimize the perceived quality and coding efficiency during compressing the generated LDR images. The whole framework is based on the sparse coding theory and our prior proposed image quality assessment metric, i.e. SMTI. The divisive normalization transform (DNT) is employed to optimize the SMTI index during compression. In the following, objective and subjective experiments are both conducted to evaluate the performance of the proposed method, revealing that the proposed methods outperforms the anchor.



(a) (b) (c) (d)

Fig. 4. RD curve comparison between the anchor and the proposed method. The test four images are from the TMQI database [16].



(a) original (b) anchor (0.56 bpp, 0.58) (c) proposed (0.55 bpp, 0.61)

(d) original (e) anchor (0.54 bpp, 0.59) (f) proposed (0.53 bpp, 0.60)

Fig. 5. Visual quality comparison of the reconstructed images. Image (a)(d) are the tone-mapped LDR images, (b)(e) are the ones reconstructed by the anchor method and image (c)(f) are reconstructed by the proposed scheme. The values below each image denotes that rate and SDS index, respectively.

ACKNOWLEDGMENT

This work was supported in part by the National Basic Research Program of China (973 Program, 2015CB351800), and National Natural Science Foundation of China (61322106, 61421062), and Shenzhen Peacock Plan, which are gratefully acknowledged.

REFERENCES

[1] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. Bull, "High dynamic range video compression exploiting luminance masking," *IEEE Trans. Circuits Syst. Video Technol.*, 2015.

[2] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, April 2013.

[3] K. Gu, S. Wang, G. Zhai, and S. Ma, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, 2016.

[4] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang, "Details preservation inspired blind quality metric of tone mapping methods," *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, pp. 518–521, 2014.

[5] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3086–3097, 2015.

[6] L. Xie, X. Zhang, S. Wang, X. Zhang, and S. Ma, "Quality assessment of tone-mapped images based on sparse representation," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, 2016(to appear).

[7] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[8] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, 2014.

[9] X. Zhang, S. Wang, K. Gu, T. Jiang, S. Ma, and W. Gao, "Sparse structural similarity for objective image quality assessment," in *IEEE Int. Conf. Systems, Man, and Cybernetics*, 2015.

[10] S. Ma, X. Zhang, S. Wang, H. Sun, and W. Gao, "Entropy of primitive: From sparse representation to visual information evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, 2016.

[11] X. Zhang, S. Wang, S. Ma, and W. Gao, "Towards accurate visual information estimation with entropy of primitive," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, 2015.

[12] J. Zepeda, C. Guillemot, and E. Kijak, "Image compression using sparse representations and the iteration-tuned and aligned dictionary," *IEEE J. Sel. Topics in Signal Processing*, vol. 5, no. 5, pp. 1061–1073, 2011.

[13] 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, 2006.

[14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[15] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area mt," *Vision research*, vol. 38, no. 5, pp. 743–761, 1998.

[16] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, 2013.