IMAGE DECOLORIZATION BASED ON INFORMATION THEORY

Cosmin Ancuti [†*], *Codruta O. Ancuti* [*], *Miquel Feixas*[†], *Mateu Sbert*[†]

[†] Institute of Informatics and Applications, University of Girona, Spain
[*] ETcTI, Universitatea Politehnica Timisoara, Romania

ABSTRACT

In this paper we introduce a novel non-linear mapping technique to effectively decolorize images. Designed in a multi-scale fusion fashion, we first derive three input images represented by the color channels R,G and B. In order to transfer to the decolorized image only the relevant features of the derived inputs, we define two weight maps based on information theoretic approaches. The first weight map extracts visually salient regions based on a information maximization strategy while the second weight map filters the amount of local variation of each derived input computing local entropy per patch. Finally, to reduce the local distortions that might be introduced by the weight maps discontinuities, our decolorization strategy is designed in a multi-scale fusion. We also introduce a blind measure to accurately evaluate image decolorization methods. Our comprehensive qualitative and quantitative validation demonstrates that our method yields very competitive results.

I. INTRODUCTION

Grayscale images contain tones of neutral gray ranging from white to black. They are still demanded in many applications such as compression, visualization of medical imaging, aesthetical stylization, monochrome printing and single-channel image processing. Conventionally, the grayscale (decolorized) images are computed as the luminance channel of different color spaces (e.g. $CIEL^*a^*b^*$, *YCbCr*, *HSL/HSV*). Basically, the standard conversion employs a simple linear transformation that compresses the three color-channel information into a single channel image version. However, the linear mapping disregards initial visual cues such as chromatic information, local contrast or salient regions. In consequence, such standard grayscale conversions yield non-perceptually accurate images that suffer from loss of details, structure and content miss-identification (please observe the L^* channel shown in figure 1).

Since decolorization is a fundamental operation, various solutions [1], [2], [3], [4], [5], [6], [7] have been introduced in the literature in the last decades. The existing techniques can be roughly classified in two main classes: local [8], [3], [9] and global [1], [2], [4], [6], [10], [11], [12] mappings. Local mapping solutions are more robust to identifying and preserving the local features, but they are prone to distort the appearance of color regions characterized by low variance. On the other hand, global mapping techniques shown to be more effective to map over the entire image the same color to the same gray level (crucial step in image decolorization).

In this paper we introduce a novel decolorization strategy that filters and preservers the most significant features of the original color image based on both global and local information. Our solution builds on a fusion strategy that blends several images derived from the initial colored input. We first derive three input images represented by the color channels R (red), G (green) and B (blue). Then, we define two weight maps in order to transfer only relevant features of the derived inputs to the grayscale output image. The first weight map identifies prominent regions in each of the input derived image based on a information maximization strategy while the second weight map filters the amount of local variation of each derived input computing local entropy per patch. Finally, to reduce the local distortions that might be introduced due to the weight maps discontinuities, our decolorization operator employs a multi-scale fusion strategy, using a Laplacian pyramid decomposition of the inputs combined with the Gaussian pyramid of normalized weights. Multi-scale fusion is a well-studied topic in image processing and has been used in numerous enhancing applications [13], [14], [15], [16], [17], [18], [19], [20].

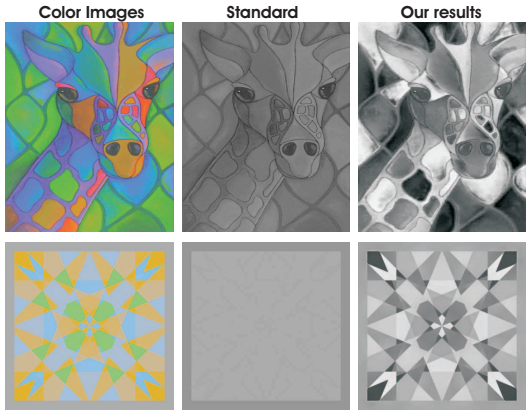

Fig. 1. *In contrast to standard decolorization technique (middle column), our method (last column) is able to preserve the global appearance of the original color image (first column).*

While multi-scale fusion has been recently used for image decolorization [12], the first contribution of this work is the original paradigm based on two measures that use information theory concepts. Information theory has been successfully applied in various image processing applications [21] ranging from enhancement [22], [23], segmentation [24], [25], saliency [26], [27], [28], shape matching [29], face recognition [30], HDR imaging [31] to video coding [32]. However, to the best of our knowledge **this is the first image decolorization concept that is built on information theory**. Moreover, in contrast to our previous fusion approach [12], that extended the HDR imaging approach of [13], here we reduce the complexity of the fusion framework by employing only three derived inputs and two (different) weight maps. We also demonstrate that the proposed version overcomes the previous fusion-based method of [12].

Additionally, as a **second contribution we introduce a new**

blind measure to evaluate image decolorization techniques. Built on the well-known SSIM [33] our measure besides contrast, luminace and structure similarity measures, penalizes the reversal of original gray tonal order). Finally, we perform an extensive qualitative and quantitative validation that demonstrates that our technique is competitive yielding comparative and even better results compared to the existing decolorization techniques.

II. THE PROPOSED NON-LINEAR DECOLORIZATION APPROACH

The standard decolorization approach performs a linear mapping ($L = w_R R + w_G G + w_B B$), where R, G, B are the color channels and w_R, w_G, w_B are the contribution weights to the final results. The weights are simple scalar parameters that represent the impact ratio of each color channel to the final grayscale result. On the other hand, for non-linear mapping techniques, the weights are vectors and as a result can be defined more appropriate for this operation. However, for both linear and non-linear mappings, in order to maintain the range consistency of the result, the weights are normalized.

Linear mapping, due to the uniform contribution of all pixels, shown to be less effective to depict accurately the original color image appearance in the decolorized output. On the other hand, if the weights are properly defined, non-linear mapping techniques shown higher robustness to preserve in the same time the local and global features. Unfortunately, for non-linear strategies the weights might introduce additional unpleasing local artifacts due to the transitions of the weights that in general do not correlate with the transitions of the input images.

In this paper we introduce an effective non-linear mapping and to overcome introducing such artifacts, we design our algorithm based on the multi-scale fusion principle that is detailed in the following subsections.

II-A. Derived Inputs and Weight Maps

While the color-to-grayscale can be seen as a compression operation (reducing the image information content from three color channel to only one luminance channel), we first decompose the color image in three inputs represented by its color channels R, G and B. Next, we derive two measures that aim to preserve the global and local information of the original image. Finally these measures (weight maps) are combined effectively in a fusion framework. The two measures are built on information theory and are depicted in the following paragraphs.

Global weight map filters the dominant/salient values in each of the input derived images. The human visual system (HSV) is highly sensitive to the global contrast that is related with the saliency of the objects in the scene [34]. In the seminal work of Attneave [35] it is shown that there is a significant redundancy in human's visual stimuli and also the salient regions of a scene are those where HVS make the greatest errors in guessing. Or, in other words, how unexpected is the content of a local region compared to its neighborhood regions. In the context of finding salient region, as demonstrated in [26], [27], extracting visually-prominent regions is similar with quantifying self-information of small patches in the image. Inspired by [26], a set of independent features is extracted from a set of local patches (31×31) randomly sampled from more than 3500 natural images. Based on these sets of local patches,

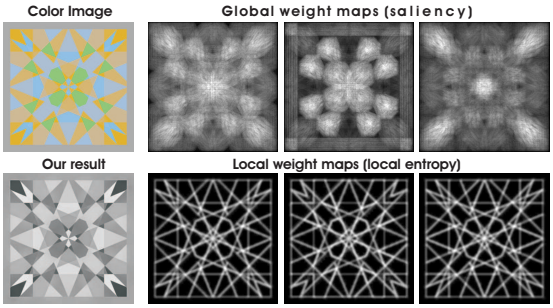

Fig. 2. *The weight maps corresponding to the three derived inputs.*

a set of basis coefficients are learned employing the Independent Component Analysis (ICA) from each local patch (the process of ICA learning selected randomly 100 patches from each image). The independent coefficients computed in every local neighborhood yield a kernel density estimate. A joint likelihood of a particular region is computed based on the product of all its individual likelihoods. Finally, to obtain our saliency (global) weight map (denoted in our fusion framework with $\mathcal{W}^1$) the computed joint likelihood $p(x)$ is converted based on the self-information measure computing $-log(p(x))$.

Local weight map (our second weight map denoted with $\mathcal{W}^2$) is computed in our framework based on the local entropy [36] that aims to filter the amount of local variation of each derived input. Based on the Shannon's theorem [37], the entropy represents the amount of uncertainty about an event associated with a given probability distribution. If we consider images as a realization of random variables, the image entropy measures the level of randomness of the image and can be defined as:

$$\mathcal{E}_I = \sum_{l=1}^{L} p_l log(1/p_l) \tag{1}$$

where p_l is the probability of pixel intensities value l to appear in the image I and L is the total gray intensities of the image. In our approach the local entropy is defined in a small patch Ω and is computed as:

$$\mathcal{E}_\Omega = \sum_{j=1}^{L} p_j log(1/p_j) \tag{2}$$

where p_j denotes the probability of pixel intensities value j to appear in the local patch Ω. To weight the local variation of each derived input we use a patch size of 5×5 and the threshold θ is computed automatically based on the maximum value of the entropy $\theta = \alpha \max(\mathcal{E}_\Omega)$ (with default parameter α=0.85).

II-B. Multi-scale Fusion

The fusion process blends per-pixel the three derived inputs (the color channels R, G, B) guided by the information of the two weight maps previously defined. Straightforwardly, the process of fusion can be expressed as: $\mathcal{F}(x,y) = \sum_k \bar{\mathcal{W}}^k(x,y)\mathcal{I}^k(x,y)$, where k is the index that counts the number of the derived inputs $\mathcal{I}^k$ (k=3 in our framework). This simple fusion operation is per pixel, and for each pixel location (x,y) the result is obtained by simply summing the corresponding locations of the inputs $\mathcal{I}^k$ balanced by the normalized weight maps $\bar{\mathcal{W}}^k$.

However, as shown in [13], this naive fusion implementation yields inconsistent results characterized by unpleasing halos artifacts close to the edges. To overcome this problem we employ a multi-scale pyramidal refinement strategy [38]. In our framework, for each derived input, a Laplacian multi-scale pyramid is computed by applying the Laplacian operator at different scales. Similarly, a Gaussian pyramid of the weight maps is built by applying Gaussian to each normalized weight map $\bar{W}$ at different scales. The blending process between the Laplacian pyramid of the inputs and Gaussian pyramid of normalized weights is performed at each level independently. Mathematically, for the pyramid scale level l the fused result is expressed as:

$$F^l(x,y) = \sum_k G^l\left\{\bar{W}^k(x,y)\right\} L^l\left\{I^k(x,y)\right\} \quad (3)$$

where $L^l\{I\}$ represents the Laplacian of the input I at scale l. Similarly, for the normalized weight map of the $\bar{W}$ we have defined the Gaussian $G^l\left\{\bar{W}\right\}$. To obtain our final decolorized image, we sum up the fussed contribution of all pyramid levels F^l.

III. RESULTS AND DISCUSSIONS

We extensively tested our new decolorization approach for various natural and synthetically generated color images. In our experiments we considered several state-of-the-art decolorization techniques including the techniques of Bala et al. [8], Rache et al. [2], Gooch et al. [1], Grundland and Dodgson [4], Smith et al. [3], Kim et al. [5], Lu et al [10] and Ancuti et al. [12]. Figure 4 presents several color images and decolorized versions generated by the analysed techniques. For the entire set of 24 images the reader is referred to the supplementary material (https://drive.google.com/file/d/1WG6O3zjKBN-KrcStDFSjZ0OUvIPrASg6/view?usp=sharing). Qualitatively, it is challenging to classify which operator works better. On a first glance, the local methods of [8], [2], [10] seem to darken some of the images. Also, the method of Smith et al. [3] introduce some artifacts close to edges due to the unsharp filter-based strategy. Moreover, compared with the recent fusion-based approach of Ancuti et al. [12] the proposed technique is able to preserve better the global contrast (see for instance the results shown in the first three rows of Fig. 4). Overall, we can conclude that our technique yields accurate results having the advantage to be less prone to artifacts compared to the local mapping techniques and yielding competitive results compared with the global mapping techniques, as well.

III-A. Quantitative Validation

Besides the qualitative evaluation we introduce a quantitative measure to validate the decolorization techniques. We built on the recent work of Ma et al. [39] that uses the well-known SSIM [33] for color-to-grayscale problem. Similar as in [39] we found that the three SSIM-derived measures are also important in the decolorization evaluation. However, a crucial characteristic of image decolorization has been ignored by the measure of Ma et al. [39]: the decolorized techniques should not reverse the original gray tonal order (e.g white has to remain white and black has to remain black also in the decolorized output). This has been observed also by the extensive perceptual user-study of Cadik [40] where users highly penalized all the examples with white background mapped

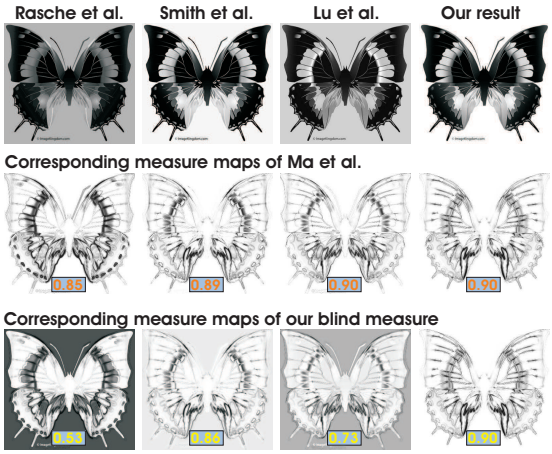

Fig. 3. *The first row shows the decolorized results of the color image with a white background shown in the 7th row of Fig. 4. The second row presents the corresponding measure maps and the quantitative values of the measure of Ma et al. [39]. The bottom row presents the corresponding measure maps and the quantitative values of our measure. Obviously, in contrast to the method of Ma et al. [39], our blind measure (BW_{SSIM}) penalizes the reversal of the original gray-tone (e.g the background should remain white also in the decolorized image).*

to a different shade of gray (e.g *butterfly* and *written white sheet* examples shown in Fig.4).

As a result, we define an additional measure map ($\mathcal{G}$) that has to penalize the reversal of original gray tonal order. This map verifies if the gray-tones (including black and white) are preserved, and vice-verse (colorful locations cannot be mapped to white or black values since although this might increase the contrast the output is not realistic). To implement this measure map we simply verify per pixel if the decolorized value is higher than the minimum and lower than the maximum of all the color channels values:

$$\mathcal{G} = \max[1 - 2 \cdot |D - \min(I_{max}, \max(I_{min}, D))|, 0] \quad (4)$$

where D is a decolorized output image, $I_{min} = \min(R,G,B)$ and $I_{max} = \max(R,G,B)$ with R,G,B are the color channels of the original color image. The image values are considered normalized in the $[0,1]$ range and the measure map returns values in the same range (as the other three SSIM-derived measures) with lower values that penalizes the result. If a pixel value of the initial color image does not contains chromatic information, it means that all it's color channels values are identical, and the measure map returns 1. On the other hand, if a pixel value contains chromatic information, the equation 4 verifies if the contrast is increased outside the local range (I_{min},I_{max}) which often implies loss of details and unrealistic mapping to black or white values (undesired).

Finally, our blind measure, denoted BW_{SSIM} (*Black-White Structure Similarity Index Measure*), is computed per pixel location (x,y) by multiplying four measure maps:

$$BW_{SSIM}(x,y) = \mathcal{L}(x,y) \cdot \mathcal{C}(x,y) \cdot \mathcal{S}(x,y) \cdot \mathcal{G}(x,y) \quad (5)$$

where the first three maps are derived from SSIM and defined as in [39] ($\mathcal{L}$ is luminance measure, $\mathcal{C}$ is the contrast measure and $\mathcal{S}$ is the structure similarity measure) while $\mathcal{G}$ represents our measure map that penalizes the reversal of gray tones defined by equation 4.

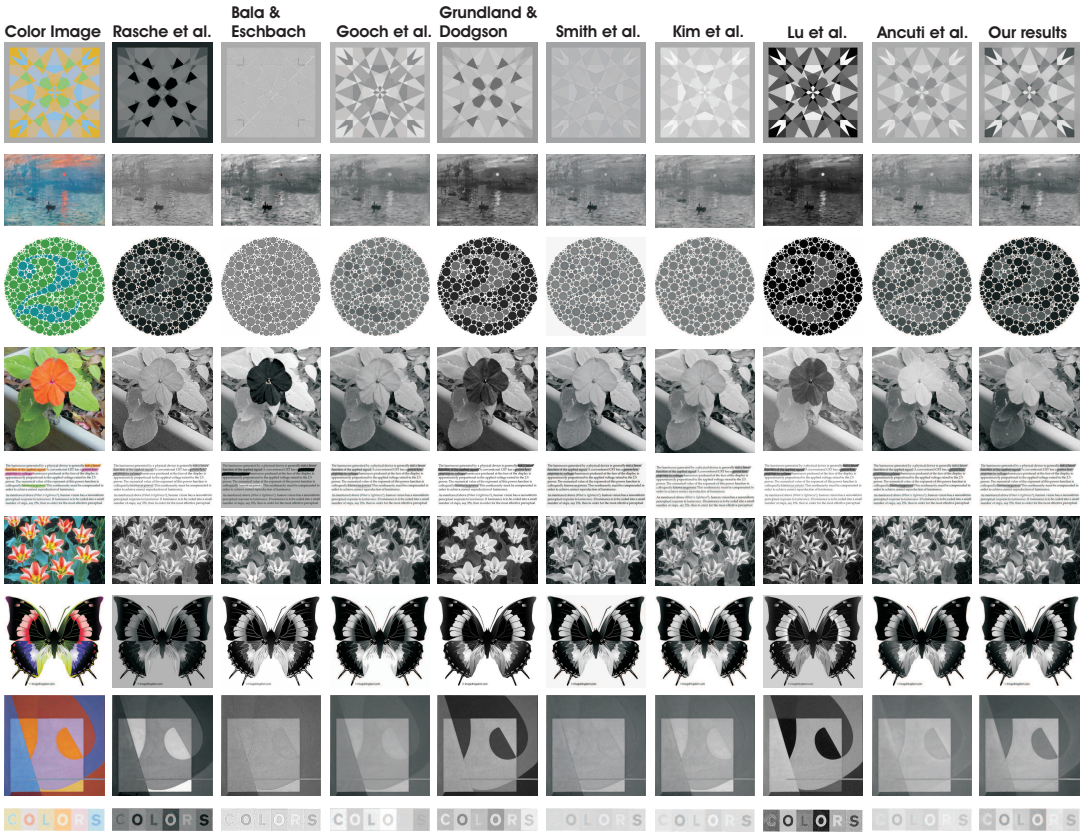

Fig. 4. *Comparison with the local and global decolorization techniques. We considered several state-of-the-art decolorzation techniques including the techniques of Bala et al. [8], Rache et al. [2], Gooch et al. [1], Grundland and Dodgson [4], Smith et al. [3], Kim et al. [5], Liu et al [10] and Ancuti et al. [12]*

Methods	Rache [2]	Bala [8]	Gooch [1]	Grundland [4]	Smith [3]	Kim [5]	Lu [10]	Ancuti [12]	Ours
BW_{SSIM}	0.8258	0.8029	0.8508	0.8483	0.8613	0.8325	0.8308	0.8496	**0.8642**

Table I. **Quantitative evaluation of decolorization methods.** We processed all 24 images of the Cadik's [40] dataset. Several images of the dateset and results of different methods [8], [2], [1], [4], [3], [5], [10], [12] are shown in Fig 4. This table presents the average values of the BW_{SSIM} measure for the considered decolorization techniques.

The importance of our blind measure is exemplified in Fig. 3. It can be seen that our measure correctly penalizes in this case (the *butterfly* image has a white background, please refer to the original color image shown on the 7th row of Fig. 4) the results yielded by the methods of Rache et al. [2] and Lu et al. [10] that are not able to map correctly the white background. This is clearly observed by comparing both the corresponding measure maps and the measure values of the Ma et al. [39] and our blind measure for this example.

Quantitatively, we employ the 24 sets of images used in the perceptual evaluation of Cadik [40]. Several images and the decolorized results of the analyzed methods [8], [2], [1], [4], [3], [5], [10], [12] are shown in Fig.4. Table I presents the quantitative evaluation based on BW_{SSIM} (the table presents the average values over the entire set of 24 images). The quantitative evaluation reveals that indeed our method is competitive and yields results with less local distortions both compared with local and global approaches.

IV. CONCLUSIONS

In this paper we introduce a novel decolorization technique designed in a multi-scale fusion fashion. To transfer only the relevant features of the derived inputs, we define two weight maps based on information theoretic approaches (a global and a local weight map). We also introduce a novel blind measure to accurately evaluate the decolorization techniques. The comprehensive validation demonstrates that our method yields very competitive results compared to the existing techniques.

Acknowledgments: *Part of this work has was supported by 2020 European Union Research and Innovation Horizon 2020 under the grant agreement Marie Sklodowska-Curie No 712949 (TECNIOspring PLUS), as well as the Agency for the Competitiveness of the Company of the Generalitat de Catalunya - ACCIO: TECSPR17-1-0054. Part of this work was supported by research grant GNaC2018 - ARUT, no. 1361-01.02.2019, financed by Politehnica University of Timisoara and also by Spanish project from Ministery of Economy and Competitivity, MINECO, TIN2016-75866-C3-3-R.*

V. REFERENCES

[1] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch, "Color2gray: salience-preserving color removal," *SIGGRAPH, ACM Trans. Graph.*, vol. 24, no. 3, pp. 634–639, 2005.

[2] K. Rasche, R. Geist, and J. Westall, "Re-coloring images for gamuts of lower dimension," *Computer Graphics Forum*, 2005.

[3] K. Smith, P.-E. Landes, J. Thollot, and K. Myszkowski, "Apparent greyscale: A simple and fast conversion to perceptually accurate images and video," *Computer Graphics Forum*, 2008.

[4] M. Grundland and N. A. Dodgson, "Decolorize: Fast, contrast enhancing, color to grayscale conversion," *Pattern Recognition*, vol. 40, no. 11, 2007.

[5] Y. Kim, C. Jang, J. Demouth, and S. Lee, "Robust color-to-gray via nonlinear global mapping," *ACM TOG*, 2009.

[6] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Enhancing by saliency-guided decolorization.," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[7] H. Du, S. He, B. Sheng, L. Ma, and R. W. H. Lau, "Saliency-guided color-to-gray conversion using region-based optimization," *IEEE Transactions on Image Processing*, 2015.

[8] R. Bala and R. Eschbach, "Spatial color-to-grayscale transform preserving chrominance edge information.," in *Color Imaging Conf.*, 2004.

[9] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "Image and video decolorization by fusion," *Asian Conference on Computer Vision*, 2010.

[10] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization with perception-based quality metrics," *International Journal of Computer Vision (IJCV)*, 2014.

[11] Wei Wang, Zhengguo Li, and Shiqian Wu, "Color contrast-preserving decolorization," *IEEE TIP*, 2018.

[12] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and M. Sbert, "Decolorization by fusion," *IEEE Access Journal*, 2018.

[13] T. Mertens, J. Kautz, and Frank Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, 2009.

[14] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22(8), pp. 3271–3282, 2013.

[15] C. O. Ancuti, C. Ancuti, T. Haber, and P. Bekaert, "Fusion-based restoration of the underwater images," in *IEEE ICIP*, 2011.

[16] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, R. Garcia, and A.C. Bovik, "Multi-scale underwater descattering," in *ICPR*, 2016.

[17] C. O. Ancuti, C. Ancuti, and C. De Vleeschouwer, "Effective local airlight estimation for image dehazing," in *IEEE ICIP*, 2018.

[18] M. Grundland, R. Vohra, G. P. Williams, and N. A. Dodgson, "Cross dissolve without cross fade : Preserving contrast, color and salience in image compositing," *Computer Graphics Forum (EUROGRAPHICS)*, 2006.

[19] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. on Image Processing*, vol. 24, no. 10, 2015.

[20] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Transactions on Image Processing*, 2018.

[21] M. Chen, M. Feixas, I. Viola, A. Bardera, H.W. Shen, and M Sbert, *Information Theory Tools for Visualization*, CAK Peters/CRC Press, 2016.

[22] Turgay Celik, "Spatial entropy-based global and local image contrast enhancement turgay celik," *IEEE Transactions on Image Processing*, 2012.

[23] Yi Niu, Xiaolin Wu, and Guangming Shi, "Image enhancement by entropy maximization and quantization resolution upconversion," *IEEE TIP*, 2016.

[24] J. Kim, J. W. Fisher, A. Yezzi, M. etin, and A. S. Willsky, "A nonparametric statistical method for image segmentation using information theory and curve evolution," *IEEE Transactions on Image Processing*, 2005.

[25] Z. Xu, B. S. Shin, and R. Klette, "Accurate and robust line segment extraction using minimum entropy with hough transform," *IEEE Transactions on Image Processing*, 2015.

[26] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," *Advances in Neural Information Processing Systems*, 2006.

[27] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, 2009.

[28] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[29] E. Hasanbelliu, L. S. Giraldo, and J. C. Principe, "Information theoretic shape matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2014.

[30] Y. Wang, Y. Y. Tang, and L. Li, "Robust face recognition via minimum error entropy-based atomic representation," *IEEE Transactions on Image Processing*, 2015.

[31] Q. Zhao, M. Sbert, M. Feixas, and Q. Xu, "Multi-exposure image fusion based on information-theoretic channel," *IEEE International Conference on Image Processing (ICIP)*, 2018.

[32] Y. Xue and Y. Wang, "A novel video coding framework using a self-adaptive dictionary," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Img. Proc.*, vol. 13, no. 4, 2004.

[34] Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, and Shi-Min Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2015.

[35] F. Attneave, "Some informational aspects of visual perception," *Psychological Review*, 1954.

[36] Wenzhan Dai and Kangtai Wang, "An image edge detection algorithm based on local entropy," *IEEE ICIT*, 2007.

[37] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, 1948.

[38] P. Burt and T. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, 1983.

[39] K. Ma, T. Zhao, K. Zeng, and Z. Wang, "Objective quality assessment for color-to-gray image conversion," *IEEE Transactions on Image Processing*, 2015.

[40] M. Cadik, "Perceptual evaluation of color-to-grayscale image conversions," *Computer Graphics Forum*, vol. 27, no. 7, 2008.