IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2018, accepted October 4, 2018, date of publication October 16, 2018,
date of current version November 19, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2876373*

# Decolorization by Fusion

**COSMIN ANCUTI [1,2], CODRUTA O. ANCUTI [2], CHRISTOPHE DE VLEESCHOUWER [3], AND MATEU SBERT [1,4]**

[1]Institute of Informatics and Applications, University of Girona, 17004 Girona, Spain
[2]MEO, Politehnica University of Timişoara, Timişoara 300006, Romania
[3]ICTEAM, Université catholique de Louvain, 1348 Ottignies-Louvain-la-Neuve, Belgium
[4]Department of Computer Science and Technology, Tianjin University, Tianjin 300072, China

Corresponding author: Cosmin Ancuti (cosmin.ancuti@gmail.com)

**ABSTRACT** Decolorization aims at converting color images into grayscale images while preserving their original contrast and color discriminability. In this paper, we introduce an original fusion-based decolorization approach. Our algorithm employs as inputs the three color channels *R*, *G*, and *B*, and an additional input related to the Helmholtz-Kohlrausch effect. To blend those inputs, we adopt a multiscale fusion strategy to prevent the artifacts arising from a pixel-wise application of weight maps and use several weight maps that respectively control saliency, exposure, and saturation. The new operator has been tested successfully on a large data set of both natural and synthetic images. It is competitive compared with modern optimization-based methods in terms of decolorized image visual quality while offering the advantage of being computationally simple, and temporally consistent when decolorizing video sequences.

**INDEX TERMS** Image decolorization, multi-scale fusion, color-to-grayscale, perception.

## I. INTRODUCTION

Nowadays, digital cameras are commonplace and the pictures/images are most often captured and represented in color. However, there are still important applications such as display to viewers with color-deficient vision, compression, visualization of medical imaging, aesthetical stylization, and *black and white* printings that require reliable grayscale/monochrome versions of the images. The standard color-to-grayscale conversion employs the luminance channel only and is still widely used for common purposes (e.g. printing, various image processing algorithms). In many cases, however, a decolorized image obtained by mapping isoluminant pixels to the same gray level intensity is not able to preserve the original appearance of the color input, as the global appearance and original spatial distribution are not well preserved (illustrated in FIGURE 1). This limitation of the standard transformation is mainly due to the fact that isoluminant locations are mapped onto the same gray intensity in the output.

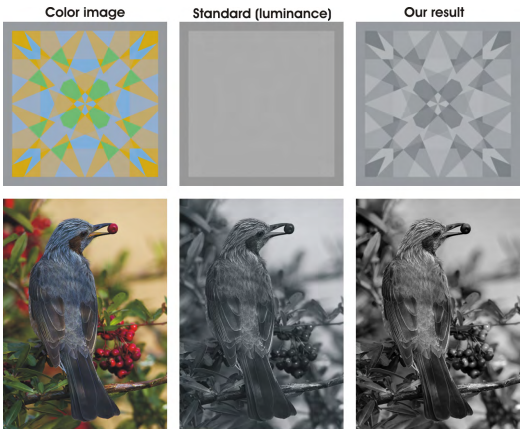


**FIGURE 1.** In contrast with a standard color-grayscale approach that would just keep the luminance as the decolorized grayscale image, our method seeks to preserve the global appearance and contrast present in the initial color image.

The operation of color-to-grayscale conversion has thus been addressed with more sophisticated approaches, based on a variety of perspectives: gamut-mapping [1], color dimensionality reduction [2]–[4], finding the color axis with predominant chromatic contrast [5], [6], decolorization targeting contrast/saliency preservation [7]–[13]. Beyond their diversity, previous methods follow either a global or a local strategy to derive the mapping function converting color pixels into gray-level values.

In this paper we present a novel decolorization method, that builds on the fusion of multiple images derived from the initial colored input. Image fusion combines several input

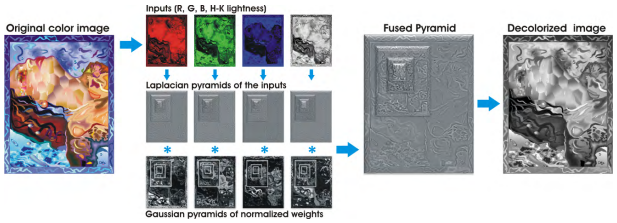


**FIGURE 2.** Overview of the proposed fusion-based decolorization approach. We derive four input images (*R*, *G*, *B* and $H-K$ lightness) from the original color image and, for each of them, three weight maps that are combined in a single normalized map. Employing a multi-scale image fusion approach, those four inputs yield the decolorized output.

images into a single output one. This is done by associating a weight map to each input so as to only transfer its relevant features to the output. Our algorithm employs four input images. The first three correspond to the three *R*, *G*, *B* channels. As a fourth input, our method derives an image that reflects the color contrast in a way that accounts for the Helmholtz-Kohlrausch (H-K) effect, i.e. the property of human perception that makes colored light appear brighter than white light of same luminance. This fourth input helps to better preserve the global appearance of the image, as it enforces distinct gray-shades for (adjacent) distinct colors. The weight maps associated to each input image are computed from three popular local contrast metrics: (a) a saliency map that preserves the most prominent pixels, i.e. the pixels that differ from the image average; (b) an exposedness weight map that advantages well-exposed regions; and (c) a chromatic weight map which favors the color saturated regions. Moreover, to reduce the local distortions that might be introduced by the weight maps discontinuities, our approach is designed in a multi-scale fashion.

In final, our method provides a straightforward technique that avoids employing color quantization or cost function optimization (which may be computationally expensive while risking not converging to a global extremum, and generally lacking of temporal consistency). In terms of image quality, our method appears to be competitive with state-of-the-art optimization-based image decolorization methods. It however offers the additional and unique (among high quality methods) advantage of providing both spatially and temporally consistent decolorization. Moreover, with single scale approximation of the multiscale fusion [cite here your single scale fusion paper], the pixel-wise processing of inputs and weight maps makes it computationally efficient, and suited to parallel implementation. Together, all those features make it especially suited to video processing.

The presented fusion-based decolorization strategy builds on our previous conference paper [14] presenting a more detailed and coherent body of work. In addition, here we demonstrate quantitatively that our technique is able to preserve important global and local features of the color image. We also perform a comprehensive perceptual validation based on the preference of real observers. Moreover, in this work we introduce for the first time a quantitative strategy to assess the temporal coherence for video decolorization.

The remainder of the paper is structured as follows. Section II surveys the related work. Section III introduces our decolorization method. Section IV validates our approach and provides a quantitative and qualitative analysis of recent decolorization operators, including for video sequences. Finally, Section V concludes.

## II. RELATED WORK

Recently, the problem of decolorization or grayscale image conversion has received an increasing amount of attention. Various local and global mapping techniques have attempted to solve this dimensionality reduction problem.

With a **global mapping** scheme, pixels with the same color are converted to the same grayscale. Global adjustment methods typically compute the mapping either (i) to maximize the grayscale image variance, or (ii) to minimize the difference between the contrasts measured in the graylevel and colored images. Among the first class of methods, the works in [15] and [16] define the global mapping to be a linear combination of RGB channels that is optimal in the sense that it maximizes the variance of the output grayscale image, while preserving the image brightness. Compared to [15] and [16] adds an additional term to the objective function so as to promote a mapping that selects the minimum norm solution among all feasible solutions. Among the methods preserving the contrast between color and grayscale images, the method of Rasche *et al.* [3] formulates the optimization based on a set of landmarks whose colors have order-constrained luminance values. The method is known to scale poorly with the number of landmark colors [17]. Kim *et al.* [18] do not consider landmarks. They optimize a nonlinear parametric global mapping function so as to minimize, over the whole image, the difference between the local color and the grayscale image gradients. The mapping function is chosen to achieve lightness fidelity, thereby favoring the preservation of brightness ordering. Lu *et al.* [7], [19] adopt a similar global energy functional minimization. They use a second order multivariate polynomial parametric model as a global mapping function, and explicitly relax the strict preservation of brightness ordering constraint while maximizing the preservation of the original color contrast. Reference [8] fundamentally follows [7], [18], and [19] in the sense that it computes an optimal global mapping that preserve the gradient between color and gray images. Overall, the comparative analysis provided in our validation section also reveals that, beyond their computational complexity, those optimized global mapping fail in rendering properly the colored scene where local chromatic contrasts have a strong perceptual impact (see FIGURE 11, sets 2/18 for Kim *et al.* [18], and sets 5/21 for Lu *et al.* [7], [19]).

Still among the global mapping techniques, the approach of Grundland and Dodgson [5] does not target grayscale image contrast maximization but employs a dimensionality reduction using predominant component analysis aiming at

identifying the axis with optimal chromatic contrast rather than the axis with maximal variance. Alternatively, to reduce the three colors to a single dimension channel, Yoo *et al.* [4] fundamentally rely on the derivation of a sequence of colors by clustering the colors along an elongated curve with locally maximal saturation in the *CIELab* color space. They then map this curve to a monotonic sequence of gray levels. Our validation reveals that those approaches might occasionally fail in capturing significant color contrasts (even for the very good and effective method in [5] see set 23 in FIGURE 11).

Differently from contrast-based optimization and dimensionality reduction, [12] simply proposes to enhance the luminance contrast based on some well-chosen chrominance information. The approach accounts for the opponency theory, which states that our vision is dominated by black/white, red/green, and yellow/blue perception processes.

At the opposite of global approaches, the **local adjustment** methods adapt the color-to-gray mapping as a function of the local distribution of colors. As for global approaches, a first category of local methods aims at maximizing the variance or preserving the contrast in the grayscale image. The approach in [20] combines the values of the red, green, and blue channel pixel by maximizing the local variance and in the same time preserving the brightness of the color image. The differences among the local transformations at nearby pixel locations are minimized based on an total variation (TV) approach. Gooch *et al.* [2] attempt to preserve the initial image contrast by adjusting the gray values in each pixel position to minimize the sum of squared difference between the grayscale and color contrast computed in each pixel. The gradient descent convergence process, initiated by the conventional luminance image, is likely to remain stuck in a local optimum in presence of highly textured images. Reference [9] follows [2] in that it minimizes an energy function reflecting the difference between grayscale and color contrast, but differs from [2] in that it derives the color contrast from the perceptually dominant channel among hue, saturation, and luminance, and in that it estimates the energy function on a subset of pixels, before propagating the result to other pixels thereby limiting the perceptual impact of spatial inconsistencies. Lau *et al.* [21] first project the color image to the luminance space, and adopt a cluster-based approach to improve the luminance contrast. Therefore, the method groups pixels into clusters according to their color and spatial similarities, solves for new cluster grayscale values in the target space with an optimization that aims at preserving original chromatic contrast, and transfers those changes of cluster values back to each pixel. When the number of clusters increases, the method reduces to [2], and suffers from the same drawbacks.

A second category of local approaches reconstructs the grayscale image through gradient field integration. Neumann *et al.* [17] compute a chromatic gradient field both from the *CIELab* and *Coloroid* [22] color spaces. They then compute the integral-consistent gradient field that is the nearest to the chromatic one, and reconstruct the gray

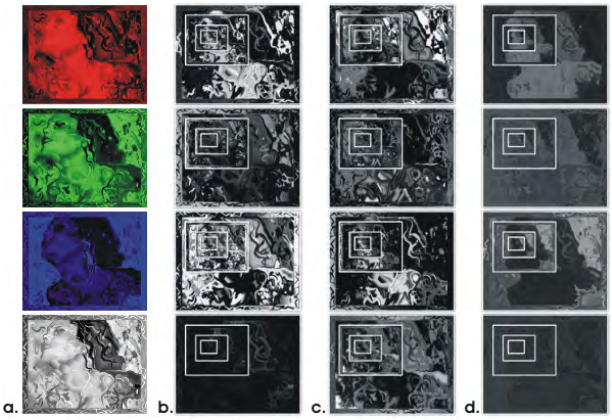


**FIGURE 3.** a. The derived four inputs (*R*, *G*, *B* and $H-K$) and the corresponding Gaussian pyramids of the weight maps (b. saliency; c. exposedness; d. chromatic).

image by simple integration. In final, the consistency constraint however appears to limit the local adaptation of the color to gray mapping. The multispectral method of Alsam and Drew [23], defines the gray image gradient based on the maximum gradient among the color channels gradient, which is a simplification of the dominant eigenvalue considered in [24], and derives the gray image through an iterative procedure. Reference [25] follows a similar approach, using a Poisson equation solver to derive the mapping from the gradient field. Finlayson *et al.* [26] also process the image in the gradient domain and then re-integrate the gradient field to reconstruct the image. To avoid the artifacts inherent to the ill-pose nature of reconstructing a non-integrable field, they propose to learn a relationship to map the gradients of the 3 colors and their 6 pairwise products onto the desired chromatic gradient, and use this relationship to map the colors and pairwise products directly to the output gray value. In final, the approach results thus in a global mapping strategy, and does not benefit from the advantages of local approaches.

The third and last category of local methods drives local contrast enhancement based on some chromatic filtering process [1], [6], [10], [11]. The strategy introduced in [1] combines the luminance channel with a high-pass filtered chroma channel. Reference [6] also merges the luminance and a chromatic contrast/gradient information, but estimate the color contrast/gradient by computing difference of Gaussians according to an image-dependent predominant chromatic orientation. Smith *et al.* [11] build on the Laplacian pyramid to adjust the grayscale conventional luminance image according to the local chromatic contrast. It relies on the lightness measure of Fairchild and Pirrotta [27] to account for the Helmholtz-Kohlrausch effect when estimating the local chromatic contrast. In that sense, it is close from our approach. Our validation section however demonstrate that our way to weight the color and H-K components generally performs better than Smith *et al.* [11].

Some more previous works address a related but distinct problem than visually consistent decolorization. For exam-

ple, [28] challenges the inverse problem of estimating the original RGB image from a black-and-white halftone image with homogeneously distributed dot patterns. Besides, de Queiroz and Braun *et al.* [29] map the color information onto imperceptible high-frequency textures of the output graylevel image, so that they can be recovered with an appropriate decoder. Their work has proven to be a practical solution for office documents, but obviously does not preserve the initial color contrast in the grayscale image. As a last example, the work in [30] considers document segmentation, by earning a filter that maximizes the output on text pixels while minimizing it on background pixels.

In contrast with most previous techniques, our decolorization algorithm does not require an optimization step before mapping colors to gray values. It employs a straightforward multi-scale fusion strategy. When computational complexity and data transfer become an issue, the multi-scale process can be effectively approximated with a single scale procedure [31]. Fusion-based decolorization has initially been considered in [32] to merge the three R, G, and B color components. In contrast, by introducing an additional input that accounts for colored lightning perception and adopting appropriate weight maps, our approach preserves visual appearance and demonstrates spatial and temporal consistency when decolorizing both images and videos. The qualitative and quantitative analysis provided in Section IV reveals that our method compares favorably to a representative set of the above methods.

Image fusion is a technique that blends data from multiple sources. It has been successfully applied in different fields [14], [24], [33]–[47], and significant efforts have been made in recent years to develop efficient and effective image fusion methods for a variety of applications from all-in-focus imagery to dehazing and HDR imaging. Even though we built on the fusion principle, our approach presents several distinctive features that allow compressing the three-dimensional color image in a grayscale version that preserves the original contrast and details, by employing inputs and image quality measures that are specifically designed for our decolorization task.

## III. FUSION-BASED DECOLORIZATION APPROACH

In this work we start from the principle that the presentation of an image in black-and-white should be tightly connected to color perception. However, notions like color saliency and color contrast are difficult to measure and integrate in a mapping of pixels color values. For this reason, we have opted for an image fusion approach, combining the Helmholtz-Kohlrausch ($H-K$) lightness predictor [27] with the three $R$, $G$, $B$ color components, based on a set of pixel weights that are determined by the product of saliency, pixel exposure, and chromatic quality metrics. An overview of our approach is given in FIGURE 2. The pixel-wise locality of our approach makes it both spatially and temporally consistent. This is in contrast with region-based local approaches, which process colors differently in case of distinct neighborhood, but also with global approaches, which derive different mapping schemes when the image (partly) changes, thereby inducing temporal inconsistencies.


**FIGURE 4.** Importance of the Helmholtz-Kohlrausch ($H-K$) input to improve the contrast of our decolorized result.

The rest of this section introduces the $H-K$ input formally, defines the weight maps, and surveys the fusion process.

### A. HELMHOLTZ-KOHLRAUSCH CHROMATIC ADAPTED LIGHTNESS

The Helmholtz-Kohlrausch ($H-K$) effect results from the human visual system behavior when facing a colored lighting. It refers to the entoptic phenomenon that makes colored light appears brighter to us than white light of the same luminance. Basically, the intense perceived saturation of spectral hue is interpreted as part of the original luminance of the color.

As observed in [11], the $H-K$ effect is useful to solve ambiguities introduced by the isoluminant colors. In practice, the most colorful patch among two isoluminant ones is mapped onto a brighter intensity. Formally, to define the associated fusion input channel, we propose to use the Fairchild's chromatic lightness metric [27], which predicts the $H-K$ effect, by defining a perceived luminance $L_{H-K}$ in the *CIE* $L^*c^*h$ color space by the expression:

$$L_{H-K} = L^* + (2.5 - 0.025L^*)(0.116\left|sin\left(\frac{h-90}{2}\right)\right| + 0.085)c^* \quad (1)$$

where $L^*$, $c^*$, $h$ respectively denote the luminance, chroma and hue color components. This $L_{H-K}$ predictor has been employed previously in the work of Smith *et al.* [11]. However, as discussed, we used this predictor in a different fashion and Section IV reveals that our way to exploit the $L_{H-K}$ predictor performs generally better than Smith *et al.* [11] (see sets 2, 8, 12 in FIGURE 11).

### B. WEIGHT MAPS DEFINITION

In the following, we define and motivate the weights used by our fusion technique. First, since image decolorization aims at maintaining the distinctiveness of local differential structures depicted by initial color images, our approach detects the visually important features based on a saliency measurement. We have considered the metric defined in [48] to estimate the saliency. The second weight map of the algorithms detects and maintains the overall image appearance into the final result by estimating the exposedness for each input. This gain map ensures the preservation of the achromatic

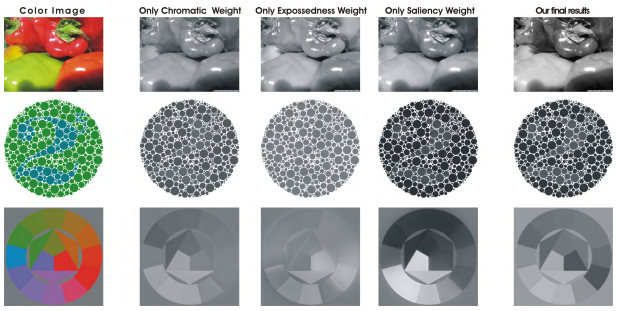


**FIGURE 5.** The weights impact (for three set of images) when considered separately in our fusion-based framework.

tones presented in the color input as demonstrated in the validation section.

Finally, the third weight map is motivated by the observation that the luminosity perceived by human subjects is affected by the color of light and also by the saturation. More saturated colors are perceived brighter. This adaptation of the visual-human system has been exploited by artists to compensate for the difference between the real world wide range luminance and the reflectance of existing pigments. Increasing color saturation can produce illusory sensations of brightness, since details visibility changes when regions are presented more saturated [11]. Based on these observations it is desirable that more saturated regions appear brighter in decolorized images.

In the rest of the section, we present a formal definition of each of those three weight maps.

***The saliency weight map*** ($\mathcal{W}_S$) refers to the the degree of prominence with respect to the linked regions. Among the numerous works dealing with saliency estimation, we have been interested in solutions that are able to extract a full resolution saliency maps with well-defined boundaries of salient objects while being also computationally effective. As a result, for this measurement, we opted for the recent saliency algorithm of Achanta *et al.* [48]. In this approach the saliency weight at pixel position $(x, y)$ of input $I^k$ is expressed on a per-pixel basis as:

$$\mathcal{W}_S(x, y) = |I^k_{\omega_{hc}}(x, y) - I^k_{\mu}| \tag{2}$$

where $I^k_{\mu}$ is the arithmetic mean pixel value of the input $I^k$ while $I^k_{\omega_{hc}}$ is the blurred version of the same input (see [48] for more details). This map prevents introducing artifacts since neighboring comparable values are assigned to similar saliency weights. The fact that the saliency map depends on the image average (a global parameter) could potentially induce a lack of temporal consistency when decolorizing videos. To circumvent this potential issue, and to preserve the temporal coherence in our decolorized sequence, instead of computing a mean value for each frame, we propose to simply compute a global mean value for the entire sequence. For long sequences, the average on a large sliding window could of course be considered.

***The exposedness weight map*** ($\mathcal{W}_E$) aims at avoiding an over- or underexposed location appearance by giving more weight to the inputs lying around their mid-range. Inspired by the approach of Mertens *et al.* [42], who employs a similar measure but in the context of multi-exposure fusion, this weight map is expressed as a Gaussian-modeled distance to the average normalized range value:

$$\mathcal{W}_E(x, y) = exp\left(-\frac{(I^k(x, y) - 0.5)^2}{2\sigma^2}\right) \tag{3}$$

where $I^k(x, y)$ is the value of the pixel location $(x, y)$ of the input image $I^k$, and the parameter $\sigma = 0.25$. This mapping conserves those tones that are characterized intermediate exposedness, neither over- or under-exposed, keeping the original appearance of the input.

***The chromatic weight map*** ($\mathcal{W}_C$) enables our conversion algorithm to adapt to the chromatic information by boosting the color of the highly saturated regions. Basically, for each pixel chromatic weight is computed as follows:

$$\mathcal{W}^k_C(x, y) = \left[I^k(x, y) + \delta.S(x, y)\right]^2 \tag{4}$$

where $k$ indexes the input, $S(x, y)$ denotes the saturation in position $(x, y)$ for the HSI color space, and $\delta$ denotes a small percentage value, set to one percent in all the results provided in this manuscript. In consequence, for each channel, the pixels with high values on that channel will be identified as higher contributors for the chromatic weight associated to that particular channel. This can be observed in FIGURE 7. Note that, because the product of the weights are normalized across the inputs (see next section), the relative values of the chromatic weights assigned to the different inputs are more meaningful/relevant than their actual absolute values. For example, white pixels assign a same chromatic weight to all color components, and the actual impact of the chromatic weight for those pixels actually vanishes after normalization.

### C. COMBINING THE WEIGHT MAPS

We have processed and analyzed a large and diverse set of images. It appeared that none of the weight map was significantly and constantly more relevant than the others. Hence, an aggregated weight map $\mathcal{W}^k$ for input k is simply computed by multiplying the processed weight maps $\mathcal{W}^k_S$, $\mathcal{W}^k_E$, $\mathcal{W}^k_C$. This ensures that the weight of a pixel will be small as soon as one of the weight map is small in that pixel. To yield consistent results that exploit the entire image dynamic range, we normalize the resulted aggregated weight maps across inputs ($\bar{\mathcal{W}}^k(x, y) = \mathcal{W}^k(x, y) / \sum_k \mathcal{W}^k(x, y)$, see FIGURE 8). This operation imposes that the sum of the normalized weight maps is equal to one in each pixel location.

### D. MULTI-SCALE FUSION OF THE INPUTS

Having defined the inputs ($R$, $G$, $B$ color channels and $H - K$ chromatic adapted lightness) and the weight maps, we now present how this information is blended through a multiscale fusion strategy. As previously mentioned, the purpose of

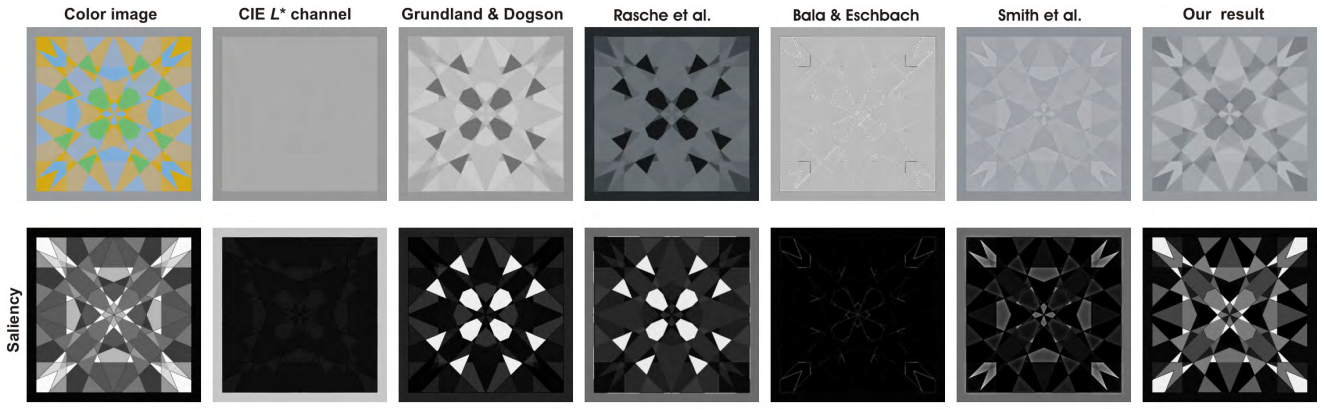


**FIGURE 6.** Saliency preservation of different decolorization strategies. This figure is based on the saliency metric defined in [48].

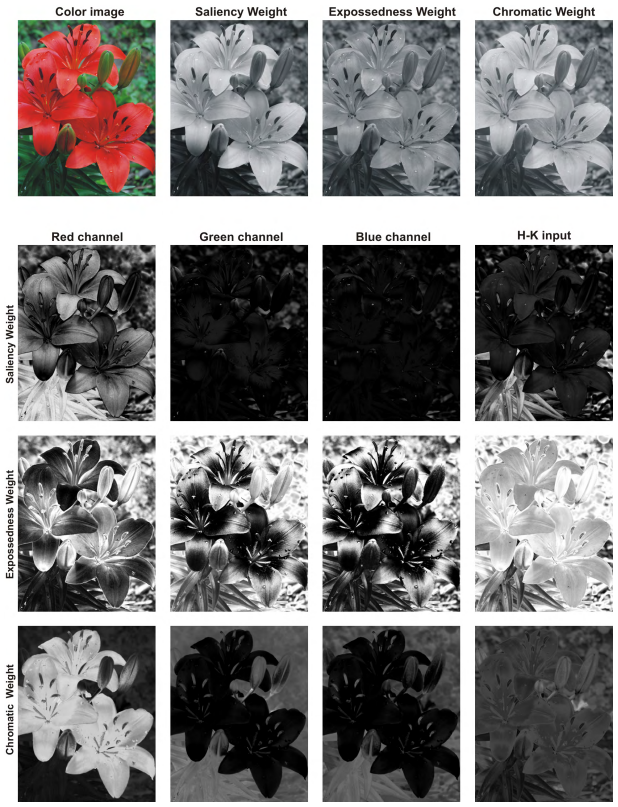


**FIGURE 7.** The impact of each weight map (top row) and the weight maps computed for each input separately.

the fusion process is to transfer each input image to the output, in proportion to its associated weight map value. A naive and straightforward implementation of this principle would consist in computing the fused image $\mathcal{F}$ every pixel location $(x, y)$ as:

$$\mathcal{F}(x,y) = \sum_{k=1}^{K} \bar{\mathcal{W}}^k(x,y)\mathcal{I}^k(x,y) \tag{5}$$

where $k \leq 4$ is the input index and $\bar{\mathcal{W}}^k$ are the normalized weight maps that influences the inputs.

However, applying Eq. 5 directly may introduce important haloing artifacts (see FIGURE 9), mainly in locations close to strong transitions between weight maps. This issue is solved by multi-scale decomposition strategies that use linear [49], [50] or non-linear filters [51]–[53]. Even though the class of non-linear filters are competitive to preserve edges, the linear filters are computationally more effective. In this work we have opted for the multi-scale Laplacian pyramid decomposition [49].

In the multi-scale Laplacian pyramid decomposition, every input image $\mathcal{I}^k$, is decomposed into a Laplacian pyramid while the normalized weight maps $\bar{\mathcal{W}}^k$ are decomposed using a Gaussian pyramid. Assuming that both the Gaussian and Laplacian pyramids have the same number of levels, the mixing of the Laplacian inputs with the Gaussian normalized weights is performed independently at each level, yielding finally the $l^th$ fused pyramid level to be:

$$\mathcal{F}^l(x,y) = \sum_{k=1}^{K} G^{l-1}\left\{\bar{\mathcal{W}}^k(x,y)\right\} L^l\left\{\mathcal{I}^k(x,y)\right\} \tag{6}$$

when $l < N$ and

$$\mathcal{F}^N(x,y) = \sum_{k=1}^{K} G^{N-1}\left\{\bar{\mathcal{W}}^k(x,y)\right\} L^N\left\{\mathcal{I}^k(x,y)\right\} + \sum_{k=1}^{K} G^N\left\{\bar{\mathcal{W}}^k(x,y)\right\} G^N\left\{\mathcal{I}^k(x,y)\right\} \tag{7}$$

when $l = N$, with $l$ denoting the pyramid level index.

The final decolorized image is obtained by summing the fused contribution of all levels, each level being upsampled at the native image resolution:

$$\mathcal{D}(x,y) = \sum_{l}^{N} \mathcal{F}^l(x,y) \uparrow^d \tag{8}$$

where $\uparrow^d$ is the upsampling operator with factor $d = 2^{l-1}$. A similar fusion process has been applied previously

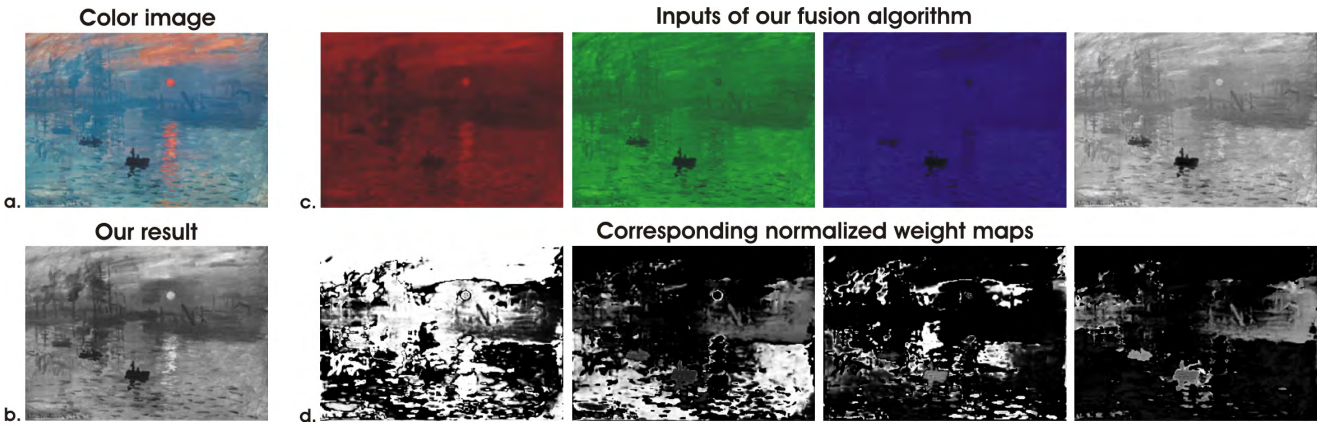


**FIGURE 8.** From color to gray: from the original color image (a), we obtain our decolorized result (b) by applying an image fusion approach, using the four inputs (c), weighted by the corresponding normalized weight maps (d).

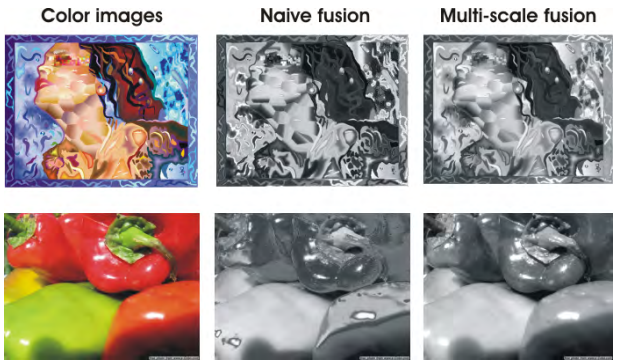


**FIGURE 9.** Compared with the multi-scale fusion, the naive approach introduces haloing artifacts.

for other applications such as image enhancement [40], HDR imaging [42], underwater imaging [44], [47] and image dehazing [43].

Applying the fusion process independently at every scale level the potential artifacts due to the sharp transitions of the weight maps are significantly reduced since it somehow respects simplistic models of the human eye.

## IV. EXPERIMENTAL RESULTS

This section validates our approach. After having motivated the use of our multiple weight maps, we present a number of results that demonstrate that our approach is competitive with recent state-of-the-art optimization-based approaches in terms of still image decolorization quality. Therefore, an initial qualitative assessment (Section IV-B) is completed by a quantitative evaluation (Section IV-C) and a perceptual study (Section IV-D). In a second step, temporal consistency is investigated in synthetic and natural video decolorization scenarios (Section IV-E). When dealing with video, our approach appears to outperform the previous approaches that are equivalent in terms of still image decolorization quality. This is a strong advantage of our approach, in addition to its simplicity.

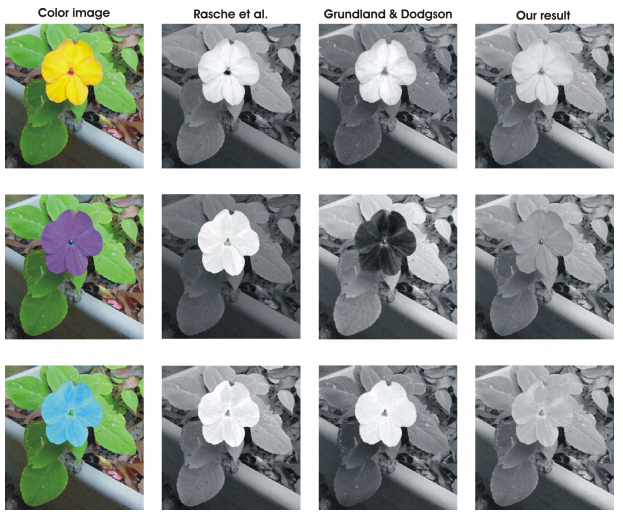


**FIGURE 10.** Consistent color-to-gray mapping. Please observe how differently the considered methods map the leaves and the flower. Our operator is more consistent compared with the techniques of [3] and [5] being able to mapinto the same grayscale level the leaves while the flower is converted into different grayscale levels.

### A. WEIGHT MAP ASSESSMENT

To motivate our weight maps merging process, and assess the impact of each weight map we have decolorized the entire set of 24 color images considered in our evaluation by using only one weight map at a time (please refer to FIGURE 5) and have compared them to the decolorized image obtained based on the combined weight maps. We observe that the combination of weight maps achieves the more consistent results by being the most accurate in reproducing the structures/textures/edges that are visible in the color image. In contrast, the chromatic and exposedness weight maps fail in rendering some significant color differences, while the saliency map misses some details and textures by inducing some undesired saturation.

### B. QUALITATIVE VISUAL ASSESSMENT

In FIGURE 10 are presented several versions of the same image, in which the flower has been colored differently at

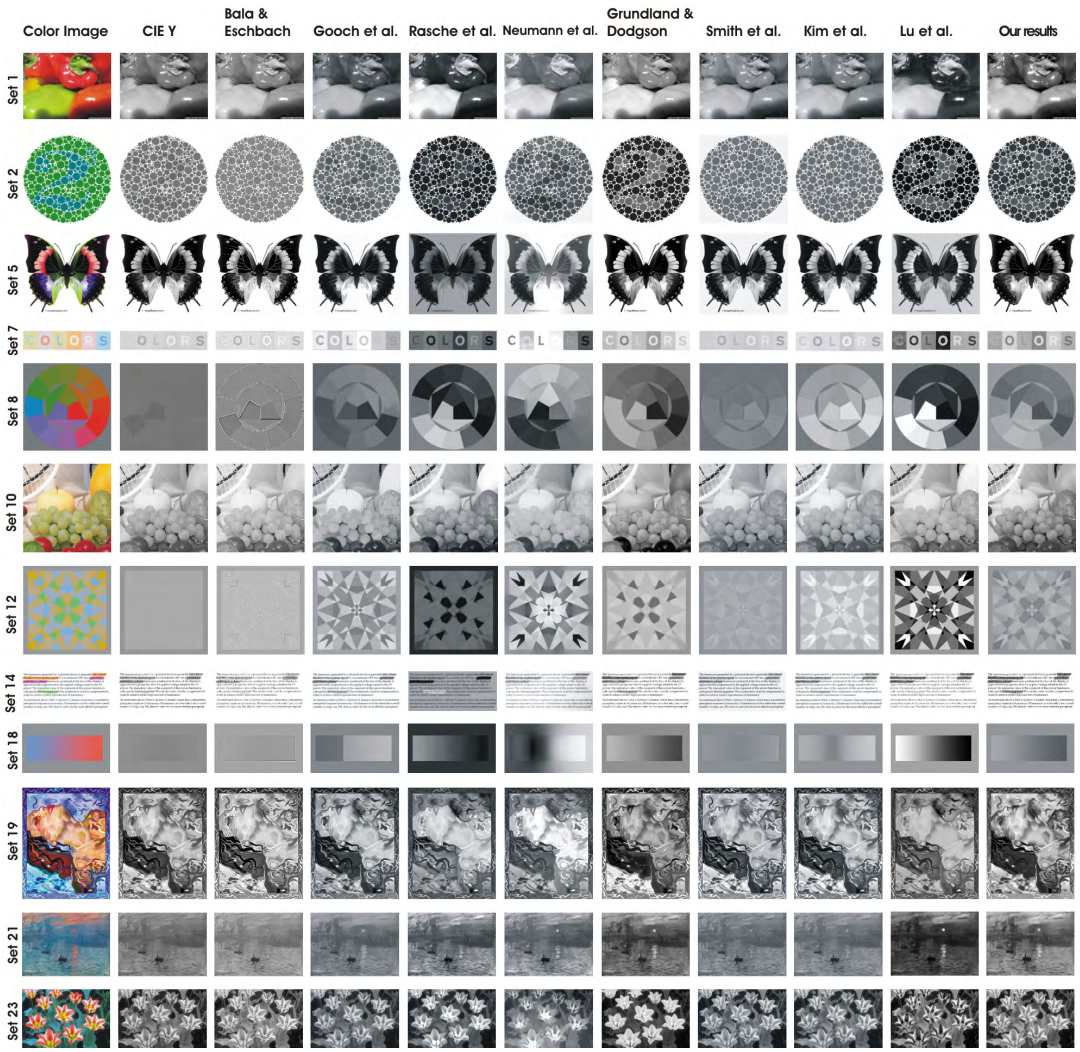


**FIGURE 11.** Comparative results. From left to right the grayscale results obtained by applying *CIEY*, Bala and Eschbach [1], Gooch *et al.* [2], Rasche *et al.* [3], Neumann *et al.* [17], Grundland and Dodgson [5], Smith *et al.* [11], Kim *et al.* [18], Lu *et al.* [19] methods and our fusion-based operator.

each instance. As can be observed, the global approach of Grundland and Dodgson [5] generate dissimilar gray levels for the same region on different instances (note the leaves and the background mapping). Differently, our technique and [11] yield more consistent outputs, while still differentiating the flower colors. We conclude that our strategy has the advantage to preserve both global and local characteristics, therefore being consistent for this challenging case.

### C. QUANTITATIVE EVALUATION

Our approach has been tested extensively for a large set of images. FIGURE 11 shows comparative results against several state-of-the-art decolorization operators, including the ones of Bala and Eschbach [1], Gooch *et al.* [2], Rasche *et al.* [3], Grundland and Dodgson [5], Neumann *et al.* [17], Smith *et al.* [11], Kim *et al.* [18], and Lu *et al.* [19]. As a general observation, despite its relative simplicity, our proposed solution compares favorably to all other operators.

An in-depth critical analysis of those figures reveals that our method fails to properly render the color contrast on set 6 in FIGURE 11. However, most other methods suffer from the same problem for this image, and the method of Grundland and Dodgson [5], which performs reasonably well on this image, suffers from other kinds of artifacts (see the lack of contrast on the top-right corner in set 4 or the lack of grayscale variety in the flowers depicted in set 23). We conclude that no other method is constantly better than ours when dealing with still image decolorization. Hence, our method is certainly a good alternative, given its computational simplicity.

Even whether the visual perception of decolorized image investigated through a subjective perceptual study in the next subsection remains the definitive argument to evaluate a method, for practical reasons, measuring quantitatively the

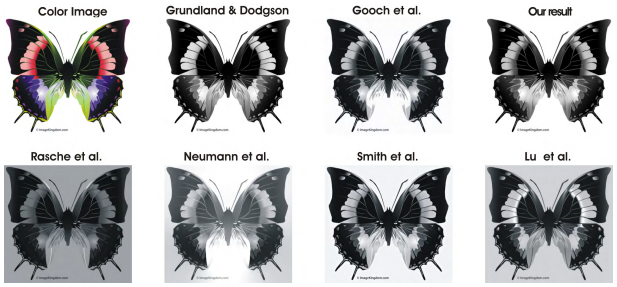


**FIGURE 12.** Black and white preservation of different operators. As can be observed the operators shown in the bottom row distort significantly the original white color regions.

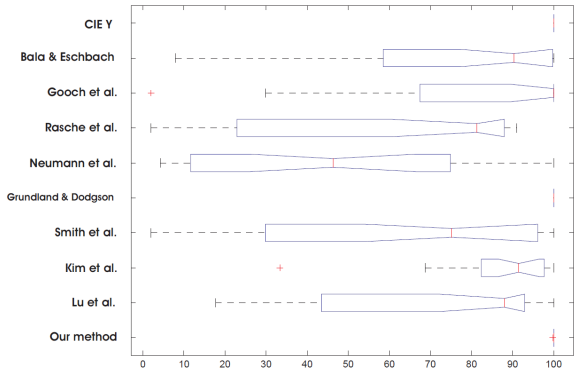


**FIGURE 13.** Quantitative evaluation using the index $\delta_e$ that verifies the preservation of the extreme values (black and white colors) in the decolorised image versions. The results have been interpreted statistically using analysis of variance (ANOVA) [54].

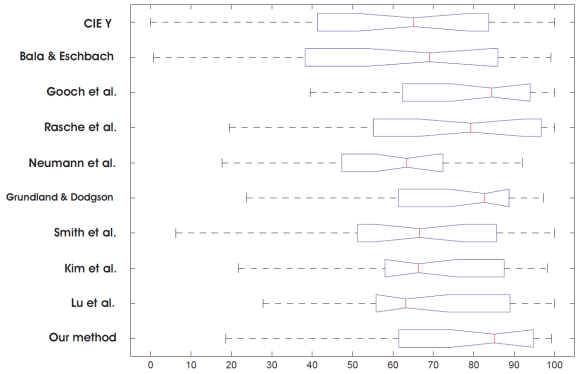


**FIGURE 14.** Quantitative evaluation using the index $\delta_s$ that estimates the level of saliency preservation by comparing the saliency maps of the reference color image with the saliency maps of decolorised versions. The results have been interpreted statistically using analysis of variance (ANOVA) [54].

most significant factors that affect the visual perception of transformed images is relevant.

Since in the literature there is no specialized metric generally accepted to measure the accuracy of a color-to-grayscale transformation. We first introduce and motivate the metrics we have considered in our quantitative comparative study. In short, we propose to define two quality indices that measure, in the decolorized image, the preservation of the extreme pixel values and of the color saliency map between the color and decolorized images.

We first define the extreme values index $\delta_e$ to verify the preservation of the location of achromatic values (black and white pixels) between the original and the decolorized image. This index is relevant because the most evident and undesirable visual distortion of a decolorization transform is probably the conversion of the original black and white values (extreme values that appear in the original color image) into mid-tones (this kind of conversion error is shown in FIGURE 12).

This index $\delta_e$ is defined as:

$$\delta_e = N_{G_{BW}}/N_{C_{BW}} \tag{9}$$

where $N_{G_{BW}}$ is the number of pixels in the original color image that are black or white and $N_{C_{BW}}$ counts the number of those pixel locations that present black and white values in the converted grayscale image. The results have been interpreted statistically through variance analysis (ANOVA) [54] in FIGURE 13 for the 24 images the have been used in [55], and for the same SoA decolorization methods considered in FIGURE 11.

Our second metric is based on the observation (see FIGURE 6) that the saliency map appears to provide a good reference cue to evaluate the perceptual quality of the decolorization. A conversion that loses most of the saliency structure is definitely not appropriate (see the standard conversion in FIGURE 6). Moreover, the preservation of the saliency clearly facilitates fast image understanding since it impacts the way our brain speeds-up the localization and selection of relevant items in a scene. Therefore, we propose to introduce a second decolorization quality index to complete the extreme value index. This second index assesses saliency preservation during decolorization, and is denoted $\delta_s$. As a preliminary step to the $\delta_s$ computation, first we compute a saliency map [48] for both the original and decolorized images. In color images, the saliency map $Sal(x, y)$ computation is performed in $L^*a^*b^*$ color space using:

$$Sal(x, y) = \left\| \mathbf{I}_{\omega_{hc}}(x, y) - \mathbf{I}_{\mu} \right\| \tag{10}$$

where $\mathbf{I}_{\mu}$ is the mean image feature vector, $\mathbf{I}_{\omega_{hc}}(x, y)$ is the corresponding image pixel vector value in the Gaussian blurred version (using a $5 \times 5$ separable binomial kernel) of the original image, and $\|.\|$ is the L2 norm.

The saliency index $\delta_s$ is then simply defined as:

$$\delta_s = N_{Sal_G}/N_{Sal_C} \tag{11}$$

where $N_{Sal_C}$ represents the number of pixels of the color image saliency map that lie above 70% of the maximal saliency value, while $N_{Sal_G}$ characterizes the number of pixels of the decolorized image saliency map that lie above 70% of the maximal saliency value.

Analyzing the results of the extreme value index $\delta_l$ (the graphic shown in FIGURE 13) it can be observed that our

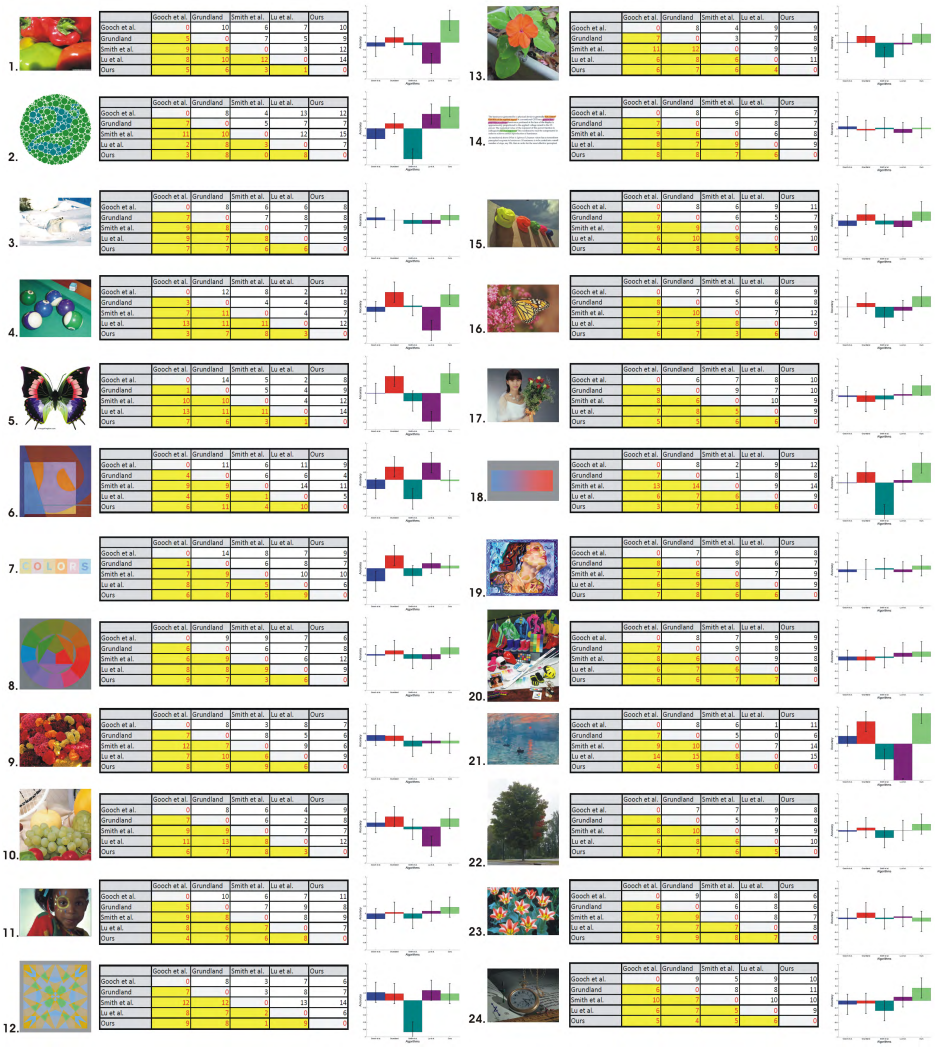

**FIGURE 15.** Perceptual evaluation. The frequency matrices and the accuracy scores with 95% confidence intervals obtained by applying Thurstone's law for every of the 24 images. The 95% confidence intervals $CI = \pm 0.2778$.

operator together with CIE Y but also the operators of Grundland and Dodgson [5], Gooch *et al.* [2], Kim *et al.* [18] preserve the white and black pixel values. This characteristic can be observed in FIGURE 12 but also by a close inspection of the results corresponding to the 24 images part of them shown in FIGURE 11. On the other hand, the saliency index $\delta_s$ (the graphic shown in FIGURE 14), divides roughly the considered operators in two main classes: those that preserve quite well the saliency (Grundland and Dodgson [5], Gooch *et al.* [2] and our operator) and those that lose significantly this important global information (e.g. Neumann *et al.* [17], Bala and Eschbach [1], Smith *et al.* [11]).

Moreover, by comparing our quantitative validation with the perceptual study of Cadik [55], we have observed that the results of the two assessment methodologies are quite correlated and lead to the same conclusions. Indeed, algorithms with high degree of saliency preservation and small local distortion of extreme values obtain the highest perceptual scores in [55]. For example in FIGURE 13 and FIGURE 14, the methods of Grundland and Dodgson [5] and Gooch *et al.* [2] are among the best methods according to our proposed criteria. They were also preferred ones in the perceptual study of Cadik [55]. In contrast, the approaches such as the ones of Rasche *et al.* [3] and Bala and Eschbach [1] preserve relatively well the original saliency information but introduce a high degree of extreme values distortion, which makes them to be classified less perceptually accurate in the study of Cadik [55]. Similarly, *CIEY* does not introduce important local extreme values distortion but is not able to preserve the global appearance in many cases.

### D. PERCEPTUAL EVALUATION

Since the problem of color-to-grayscale transformation is fundamentally related to image perception, we performed a validation based on the preference of real observers.

The proposed perceptual validation is implemented in a similar manner as the strategy used in the recent study of Montagna and Finlayson [56], which consists in a pairwise comparison where each decolorized image version is compared to all the other considered color-to-grayscale versions. For this validation we used the same set of 24 color images that were analyzed previously in [55]. Following the results reported by Cadik [55] we considered the following 5 color-to-grayscale techniques: Gooch *et al.* [2], Grundland and Dodgson [5], Smith *et al.* [11], Lu *et al.* [19] and our method.

The number of methods considered in our perceptual validation has been limited to 5 to limit the load on the viewers since every volunteer had to observe a number of $n(n-1)/2$ ($n = 5$ is the number of analyzed color-to-grayscale techniques). Consequently, there were a number of 10 comparisons for each of the 24 input color images, which results in 240 choices performed by each participant. 15 volunteers were involved in our experiment. They declared to have normal color vision or corrected to normal acuity. Basically, each participant has been requested to select from a pair of grayscale images that have been displayed in a random order and on a large display (ASUS VG248QE 24-inch LED display) together with the original color image. The setup was placed in a dark room and the participants had no constraint related to the time available to perform their selections. For each pair of grayscale images (displayed together with the reference color image), the observers had to choose which version was providing the most accurate reproduction of the original color image. The participants were asked to analyze general features of the transformations (e.g. contrast, similar colors mapping, local artifacts, edges preservation). As an average, the time required to perform the test was about 80 minutes per participant.

Based on these observations we have built a frequency matrix, which represents for each technique the number of time it is preferred during each comparison. For each tested image the frequency matrices are stored as $n \times n$ raw data matrices where the value at position $(i, j)$ represents the number of users who preferred the technique of column $j$ to the method of row $i$. As in [56], to interpret the obtained frequency matrices we employ the well-known Thurstone's law of comparative judgment, case V [57]. As in the study of Morovic and Luo [58], the $n \times n$ z-score matrix is computed as a logarithmic transformation of the frequency matrix scaled with a coefficient $\chi = 0.667$. Finally, for each tested image, the accuracy score of the evaluated techniques is computed as a mean score per column of the $n \times n$ z-score matrix. Based on the empirical study of Montag [59], the 95% confidence intervals are expressed as $CI = \hat{A}\acute{s}1.96\sigma_{obs}$, where the estimated observed standard deviation has the value $\sigma_{obs} = 0.142$.

The frequency matrices and the accuracy scores with 95% confidence intervals obtained by applying Thurstone's law are shown for each of the 24 images in FIGURE 15. The overall preference score of the tested grayscale techniques is shown in FIGURE 16. Positive values indicate preferred grayscale techniques while negative scores, the opposite.

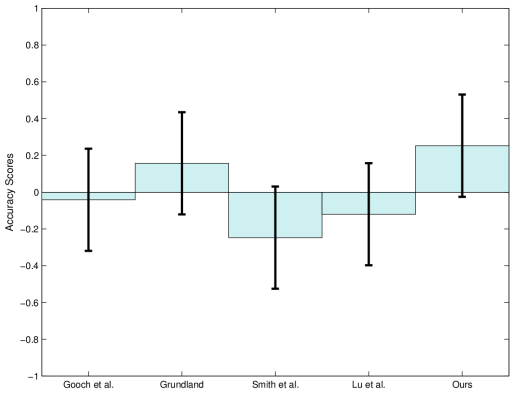


**FIGURE 16.** Perceptual evaluation. Interpretation of our dataset by applying Thurstone's law to preference judgements. On the y axis are shown the normalized scores for each method with the positive values representing preferred grayscale techniques. The error bars represent 95% confidence intervals, $CI = \pm 0.2778$.

Analyzing the results of the perceptual evaluation, it can be observed that the technique of Grundland and Dodgson [5] and ours have been generally preferred over the other methods.

A second class of preference includes the technique of Gooch *et al.* [2] and Lu *et al.* [19] while the technique of Smith *et al.* [11] seems to score lower mainly due to the poor contrast and local artifacts introduced close to the edges. Additionally, compared with our quantitative validation, the perceptual evaluation shown important correlation mostly with the $\delta_l$ descriptor.

### E. VIDEO DECOLORIZATION TEMPORAL CONSISTENCY ASSESSMENT

Now that we have shown that our method is competitive compared with recent state-of-the-art approaches in terms of still image decolorization, we investigate its behavior when processing video content. Video decolorization, which might for example be relevant for viewers with color-deficient vision, adds another dimension to the problem of image decolorization, as temporal coherence needs to be guaranteed for the entire video sequence. Formally, a temporally consistent mapping assigns similar gray levels to regions that correspond to the same colored content across time, thereby avoiding undesired flickering effects. Smith *et al.* [11] have shown that local approaches are suitable for this purpose in the sense that a local area is processed independently of its location and of the content of the entire image.

This observation is important since our approach is primarily local.

In contrast, the methods that compute an optimal mapping scheme from the (statistics of the) image to decolorize generally lack of consistency across time, since a relatively minor change in the scene might affect the mapping scheme (see for example FIGURE 13 in [10]). Those arguments in favor of our approach are confirmed by a number of decolorized natural and synthetic video sequences are provided

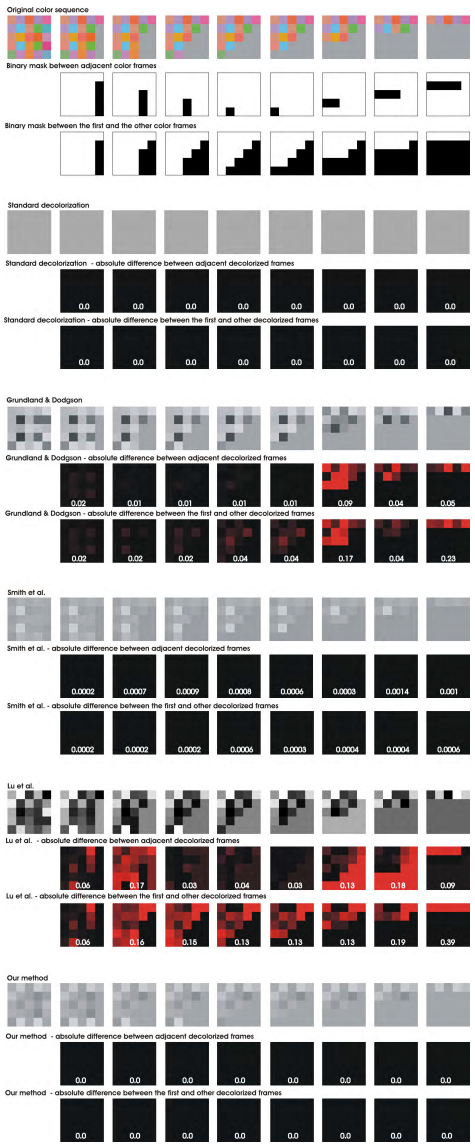


**FIGURE 17.** Decolorization temporal consistency assessment. The top row shows a sequence of frames that contains isoluminant color patches. Between two consecutive frames, the patches either keep their color or switch to a mid-gray tone. The second and third rows depict in white the pixels that do not change compared to the previous or the first frame, respectively. The remaining rows present the decolorized images and the temporal inconsistency metrics obtained when using the standard luminance-based decolorization, Grundland and Dodgson [5], Smith *et al.* [11], Lu *et al.* [7] and our method. In each of the two rows following a row of decolorized images, the temporal inconsistency is depicted in red tones (deep red = high value), and measures -on a patch basis- the mean absolute difference of the decolorized pixels compared to the previous or the first frame, respectively. This inconsistency metric is only computed between pairs of patches that have the same color in the initial images, as identified by the white pixels in the second and third rows.

at https://www.youtube.com/watch?v=i3rUiVGnTXQ). The visual observation of those sequences confirms that our method is temporally consistent. To quantify this statement compared to other decolorization approaches, Fig. 17 considers a synthetically generated sequence. The adjacent frames of the synthetic video contain the same isolluminant color patches at the same locations, except for the patches that are transformed to a mid-gray tone. As can be seen, the standard decolorization approach maps all the color patches to the same gray tone. Moreover, it can be observed that even though the technique of Grundland and Dodgson [5] decolorizes each individual frame properly, it is not able to preserve the same grayscale level for a given color patch all along the sequence. A similar observation is valid for the Lu *et al.* [7] approach. In contrast, our method decolorizes the sequence in a temporally consistent manner, while preserving the initial color contrasts.

As depicted in FIGURE 17, we quantitatively assess the temporal consistency of video decolorization by measuring the gray level consistency between pairs of decolorized frames. Therefore, we introduce an original two-step strategy. The first step identifies the pixel locations where the two investigated color image have the same color appearance, as measured by the Euclidean distance in the $CIEL^*a^*b^*$ color space.

The second step then computes the mean absolute difference of decolorized pixels over those locations where color pixels were identical. This mean absolute difference is presented in FIGURE 17 for a set of methods, including the technique of Grundland and Dodgson [5], the technique of Smith *et al.* [11] the recent approach of Lu *et al.* [7] and our method. We observed that, the techniques of Grundland and Dodgson [5] and Lu *et al.* [7] are not able to preserve the grayscale values of corresponding color patches over the adjacent frames. The method of Smith *et al.* [11], while being visually reasonably consistent, reveals some small video inconsistency when analyzed with the proposed assessment. In contrast, our approach remains consistent whatever the pairs of frames that are considered along the time.

To the best of our knowledge the above quantitative assessment methodology is the first one to evaluate the temporal coherence for video decolorization. Obviously, when considering real-life natural video sequences, it should ideally involve some kind of motion compensation between consecutive frames. Such investigation is however beyond the scope of our paper, and is left for future work.

### F. DISCUSSION

In our extensive experiments, the fusion-based operator performs generally well on still images. Compared to other methods that also perform very well on still images, it offers the additional advantage of being temporally consistent. Another important advantage of our fusion strategy is the computation time. While our optimized approach processes a $800 \times 600$ image in approximately 200 ms (Intel Core i7 CPU, 16GB RAM), Smith *et al.* [11] method takes 6.7 - 10.8 seconds for $570 \times 593$ image, Decolorize [5] -unoptimized code $-3.5$ seconds for a $800 \times 600$ image and the optimized code of Kim *et al.* [18] converts a $800 \times 600$ image in 1-2 seconds). Also computationally effective, the method of Kim *et al.* [18] is not able to solve completely the mapping limitations of [2], tending to reduce the original contrast and to lose some of the original saliency (please refer to comparative results).

## V. CONCLUSION

In this work we introduced an effective decolorization approach built on the multi-scale fusion strategy. Selecting appropriate weight maps and inputs, our multi-scale fusion strategy shown to be effective for decolorization of various color images by transferring the original image saliency and color contrast in a spatially consistent manner to a visually pleasant grayscale image. This has been demonstrated through an extensive quantitative and qualitative validation. In addition, our method is simple to implement (no optimization required, suited to parallelization) and offers the advantage of being temporally consistent. To evaluate the temporal coherence for video decolorization we introduce a quantitative assessment methodology.

## ACKNOWLEDGMENTS

This work was supported in part by the Romanian Government UEFISCDI under Project PN-III-P2-2.1-PED-2016-0940, in part by the 2020 European Union Research and Innovation Horizon 2020 under Grant Marie Sklodowska-Curie No 712949 (TECNIOspring PLUS), and in part by the Agency for the Competitiveness of the Company of the Generalitat de Catalunya under Grant ACCIO: TECSPR17-1-0054.

## REFERENCES

[1] R. Bala and R. Eschbach, "Spatial color-to-grayscale transform preserving chrominance edge information," in *Proc. Color Imaging Conf.*, 2004, pp. 82–86.

[2] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch, "Color2Gray: Salience-preserving color removal," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 634–639, 2005.

[3] K. Rasche, R. Geist, and J. Westall, "Re-coloring images for gamuts of lower dimension," *Comput. Graph. Forum*, vol. 24, no. 3, pp. 423–432, 2005.

[4] M.-J. Yoo, I.-K. Lee, and S. Lee, "Color sequence preserving decolorization," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 373–383, May 2015.

[5] M. Grundland and N. A. Dodgson, "Decolorize: Fast, contrast enhancing, color to grayscale conversion," *Pattern Recognit.*, vol. 40, no. 11, pp. 2891–2896, 2007.

[6] Z. Ji, M.-E. Fang, Y. Wang, and W. Ma, "Efficient decolorization preserving dominant distinctions," *Vis. Comput.*, vol. 32, no. 12, pp. 1621–1631, 2016.

[7] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization with perception-based quality metrics," *Int. J. Comput. Vis.*, vol. 110, no. 2, pp. 222–239, Nov. 2014.

[8] Q. Liu, P. X. Liu, W. Xie, Y. Wang, and D. Liang, "GcsDecolor: Gradient correlation similarity for efficient contrast preserving decolorization," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2889–2904, Sep. 2015.

[9] W. Zhu, R. Hu, and L. Liu, "Grey conversion via perceived-contrast," *Vis. Comput.*, vol. 30, no. 3, pp. 299–309, 2014.

[10] H. Du, S. He, B. Sheng, L. Ma, and R. W. H. Lau, "Saliency-guided color-to-gray conversion using region-based optimization," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 434–443, Jan. 2015.

[11] K. Smith, P.-E. Landes, J. Thollot, and K. Myszkowski, "Apparent greyscale: A simple and fast conversion to perceptually accurate images and video," *Comput. Graph. Forum*, vol. 27, no. 2, pp. 193–200, 2008.

[12] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Enhancing by saliency-guided decolorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 257–264.

[13] C. Ancuti and C. O. Ancuti, "Laplacian-guided image decolorization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 4107–4111.

[14] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "Image and video decolorization by fusion," in *Proc. Asian Conf. Comput. Vis.*, 2011, pp. 79–92.

[15] M. Qiu, G. Finlayson, and G. Qiu, "Contrast maximizing and brightness preserving color to grayscale image conversion," in *Proc. Conf. Colour Graph., Imag. Vis.*, 2008, pp. 347–351.

[16] Z. Jin and M. K. Ng, "A contrast maximization method for color-to-grayscale conversion," *Multidimensional Syst. Signal Process.*, vol. 26, no. 3, pp. 869–877, Jul. 2015.

[17] L. Neumann, M. Čadík, and A. Nemcsics, "An efficient perception-based adaptive color to gray transformation," in *Proc. Comput. Aesthetics*, 2007, pp. 73–80.

[18] Y. Kim, C. Jang, J. Demouth, and S. Lee, "Robust color-to-gray via nonlinear global mapping," *ACM Trans. Graph*, vol. 28, no. 5, 2009, Art. no. 161.

[19] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization," in *Proc. IEEE Int. Conf. Comput. Photogr.*, Apr. 2012, pp. 1–7.

[20] Z. Jin, F. Li, and M. K. Ng, "A variational approach for image decolorization by variance maximization," *SIAM J. Imag. Sci.*, vol. 24, no. 1, pp. 434–443, Jan. 2015.

[21] C. Lau, W. Heidrich, and R. Mantiuk, "Cluster-based color space optimizations," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1172–1179.

[22] A. Nemcsics, "Color space of the coloroid color system," *Color Res. Appl.*, vol. 12, no. 3, pp. 135–146, 1987.

[23] A. Alsam and M. S. Drew, "Fast multispectral2gray," *J. Imag. Sci. Technol.*, vol. 53, no. 6, pp. 60401-1–60401-10, 2009.

[24] D. A. Socolinsky and L. B. Wolff, "Multispectral image visualization through first-order fusion," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 923–931, Aug. 2002.

[25] X. Zheng, J. Feng, and B. Zhou, "Efficient color-to-gray conversion for digital images in gradient domain," in *Proc. 14th ACM SIGGRAPH Int. Conf. Virtual Reality Continuum Appl. Ind.*, vol. 13, Nov. 2015, pp. 85–88.

[26] G. D. Finlayson, D. Connah, and M. S. Drew, "Lookup-table-based gradient field reconstruction," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2827–2836, Oct. 2011.

[27] M. D. Fairchild and E. Pirrotta, "Predicting the lightness of chromatic object colors using CIELAB," *Color Res. Appl.*, vol. 16, no. 6, pp. 385–393, 1991.

[28] C.-H. Son, K. Lee, and H. Choo, "Inverse color to black-and-white halftone conversion via dictionary learning and color mapping," *Inf. Sci.*, vol. 299, pp. 1–19, Apr. 2015.

[29] R. L. D. Queiroz and K. M. Braun, "Color to gray and back: Color embedding into textured gray images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1464–1470, Jun. 2006.

[30] R. Hedjam, H. Z. Nafchi, M. Kalacska, and M. Cheriet, "Influence of color-to-gray conversion on the performance of document image binarization: Toward a novel optimization problem," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3637–3651, Nov. 2015.

[31] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Single-scale fusion: An effective approach to merging images," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 65–78, Jan. 2017.

[32] T. Wu and A. Toet, "Color-to-grayscale conversion through weighted multiresolution channel fusion," *J. Electron. Imag.*, vol. 23, no. 4, p. 043004, Jul. 2014.

[33] R. Raskar, A. Ilie, and J. Yu, "Image fusion for context enhancement and video surrealism," in *Proc. Int. Symp. Non-Photorealistic Animation Rendering (NPAR)*, 2004, pp. 85–152.

[34] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph*, vol. 22, no. 3, pp. 313–318, 2003.

[35] A. Agarwala *et al.*, "Interactive digital photomontage," *ACM Trans. Graph*, vol. 23, no. 3, pp. 294–302, 2004.

[36] R. R. Brinkmann, *The Art and Science of Digital Compositing*. San Mateo, CA, USA: Morgan Kaufmann, 1999.

[37] M. Grundland, R. Vohra, G. P. Williams, and N. A. Dodgson, "Cross dissolve without cross fade: Preserving contrast, color and salience in image compositing," *Comput. Graph. Forum*, vol. 25, no. 3, pp. 577–586, 2006.

[38] S. Arpa, T. Ritschel, K. Myszkowski, T. Çapın, and H.-P. Seidel, "Purkinje images: Conveying different content for different luminance adaptations in a single image," *Comput. Graph. Forum*, vol. 34, no. 1, pp. 116–126, 2014.

[39] A. G. Kirk and J. F. O'Brien, "Perceptually based tone mapping for low-light conditions," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 42.

[40] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proc. IEEE Int. Conf. Comput. Vis.*, May 1993, pp. 173–182.

[41] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proc. 15th Pacific Conf. Comput. Graph. Appl. (PG)*, Oct./Nov. 2007, pp. 382–390.

[42] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.
[43] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.
[44] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.
[45] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Night-time dehazing by fusion," in *Proc. IEEE ICIP*, Sep. 2016, pp. 2256–2260.
[46] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, R. Garcia, and A. C. Bovik, "Multi-scale underwater descattering," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 4202–4207.
[47] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
[48] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, Jun. 2009, pp. 1597–1604.
[49] P. Burt and T. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.
[50] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
[51] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph*, vol. 21, no. 3, pp. 257–266, 2002.
[52] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph*, vol. 27, no. 3, 2008, Art. no. 67.
[53] K. Subr, C. Soler, and F. Durand, "Edge-preserving multiscale image decomposition based on local extrema," *ACM Trans. Graph*, vol. 28, no. 5, 2009, Art. no. 147.
[54] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 5th ed. Boston, MA, USA: Allyn & Bacon, 2005.
[55] M. Ĉadík, "Perceptual evaluation of color-to-grayscale image conversions," *Comput. Graph. Forum*, vol. 27, no. 7, pp. 1745–1754, 2008.
[56] R. Montagna and G. D. Finlayson, "Reducing integrability error of color tensor gradients for image fusion," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4072–4085, Oct. 2013.
[57] L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.*, vol. 34, no. 4, p. 273, 1927.
[58] J. Morovic and M. R. Luo, "Evaluating gamut mapping algorithms for universal applicability," *Color Res. Appl.*, vol. 26, no. 1, pp. 85–102, 2001.
[59] E. D. Montag, "Empirical formula for creating error bars for the method of paired comparison," *J. Electron. Imag.*, vol. 15, no. 1, p. 010502, 2006.

**COSMIN ANCUTI** received the Ph.D. degree from Hasselt University, Belgium, in 2009. He was a Post-Doctoral Fellow with IMINDS and the Intel Exascience Laboratory (imec), Leuven, Belgium, from 2010 to 2012, and a Research Fellow with the Université catholique de Louvain, Belgium, from 2015 to 2017. He is currently a Senior Researcher/Lecturer with the Politehnica University of Timişoara, Romania, and the Institute of Informatics and Applications, University of Girona, Spain. He has authored over 50 papers published in international conference proceedings and journals and co-organizes the IEEE CVPR NTIRE Workshop. His area of interests includes image and video enhancement techniques, computational photography, and low-level computer vision.

**CODRUTA O. ANCUTI** received the Ph.D. degree from Hasselt University, Belgium, in 2011. From 2015 to 2017, she was a Research Fellow with the ViCOROB Group, University of Girona, Spain. She is currently a Senior Researcher/Lecturer with the Faculty of Electrical and Telecommunication Engineering, Politehnica University of Timişoara. Her main interest of research includes image understanding and visual perception. She is the first that introduced several single images-based enhancing techniques built on the multiscale fusion (e.g., color-to-grayscale, image dehazing, underwater image, and video restoration). Her work received the Best Paper Award at NTIRE 2017 (CVPR workshop).

**CHRISTOPHE DE VLEESCHOUWER** was a Senior Research Engineer with imec from 1999 to 2000, a Post-Doctoral Research Fellow with the University of California at Berkeley from 2001 to 2002 and EPFL in 2004, and a Visiting Faculty with Carnegie Mellon University from 2014 to 2015. He is currently a Research Director with the Belgian NSF and an Associate Professor with the ISP Group, Université catholique de Louvain, Belgium. He has co-authored over 35 journal papers or book chapters. holds two patents. His main interests lie in video and image processing for content management, transmission, and interpretation. He is enthusiastic about nonlinear and sparse signal expansion techniques, the ensemble of classifiers, multiview video processing, and the graph-based formalization of vision problems. He served as an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, and has been a (technical) program committee member for most conferences dealing with video communication and image processing.

**MATEU SBERT** received the M.Sc. degree in physics from Valencia University in 1977 and in mathematics from UNED, Madrid, in 1983, and the Ph.D. degree in computer science from Barcelona TU, in 1997, where he received the Best Ph.D. Award. He is currently a Professor in computer science at Tianjin University, China, and Girona University, Spain. He has authored or co-authored over 200 papers and four books, receiving more than 4000 citations to his work. His research interests include Monte Carlo, integral geometry, and information theory techniques for computer graphics, image processing, and visualization.