Universitatea Politehnica Timișoara

Habilitation Thesis
Teză de Abilitare

Image Enhancing Techniques Based on Fusion

Metode de Imbunatatire a Imaginilor Bazate pe Fuziune

Author:

Codruța Orniana Ancuți

2019

Cuprins

Partea I

Abstract

I.1 Abstract

The present habilitation thesis summarizes the research contributions of the candidate obtained between 2011 (when the candidate obtained the PhD) to this date.

The most significant research activity and the obtained results are presented structured in several parts that represent original contributions in the field of image processing and computational photography: image dehazing (day-time single image dehazing, night-time single image dehazing, dehazing evaluation dataset), underwater images enhancement, image decolorisation and single scale fusion technique for effectively merging images.

Image dehazing deals with the problem of enhancing the visibility in terms of color and details for images degraded by haze. In outdoor environments, haze phenomena appears when the light reflected from object surfaces is scattered due to the impurities of the aerosol, or due to the presence of fog and haze. The yielding hazy images are characterized by poor contrast, lower saturation and additional noise.

Day-time image dehazing. Firstly, we describe a novel single image strategy [1] that demonstrates to accurately dehaze images by only taking as input the original degraded information. Our approach is built on a fusion strategy and derives two inputs from the original image. These inputs are weighted by three normalized weight maps and finally blended in a multi-scale fashion that avoids introducing artifacts. The method is fast being straightforward to implement and shows to outperform the related operators. Our approach performs an effective per-pixel computation, that reduces the amount of artifacts compared with the patch-based methods.

Night-time image dehazing. To deal with the problem of night-time hazy scenes, we introduce a novel modality [2] to compute the airlight component while accounting for the non-uniform illumination presents in nighttime scenes. Unlike the day-time dehazing strategies that estimates a constant atmospheric light over the entire image, we estimate this by accounting the local values and patches of

varying sizes. Our work combines the advantages of using the physical model and uses noise and artifacts reduction ability by employing the fusion strategy.

Image dehazing evaluation dataset. We proposed a dataset that contains 1400+ pairs of images with ground truth reference images and hazy images of the same scene. Our dataset, called D-HAZY [3], is built on the Middlebury and NYU Depth datasets that provide images of various scenes and their corresponding depth maps. Based on the depth information and using the physical model of a hazy medium we derive a corresponding hazy scene with high fidelity. Using D-HAZY dataset, we perform a comprehensive quantitative evaluation of the recent techniques.

Enhancing Underwater Images. We describe an effective technique [4] that is able to enhance underwater images. Our strategy derives the inputs and the weight measures only from the degraded version of the image. In order to overcome the limitations of the underwater medium we define two inputs that represent color corrected and contrast enhanced versions of the original underwater image/frame, but also four weight maps that aim to increase the visibility of the distant objects degraded due to the medium scattering and absorption. The enhanced images and videos are characterized by reduced noise level, better exposedness of the dark regions, improved global contrast while the finest details and edges are enhanced significantly. In addition, the utility of our enhancing technique is proved for several challenging applications (image matching, segmentation, etc.).

Moreover, for highly scattered underwater scenes, we introduce a novel approach [5]. Following the optical underwater model, we first compute the back-scattered light by searching for the brightest location along each image patch. By simply applying the optical model using our local estimate of the back-scattered light, we are able to obtain a good degree of visual restoration, even on in extreme underwater scenes. Our results demonstrates that our proposed solution can deal with the difficult situations where most of the existing enhancing underwater approaches fail, such as when using artificial illumination at higher depths and high turbidity.

Image decolorization (color-to-grayscale) deals with the problem of converting a color image (with three-RGB-channels) into a single channel image version. Often, the standard decolorization conversion is simply employed as the luminance

channel of different color spaces. However, this simple global mapping disregards important chromatic information and therefore, in many cases, the output does not preserve the original appearance. Our grayscale transformation [6], designed in RGB color space and takes as individual inputs the three color channels (R, G, B). Our technique is guided by two weight maps that transfer in the final result the most significant information of each derived input (RGB color channels). In order to minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion.

Single scale image fusion. Image fusion plays an important role in a wide range of imaging applications. One of the main problem for all the fusion algorithms is related with the ability to use it for real-time applications. In our recent work [7] we introduced a single-scale fusion strategy that is more computational effective for blending multiple-sources images and produces similar results in comparison with the traditional multi-scale image fusion technique (MSF). Interestingly, our single-scale expression obtained from the multi-scale approximation also provides insightful cues regarding how the MSF process manipulates weights and image features to compute a visually pleasant outcome. The impact is significant, since this approach facilitates to process high resolution images more effectively.

I.2 Rezumat în limba română

Teza de abilitatre prezintă contribuțiile in domeniul cercetarii ale candidatului între anii 2011 (dată la care candidatul a obținut doctoratul) şi până în prezent.

Activitatea de cercetare precum și rezultatele cele mai semnificative obținute, sunt structurate în patru sectiuni, care reprezintă contribuții originale în domeniul procesării de imagini: recuperarea vizibilității în imaginile deteriorate de ceață, tehnici de restaurare a imaginilor subacvatice, tehnici de conversie a imaginilor color în imagini alb-negru, fuziunea imaginilor cu un singur nivel.

Tehnicile de recuperare a vizibilității din imaginile deteriorate de ceață au ca scop imbunatatira vizibilitatii în ceea ce privește culoarea și detaliile pentru

imagini degradate de ceață. Imaginile înregistrate în condiții de ceață sunt caracterizate prin contrast redus, saturație scăzută și zgomot suplimentar.

Recuperarea vizibilității in imaginile deteriorate de ceață pe timp de zi. În lucrarea [1] am introdus o nouă strategie care permite recuperarea cu precizie a detaliilor in astfel de imagini degradate considerand ca si intrare o singură imagine afectata de ceata. Abordarea se bazează pe o strategie de fuziune derivand două intrări din imaginea originală. Aceste intrări sunt ponderate cu trei mărimi caracteristice (*weight maps*) normalizate care controlează procesul de fuziune cu scopul de a reduce introducerea distorsiunilor.

Recuperarea vizibilității in imaginile deteriorate de ceață pe timp de noapte. Recent, petru astfel de scene, in lucrarea [2], am introdus o modalitate nouă de a calcula factorul iluminării folosind si noi constrângeri pentru definirea neomogenității caracteristice înregistrărilor nocturne pe timp de ceata. Spre deosebire de cazurile din timpul zilei, unde strategiile existente estimează o valoare a iluminarii atmosferice constantă pe întreaga imagine, în condiții de noapte, am derivat o procedura de estimare a acestui parametru considerand valorile locale care sunt estimate pentru diferite dimensiuni ale regiunilor locale.

Bază de date pentru evaluarea tehnicilor de restaurare a scenelor afectate de ceata. In lucrarea [3] am introdus o noua baza de date, denumita D-HAZY. Aceasta conține peste 1400 perechi de imagini formate din imagini de referință și imagini afectate de ceață. Baza de date s-a construit folosind seturile de imagini Middelbury și NYU care furnizează imagini ale diferitelor scene și hărțile lor de adâncime corespunzătoare. Scenele afectate de ceata au fost construite sintetic utilizând modelul optic de degradare a imaginiilor în condiții de ceață.

Tehnici de restaurare a imaginilor subacvatice. În lucrarea [4] am introdus o noua tehnică care permite îmbunătățirea vizibilitatii in imagini inregistrate subacvatic. Pentru a ne adapta mediului subacvatic am definit două intrări: prima intrare reprezentand imaginea initiala cu balansul de culoare imbunatatit iar a doua intrare rezultand din imaginea originala cu contrastul global îmbunătățit.

De asemenea am definit patru măsuri caracteristice (*weight maps*) care au ca scop principal ghidarea celor doua intrari derivate. Imaginile și secventele video prelucrate cu metoda noastra sunt caracterizate prin nivel de zgomot redus, o mai bună expunere a regiunilor slab iluminate, contrast îmbunătățit. Aditional, pentru cazurile subacvatice extreme (*highly-scattered*) am introdus o noua tehnica [5]. Aceasta metoda calculeaza mai întâi lumina reflectată prin găsirea regiunilor cu cea mai mare intensitate din imagine. Prin aplicarea directă a modelului optic și folosind estimarii locale a iluminării, am demonstrat ca putem obține un nivel ridicat de restaurare vizuală, chiar si pentru scenele subacvatice extreme cu iluminare neuniforma.

Tehnici de conversie a imaginilor color în imagini alb-negru se refera la conversie unei imagini color într-o imagine monocroma (alb-negru). Conversia standard utilizează canalul de luminanță a diferitelor spații de culoare (e.g. CIEL∗a∗b∗, YCbCr, HSL/HSV). Această conversie standard s-a dovedit în multe cazuri limitata. Transformarea propusa recent [6] utilizează spațiul de culoare RGB considerand ca intrări individuale ale algoritmului de fuziune cele trei canale de culoare individuale. Având în vedere scopul principal de a menține detaliile vizibile din imaginiile originale, algoritmul nostru este ghidat de două hărți caracteristice de transfer. Pentru a minimiza distorsiunile introduse de hărțile de ghidare, abordarea noastră este bazata pe o strategie de fuziune multi-scală.

Fuziunea imaginilor cu un singur nivel. Fuziunea imaginiilor joacă un rol important într-o gamă largă de aplicații in domeniul procesarii de imagini. Una dintre problemele principale este reprezentata de timpii de procesare care limiteaza utilizarea acestor methode pentru aplicatii în timp real. În practică, pentru o execuție eficientă cu imagini cu rezoluție ridicata este necesară alocarea unei memorii substanțiale dar și de putere de calcul. In lucrarea [7] am introdus un algoritm de fuziune folosind o singură scală care are avantajul de a produce rezultate comparative calitativ cu fuziunea multi-scală. Impactul acestei metode este semnificativ, deoarece facilitează procesarea mult mai eficientă a imaginilor de înaltă rezoluție.

Partea II

Contributions

II.1 Single Image-based Dehazing

Haze is an atmospheric phenomenon that significantly degrades the visibility of outdoor scenes. This is mainly due to the atmospheric particles that absorb and scatter light. This paper introduces a novel single image approach that enhances the visibility of such degraded images. Our method is a fusion-based strategy that derives two inputs from the original hazy image by applying a white balance and a contrast enhancing procedure. To blend effectively the information of the derived inputs in order to preserve the regions with good visibility, we filter their important features by computing three measures (weight maps): luminance, chromaticity and saliency. To minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion, using a Laplacian pyramid representation. We are the first to demonstrate the utility and effectiveness of a fusion-based technique for dehazing based on a single degraded image. The method performs in a per-pixel fashion, which is straightforward to implement. The experimental results demonstrate that the method yields comparable and even better results than the more complex state-of-the-art techniques, with the additional advantage of being appropriate for real-time applications.

Often, the images of outdoor scenes are degraded by bad weather conditions. In such cases, atmospheric phenomena like haze and fog degrade significantly the visibility of the captured scene. Since the aerosol is misted by additional particles, the reflected light is scattered and as a result, distant objects and parts of the scene are less visible, which is characterized by reduced contrast and faded colors.

Restoration of images taken in these specific conditions has caught increasing attention in the last years. This task is important in several outdoor applications such as remote sensing, intelligent vehicles, object recognition and surveillance. In remote sensing systems, the recorded bands of reflected light are processed [8, 9] in order to restore the outputs. Multi-image techniques [10] solve the image dehazing problem by processing several input images, that have been taken in different atmospheric conditions. Another alternative [11] is to assume that an approximated 3D geometrical model of the scene is given. In the work of Treibitz and Schechner [12] different angles of polarized filters are used to estimate the haze effects.

A more challenging problem is when only a single degraded image is available. Solutions for such cases have been introduced only recently [13–17].

II.1.1 Background Theory: Optical model

The physical model used in our approaches that are presented in the next sections, is similar to the one employed in previous single image dehazing methods [13–15, 17], initially described by Koschmieder [18]. For the sake of completeness, we provide a brief description of this model in this section.

When examining an outdoor scene from an elevated position, features gradually appear lighter and fading as they are closer towards the horizon. Only a percentage of the reflected light reaches the observer as a result of the absorption in the atmosphere. Furthermore, this light gets mixed with the *airlight* [18] color vector, and due to the scattering effects the scene color is shifted (illustrated in Fig. 1).

Based on this observation, the captured image of a hazy scene $\mathcal{I}_h$ is represented by a linear combination of *direct attenuation* $\mathcal{D}$ and *airlight* $\mathcal{A}$ contributions:

$$\mathcal{I}_h = \mathcal{D} + \mathcal{A} = \mathcal{I} * t\,(x) + A_\infty * (1 - t\,(x)) \tag{1}$$

where $\mathcal{I}_h$ is the image degraded by haze, $\mathcal{I}$ is the scene radiance or haze-free image, A_∞ is the constant airlight color vector and t is the transmission along the cone of vision. This model assumes linear correlation between the reflected light and the distance between the object and observer.

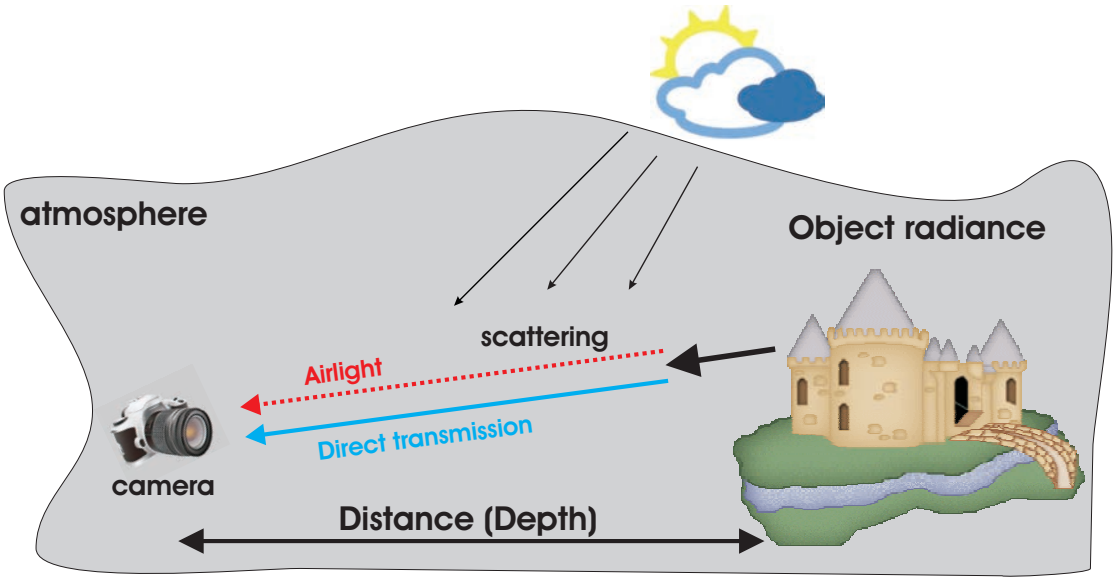

FIGURA 1: The optical model for the atmospheric phenomenon of haze.

This problem is clearly ill-posed, and requires us to recover the unknowns $\mathcal{I}$, A_∞ and $t(x)$ from only a single input image $\mathcal{I}_h$.

In a homogeneous atmosphere, the transmission t is considered to be modulated as:

$$t\left(x\right) = exp(-\beta * d(x)) \tag{2}$$

where β is the attenuation coefficient of the atmosphere due to the scattering and d represents the distance to the observer.

From equation 1, it becomes apparent that the chrominance attenuation becomes increasingly influenced by the airlight, as the optical depth increases:

$$\frac{\mathcal{A}}{\mathcal{D}} = \frac{A_{\infty} * (1 - t(x))}{\mathcal{I} * t(x)} \tag{3}$$

Theoretically, if the transmission and the airlight are known, the haze-free image can be easily computed:

$$\mathcal{I} = A_{\infty} - \left(A_{\infty} - \mathcal{I}_h\right) / t(x) \tag{4}$$

II.1.2 Haze Detection and Transmission Estimation

Existing techniques that deals with satellite images can be classified into two main categories. The first category aims to employ a physical model that aims to estimate an appropriate radiative transform model that is able to process the radiometric effect for given atmospheric conditions and spectral bands then proceeding in the subtraction from the data. Often, such approaches also require detailed auxiliary data regarding the atmospheric conditions at the time and place of the scene acquisition.

The dark object method [8] is a well-known technique within the remote sensing community. In order to recover the details and to enhance the contrast loss due to the haze influence this approach searches to detect the dark object [8]. Considering an image satellite, this is found in the area of zero reflectance and below the pixels that are characterized by the lowest reflectance values in the image (for instance shadowed areas). To find with high accuracy this value, usually all ratios histogram of considered significant bands are analyzed. The main reason for band selection is the property of the materials that are known to reflect more intensely in certain bands than in others. In addition, based on numerous observations and statistics (related with the relative reflection in different TM bands), this process of bands

Landsat TM images (Path 188, Row 32) Band composition 4-3-2; image dates:
a) 08 Septeber 1984; (b) 16 September 2010.

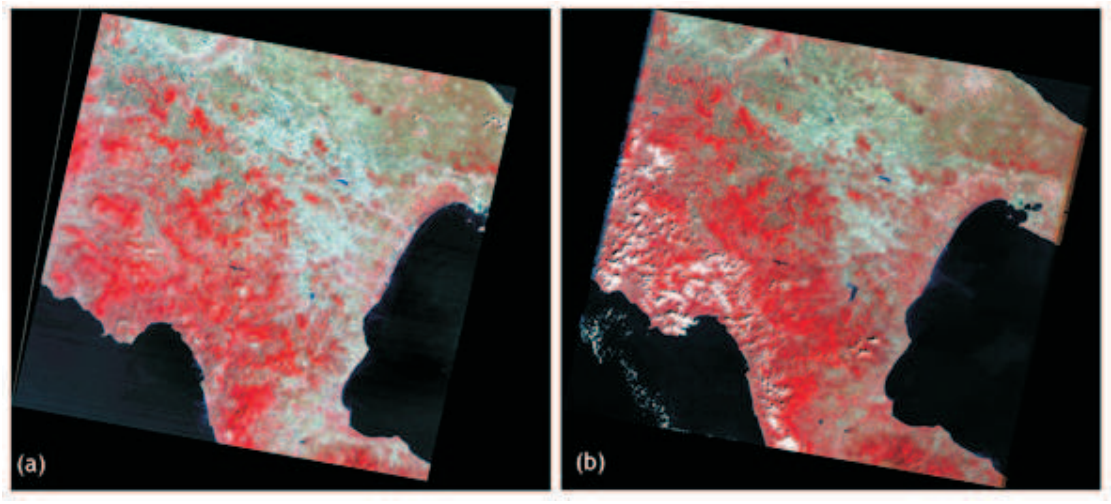

FIGURA 2: Two Landsat TM 2010 and TM 1984 cloud free images (path 188, row 32) with 30 x 30 m spatial resolution

selection can automatically classify the type of areas (e.g. vegetation, minerals, etc).

To understand the importance of image visibility adjustment it is presented an example in the context of the image registration. In the following is presented an image registration as shown in the work of [19]. First, image registration operation aims to help image conformity to another image. This implicates additional georeferencing, if the reference image is also rectified to a certain map projection. Landsat satellite images are first projected by employing several ground control points (GCPs) (in this example 26) to the Transverse Mercator (UTM) projection System (zone number: 33N; reference datum: WGS84). For geographical correction of the two images , the first order polynomial transformation model and nearest neighbor model for resampling (with a Root Mean Square (RMS) error of 0.5 pixels). Landsat satellite imageries of two different time periods Landsat TM 2010 and Landsat TM 1984 were geo-coded by "map-to-map" method.

Using the dark object subtraction model [8] the images are also radiometrically corrected. This method is considered one of the most robust and effective approach and employed by many implementations as shown in [19]. The algorithm results has been compared against the inter-calibration algorithm of NC regression normalization. The results obtained by the two compared methods has showed no significant differences in change detection over forest areas. Finally the images can be cross-classified and differences can be analyzed (see figure 2). For this shown

example, the main target has been to find changes over forest areas over the time periods.

On the other side, the process of haze removal can be performed on the basis of certain scene-based procedure without any additional data that describe the atmospheric conditions. In this category is included for example the tool SHARE (Scene-based Haze Removal) [20] implemented as extension of the method proposed by Zhang et al. [21] and Zhang and Guindon [22].

The method has the foundation on the Haze Optimized Transform (HOT) that aims to estimate the haze amount based of the certain distances(e.g. green band versus blue) which classify image pixels into so-called clear-sky line from points belonging to haze-free areas. However, this approach is affected mainly by man-made elements of the scene, that due to their spectral signature, have the effect to produce high HOT estimates and that influence the algorithm to over-correct these type of targets. For low-resolution satellite images, the solution found by Zhang et al. [21] and Zhang and Guindon [22] is to smoothly interpolate the estimated haze. This solution is not satisfactory for high-resolution data (e.g. IKONOS used by [20]). Their solution (as presented in figure 3 and 4) is to mask man-made features first as a preprocess step before applying the haze interpolation for these masked targets.

Preious presented techniques have demonstrated utility mostly for satellite images. However, remote sensing images tend to be collected at nadir, or near nadir so that the effect pf haze tends to be uniform across the image due to the fact that the ground is mostly nearly the same distance across the image, especially from space. For images affected by haze, acquired with common cameras, the presented solutions are not applicable.

More recently, He et al. [15] have presented a new derivation of dark object approach, called the dark channel prior. This prior information has its foundation on the statistics of haze-free outdoor images. Analyzing a large database of images, the authors have found that most of the local regions which are not part of the sky, are characterized by low intensity in at least one color(R, G,B) channel. Since some pixels (also called dark pixels) are always of low intensity, the minimum intensity of the patch that contains them have very low value. The dark channel for an image is defined mathematically:

The results of HOT(Haze Optimized Transform)-based haze subtraction algorithm

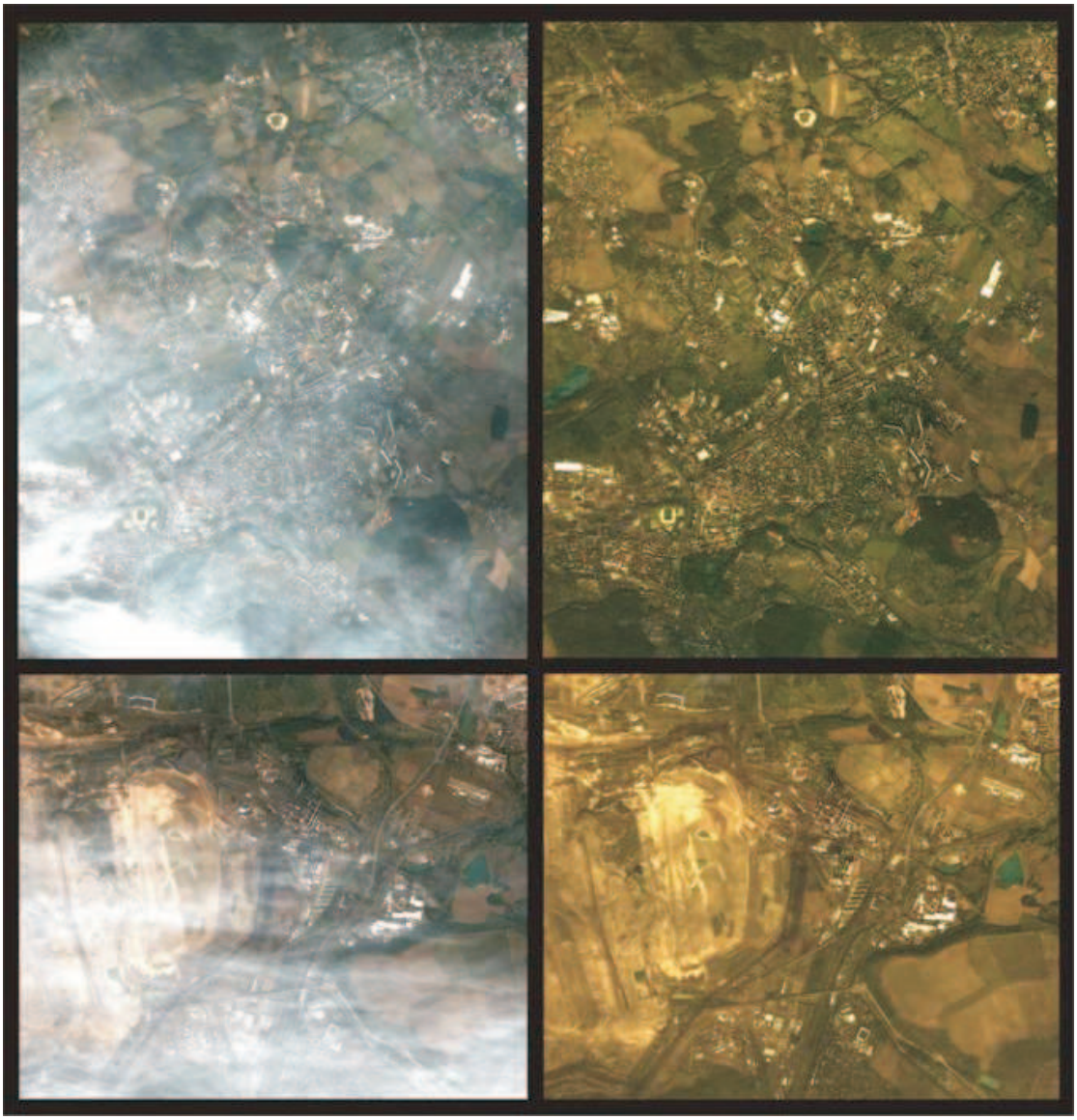

FIGURA 3: Two details of the IKONOS scene reported in the logo before (left) and after (right) the haze removal (Dark-Object Subtraction performed as well)

$$\mathcal{I}_{dark} = min_{c\in(R,G,B)}(min_{Y\in\Omega(x)}(\mathcal{I}_c(y))) \tag{5}$$

where $\mathcal{I}_c$ is the color (R, G,B) channel and $\Omega(x)$ represents a local patch centered at x. The observation of the authors is that the intensity of $\mathcal{I}_{dark}$ is low and tends to be zero, except for the sky region if the image I is actually haze-free outdoor image. The image obtained after min morphological operations are applied ($\mathcal{I}_{dark}$) is call dark channel of I.

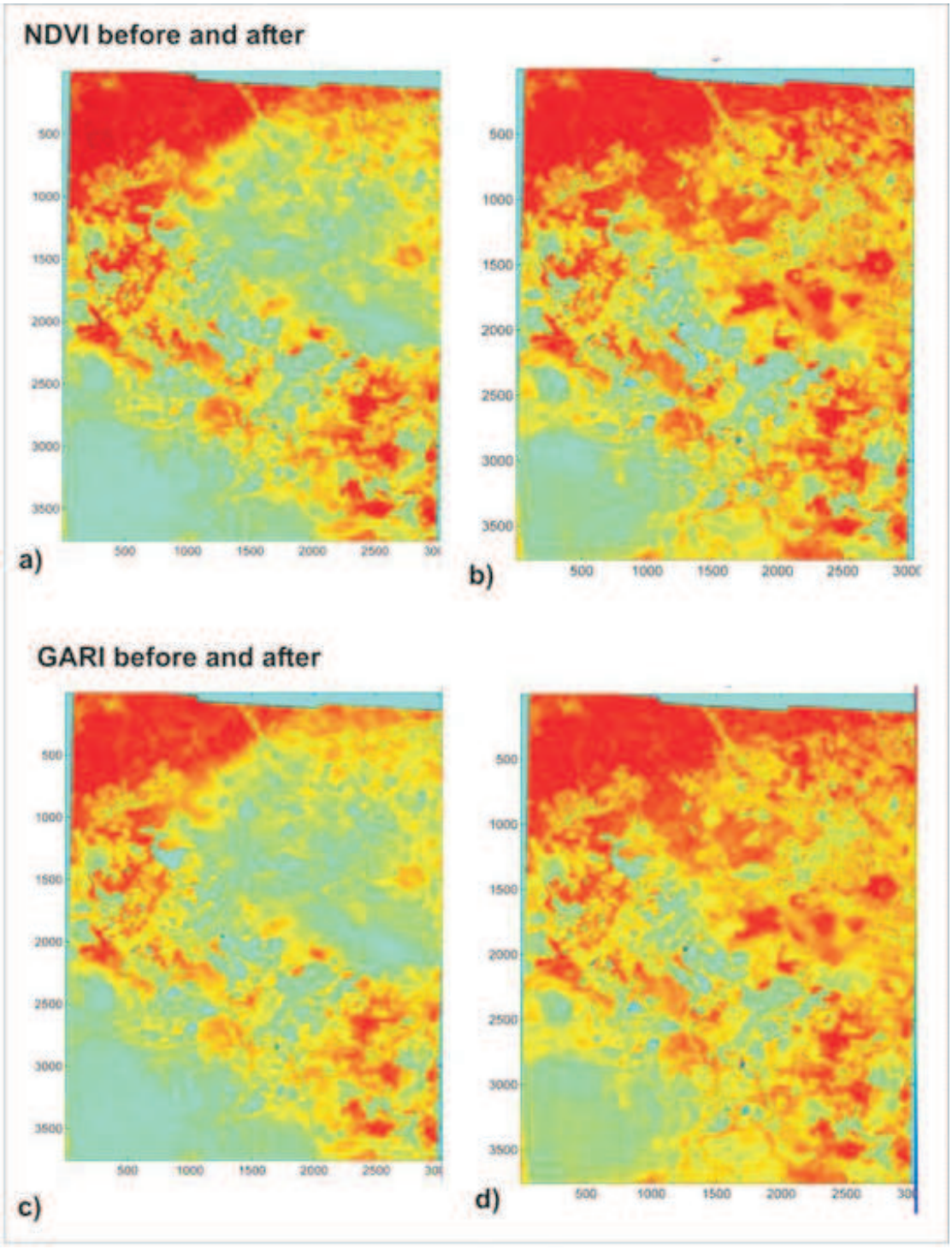

FIGURA 4: Vegetation Indices NDVI and GARI before (left column) and after (right column) the haze removal via SHARE

There are several factors that influence the intensity on the dark channel such as shadows of objects(cars, buildings, trees, rocks), colorful scene areas (e.g. green grass/plant, red/yellow/purple flowers, blue water, etc), dark objects or surfaces(e.g. trees trunk, roads and stones). To validate this assumption the authors

FIGURA 5: Dark channel example. Top: haze-free images Bottom: the corresponding dark channels.

have tested about 5000 images an manually cut out the sky regions. The images has been resized to a match a maximum withxheight of 500 pixels. For all of them the dark channel has been computed by applying the formula 5.

In the hazy images, the intensity of dark pixels in the dark channel are mainly contributed by the airlight. In other words, the dark channel computed for an haze image will have higher intensity in regions that are characterized by denser haze. Roughly, visually this dark channel is similar with the transmission that estimate the haze density over the scene (or the depth map).

A disadvantage of dark channel method is its inability to properly preserve edges, which is caused mainly by the employed erosion filter during the stage of computing the dark channel. In order to recover the refined transmission map and the latent image, this patch-based approach requires a complex postprocessing stage. By employing the dark channel prior [15], it has been shown that each patch of a natural image contains at least one point that is *dark* for non-sky or haze-free regions. The validity of this observation is mainly motivated by the fact that natural images are colorful and full of shadows [15]). On the dark channel, patches representing sky and hazy regions contain high values, as the local minimal intensity of such patches is high. Similarly, it has been observed that pixels in sky or hazy areas have high values in all color channels. These observations confirm

FIGURA 6: From left to right: input haze image, estimated rough transmission map, refined transmission map after soft matting and final haze-free image.

the assumption that values in hazy image patches vary smoothly, except at depth discontinuities.

The dark channel method built on depth map statistics is physically valid and is capable to deal with distant objects. To avoid hallo artifacts, the solution is to combine the extracted haze imaging model and a soft matting interpolation method. Thus the authors has demonstrated the possibility to recover both a hi-quality haze-free image and also to produce a good depth map (up to a scale).

Another similar approach is to maximize the image contrast. The main difference between physical based methods and contrast based methods is that the second cathegory does not require geometrical information(or depth map) in order to restore the lost details. The method of Tan [14] also propose a strategy on a similar assumption as the dark channel [15]. First, the haze-free images have more contrast and second, the airlight variation mainly depends on the distance. The method is built on a framework of Markov random fields. This approach can be efficiently optimized by choosing different techniques, such as graph-cuts or belief propagation.

This method aims to employ the chromacity information and assumes that if the object is infinitely distant, then the image chromaticity is straightforward dependent only by the atmospheric light (airlight). In addition, the authors assumed that if there is no effect of scattering particles, it can be translated mathematically, that in the absence of the airlight then the image chromaticity is constrained solely by the direct attenuation.

The algorithms requires three mains constraints to be valid: first, the dehaze image (or the haze-free image) is assumed to have higher contrast compared with the input hazy image. Second, the variation of the airlight value A depends mostly by the scene geometry, which means that objects at the same depth will record the same values of airlight A, regardless of their reflectance. As a consequence,

the airlight value for most neighboring pixels tend to be similar. This rule is not applicable to depth continuities which can impact dramatically the the airlight influence, although the number of pixels is assumed relative small. Finally, hazy images characterized mostly outdoor natural scenes. Due to this fact, the correct estimate of transmission (scene geometry) should follow the characteristics of clear-day natural images.

Summarizing the algorithm procedures: it crops from the image an nxn (7x7 or more since it depends by the size and scale of the input image) patch centered at a location x. As shown in the figure 7 the strategy is to find that airlight value that corresponds to the patch in order to maximize the edges contrast. After all of the iterations finish, the algorithm will estimate a data cost for all pixels (smoothness term), where for each pixel should find the inference, which finally produces the airlight distribution. This algorithmic part provides a data cost and a smoothness cost that supplies a complete graph in term of Markov random fields.

Currently, graph-cuts and belief propagation are considered by the authors the most efficient techniques to optimize the cost function of MRFs. The main drawbacks are represented by the fact that they still require a considerable amount of computational time (especially for high-quality input images) and when the number of labels are large. In order to reduce the complexity two strategies has been considered: first, the number of labels has been reduced (that means also the dimension of the data cost is reduced), and second, some initial estimates values for the airlight are known.

The results produced by the authors are obtained by using Iterated Conditional Modes (ICM) which uses the blurred luminance image as the initial values. However, ICM does not guarantee the global optimum, but the obtained results demonstrated that the contrast has been considerably improved. The main problems shown by the results are mainly related by the halos at depth discontinuities.

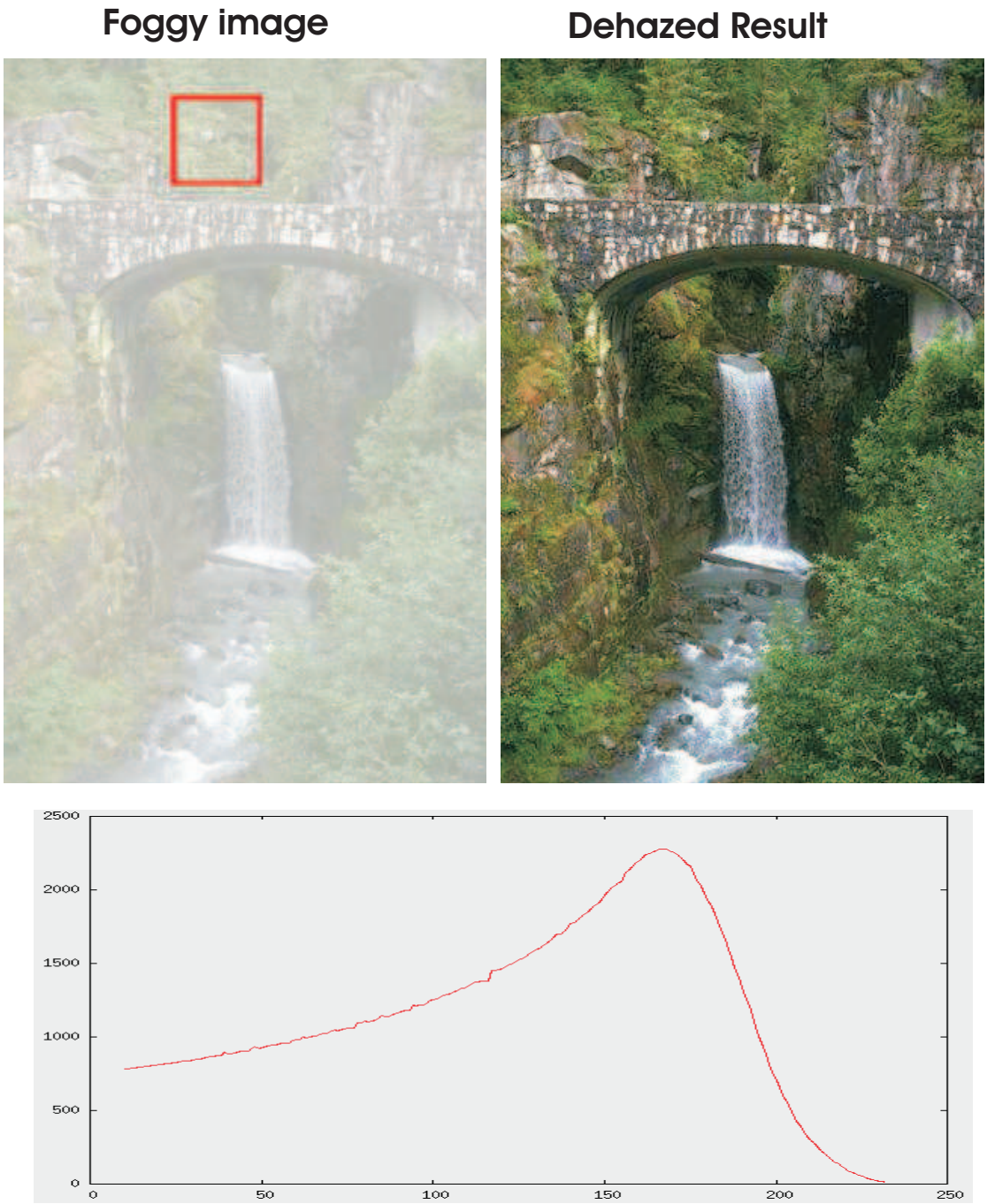

FIGURA 7: Top Left: synthetic fog applied on the right image, where A is set constant globally (=153 after the division). Right: the haze-free natural image. Bottom: Edges distribution for the region in the red rectangle with respect to A. Please notice that the peak is around A = 167 (y-axis characterizes the contrast estimated for the edges, while the x-axis is airlight values)

II.1.3 Daytime Image Dehazing by Fusion

In this section we describe an alternative single-image based strategy that is able to accurately dehaze images using only the original degraded information. The technique has been published recently by the candidate in [1]. Our technique has some similarities with the previous approaches of Tan [14] and Tarel and Hautière [17], which enhance the visibility in such outdoor images by manipulating their contrast.

However, in contrast to existing techniques, we built our approach on a fusion strategy. We are the first to demonstrate the utility and effectiveness of a fusion-based technique for dehazing on a single degraded image. Image fusion is a well studied process [23], that aims to blend seamlessly several input images by preserving only the specific features of the composite output image. In this work, our goal is to develop a simple and fast technique and therefore, as will be shown, all the fusion processing steps are designed in order to support these important features. The main concept behind our fusion based technique is that we derive two input images from the original input with the aim of recovering the visibility for each region of the scene in at least one of them. Additionally, the fusion enhancement technique estimates for each pixel the desirable perceptual based qualities (called weight maps) that controls the contribution of each input to the final result. In order to derive the images that fulfill the visibility assumptions (good visibility for each region in at least one of the inputs) required for the fusion process, we analyze the optical model for this type of degradation. There are two major problems, the first one is the color cast that is introduced due to the airlight influence and the second is the lack of visibility into distant regions due to scattering and attenuation phenomena.

The first derived input ensures a natural rendition of the output, by eliminating chromatic casts that are caused by the airlight color, while the contrast enhancement step yields a better global visibility, but mainly in the hazy regions. However, by employing these two operations, the derived inputs taken individually still suffer from poor visibility (e.g. analyzing figure 54 it can be easily observed that the second input restores the contrast of the hazy inputs, but at the cost of altering the initial visibility of the closer/haze-free regions).

Therefore, to blend effectively the information of the derived inputs, we filter (in a per-pixel fashion) their important features, by computing several measures (weight

maps). Consequently, in our fusion framework the derived inputs are weighted by three normalized weight maps (luminance, chromatic and saliency) that aim to preserve the regions with good visibility.

Finally, to minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion, using a Laplacian pyramid representation of the inputs combined with Gaussian pyramids of normalized weights.

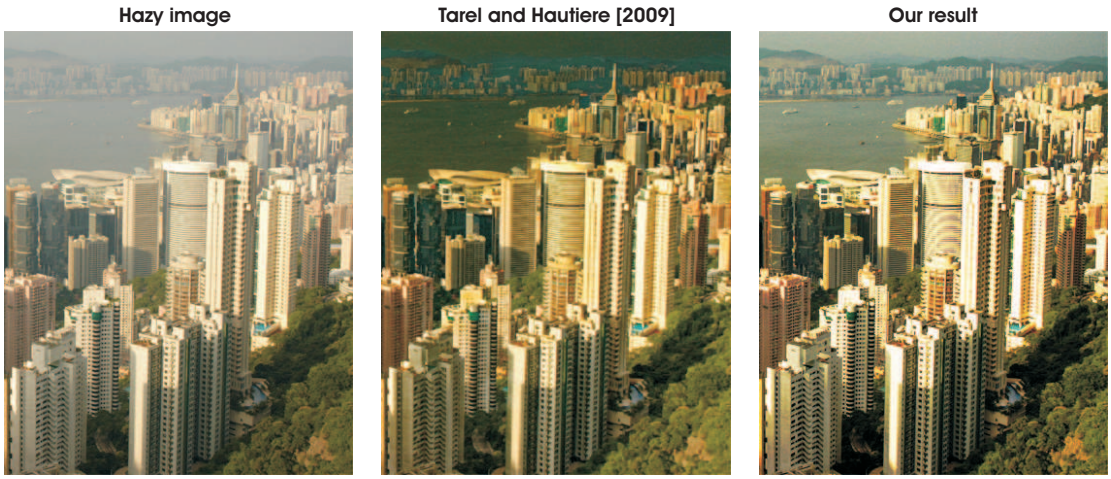

FIGURA 8: Comparison with the fast method of Tarel and Hautière [17]. Our method performs faster and yields more visually plausible results than [17] Notice the sky and sea regions.

Our technique has several advantages over previous single-image dehazing methods. First, our approach performs an effective per-pixel computation, different from the majority of the previous methods [13–15] that process patches. A proper per-pixel strategy reduces the amount of artifacts, since patch-based methods have some limitations due to the assumption of constant airlight in every patch. In general, the assumptions made by patch-based techniques do not hold, and therefore additional post processing steps are required (e.g. the method of He et al. [15] needs to smooth the transmission map by alpha-matting). Secondly, since we do not estimate the depth (transmission) map, the complexity of our approach is lower than most of the previous strategies. Finally, our technique performs faster which makes it suitable for real-time applications. Even compared with the recent effective implementation of Tarel and Hautière [17] our technique is able to restore a hazy image in less time, while showing more visually plausible results in terms of colors and details (see figure 8).

Our technique has been tested extensively for a large set of different hazy images (the reader is referred to the evaluation section). Results on a variety of hazy

images demonstrate the effectiveness of our fusion-based technique. Moreover, we perform a quantitative experimental evaluation based on the measure of Hautière et al. [24]. The main conclusion is that our approach is less prone to artifacts, yielding very similar results with the physically-based techniques such as the technique of Fattal [13], He et all. [15], Nishino et al. [25] and Kopf et al. [11]. We believe that this is a key advantage of our technique.

The rest of this section is structured as follows. In the next section, the related techniques that deal with haze removal are briefly reviewed. Next, we discuss some theoretical aspects of light propagation in such environments and present our single image based dehazing technique; the details regarding our fusion technique are discussed in this section. Finally, we report and discuss the results and present a quantitative evaluation.

Related Work. Enhancing images represents a fundamental task in many image processing and vision applications. As a particular challenging case, restoring hazy images requires specific strategies and therefore an important variety of methods have emerged to solve this problem.

Firstly, several dehazing techniques have been developed for remote sensing systems, where the input information is given by a multi-spectral imaging sensor installed on the Landsat satellites. The recorded six-bands of reflected light are processed by different strategies in order to yield enhanced output images. The well-known method of Chavez [8] is suitable for homogeneous scenes, removing the haze by subtracting an offset value determined by the intensity distribution of the darkest object. Zhang et al. [26] introduced the haze optimized transformation (HOT), using the blue and red bands for haze detection, that have been shown to be more sensitive to such effects. Moro and Halounova [9] generalized the dark-object subtraction approach [8] for highly spatially-variable haze conditions.

A second category of methods, employs multiples images or supplemental equipment. In practice, these techniques use several input images taken in different atmospheric conditions. Different medium properties may give important information about the hazy image regions. Such methods [27], [28], [10] produce pleasing results, but their main drawback is due to their acquisition step that in many cases is time consuming and hard to carry out.

Different strategies have been developed when the approximated 3D geometrical model of the scene is given. The forerunner method of Narasimhan and Nayar [29]

employs an approximated depth-map specified interactively by the users. Hautière et al. [30] designed a method for vehicle vision systems, where weather conditions are first estimated and then used to restore the contrast according to a scene structure which is inferred a priori. The *Deep Photo* [11] system uses the existing georeferenced digital terrain and urban models to restore foggy images. The depth information is obtained by iteratively aligning the 3D models with the outdoor images.

Another class of techniques exploits the properties of the airlight that is partially polarized [31], [32], [12], [33], [34]. By using different angles of polarized filters the resulting images of the same scene can be processed to estimate the haze effects. The difference between such images enables the estimation of the magnitude of the polarized haze light component. These methods have shown less robustness for scenes with dense haze where the polarization light is not the major degradation factor.

However, a more difficult case is when only a single hazy image is used as an input information. The single image dehazing is an ill-posed problem that can be solved by different strategies [13], [14], [15], [17], [25] that have been introduced only recently. Roughly, these methods can be divided into contrast-based and statistical approaches. Tan's [14] method belongs to the first category. In this case the image restoration maximizes the local contrast while constraining the image intensity to be less than the global atmospheric light value. The contrast-based enhancing approach of Tarel and Hautière [17] has shown to be a computationally effective technique, but assumes as well that the depth-map must be smooth except along edges with large depth jumps. Regarding to the second category, the technique of Fattal [13] employs a graphical model that solves the ambiguity of airlight color. It assumes that image shading and scene transmission are locally uncorrelated. He et al. [15] built their approach on the statistical observation of the dark channel [8], that allows a rough estimation of the transmission map. To refine the final depth-map, the transmission map values are extrapolated into the unknown regions, by a relatively computationally expensive matting strategy [35]. The technique of Kratz and Nishino [16], recently extended [25], is a Bayesian probabilistic method that jointly estimates the scene albedo and depth from a single degraded image by fully leveraging their latent statistical structures. Their approach models the image with a factorial Markov random field in which the scene albedo and depth are two statistically independent latent layers, which are estimated jointly.

Our method is also a single image dehazing technique. Different than previous single image dehazing approaches, our technique is built on the principle of image fusion, a well-studied topic of computational imaging that has found many useful applications such as interactive photomontage [36], image editing [37], image compositing [38, 39] and HDR imaging [40, 41]. The main idea is to combine several images into a single one, retaining only the most significant features. Even though the fusion principle has been used previously to restore hazy images, but using additionally near-infrared (NIR) image of the same scene [42], we are the first that introduce a single image dehazing technique based on the fusion principle that blends only the information existing in the input image. Our strategy bears some similarity with the recent methods of He et al. [15] and Tarel and Hautière [17]. Both of these methods can be seen as filtering solutions since the dark channel [15] can be related with an erosion problem, while [17] employed their defined *median of median filter* in order to preserve both edges and corners. However, our approach is fundamentally different since it removes the haze by simply blending the two derived inputs weighted by several measures. Our strategy combines the input information in a per-pixel fashion minimizing the loss of the image structure by a multi-scale strategy. While no post-processing steps are required it is also straightforward to implement and computationally effective.

Background Theory: Light Propagation. Due to the absorption and scattering, the light passing through the atmosphere is attenuated and dispersed. While in normal conditions (clear day) the size of air molecules is relatively small compared with the wavelength of visible light, the scattering influence might be considered insignificant. As the title presents, in this paper we are dealing with specific atmosphere conditions due to the presence of haze. As discussed in the study of McCartney [43], haze is traditionally an atmospheric phenomenon where dust, smoke and other dry particles obscure the clarity of the sky. Haze reduces visibility for distant regions by yielding a distinctive gray hue in the captured images. As will be demonstrated, our algorithm is able to deal as well with a particular case: foggy scenes. Fog is a dense cloud of water droplets, or cloud, that is close to the ground. In general, this phenomenon appears when night conditions are clear but cold, and the heat released by the ground is absorbed during the day. As the temperature of the ground decreases, it cools the air above it to the dew point forming a cloud of water droplets known as radiation fog.

Based on the Koschmieder's [18] law only a percentage of the reflected light reaches

the observer causing poor visibility in such degraded scenes. The light intensity $\mathcal{I}$ for each pixel x, that reaches the observer is described by two main additive components: *direct attenuation* and *veiling light*, also known as *airlight*:

$$\mathcal{I}(x) = \mathcal{J}(x)\ T(x) + V_\infty\ (1 - T(x)) \tag{6}$$

where $\mathcal{J}$ is the scene radiance or haze-free image, T is the *transmission* along the cone of vision and V_∞ is the *veiling* color constant. The optical model assumes linear correlation between the reflected light and the distance between the object and observer. The first component, *direct attenuation* $\mathcal{D}$, represents how the scene radiance is attenuated due to medium properties: $\mathcal{D}(x) = \mathcal{J}(x)\ T(x)$. The *veiling light* component $\mathcal{V}$ is the main cause of the color shifting, being expressed as:

$$\mathcal{V}(x) = V_\infty\ (1 - T(x)) \tag{7}$$

The value of T depicts the amount of light that has been transmitted between the observer and the surface. Assuming a homogeneous medium, the transmission T is determined as $T(x) = e^{(-\beta\ d(x))}$ with β being the medium attenuation coefficient due to the scattering, while d represents the distance between the the observer and the considered surface. Practically, the problem is to estimate from the hazy input $\mathcal{I}$ the latent image $\mathcal{J}$ when no additional information about depth and airlight are given.

Fusion-based Dehazing. In this part is presented in detail our fusion technique that employs only the inputs and weights derived from the original hazy image. The fundamental idea is to combine several input images (guided by the weights maps) into a single one, by keeping only the most significant features of them . Obviously, the choice of inputs and weights is application-dependent. By processing appropriate weight maps and inputs, we demonstrate that our fusion-based method is able to effectively dehaze images.

Inputs. As mentioned previously, the input generation process seeks to recover optimal region visibility in at least one of the images. In practice, there is no enhancing approach that is able to remove entirely the haze effects of such degraded inputs. Therefore, considering the constraints stated before, since we process only one captured image of the scene, the algorithm generates from the original image only two inputs that recover color and visibility of the entire image. The first

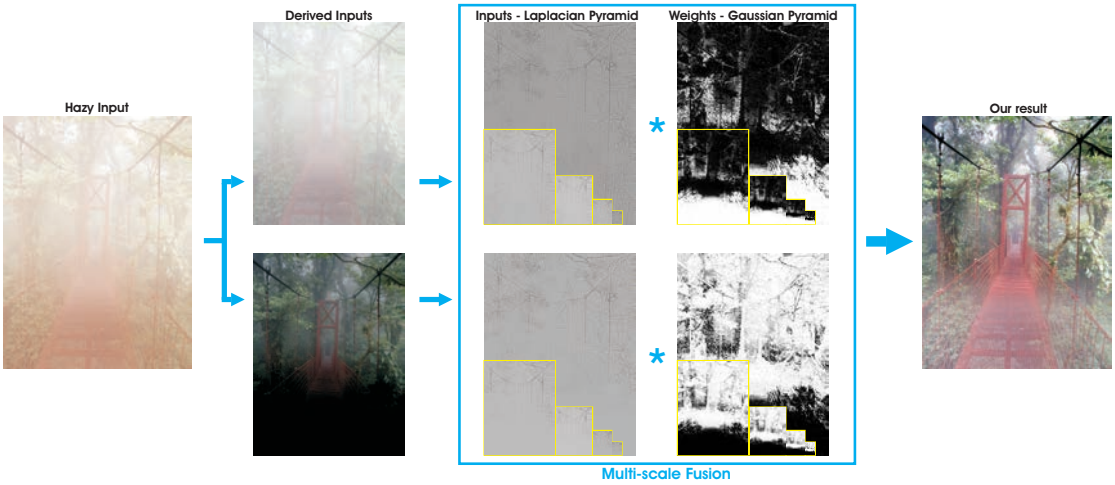

FIGURA 9: Overview of our technique. From the input hazy image are derived two enhanced versions. These two derived inputs are weighted by three normalized weight maps (luminance, chromatic and saliency) - here we show only the Gaussian of corresponding normalized maps. Finally, the Laplacian of the inputs and Gaussian of the weights are blended in a multi-scale fashion that avoids introducing artifacts. In this outline we depict our approach by using only 5 (l=5) scale levels in the Laplacian and Gaussian spaces.

one better depicts the haze-free regions while the second derived input increases visible details of the hazy regions.

Inherently inspired by the previous dehazing approaches such as Tan [14], Tarel and Hautière [17] and He et al. [15], we searched for a robust technique that will properly white balance the original image.

Our **first input** $\mathcal{I}_1$ is obtained by white balancing the original hazy image. By this step we aim a natural rendition of images, by eliminating chromatic casts that are caused by the atmospheric color.

In the last decades many white balancing approaches [44–48] have been proposed in the literature (a systematic overview of the existing methods is presented in [49]). Several specialized techniques have been experimented in the context of our problem. Since we aim for a computationally effective dehazing approach, we opted for the *shades-of-gray* color constancy technique [47]. Despite of its simplicity, this low-level approach of Finlayson and Trezzi [47] has shown to yield comparable results to those of more complex white balance algorithms (that produces reliable results based on natural image statistics [50]). The main objective of white balance algorithms, is to identify the illuminant color $e(\lambda)$ or its projection on the RGB color channels (R_e, G_e, B_e) . We use the same notations as in the original manuscript of Weijer and Gevers [51] of *grey-edges* that is just an extension

of *shades-of-gray*. Unfortunately, we observed that the *grey-edges* technique often failed when processing hazy images. We believe that this is due to the lack of local contrast in the hazy regions but also since the hazy images are characterized by a smaller amount of edges.

Given an image f, for a Lambertian surface, the intensity measured can be modeled as:

$$f(x) = \int_{\omega} e(\lambda)s(\lambda, x)c(\lambda)d\lambda \tag{8}$$

where $e(\lambda)$ is the radiance given by the light source, λ is the wavelength, $s(\lambda, x)$ denotes the surface reflectance, $c(\lambda) = [R(\lambda), G(\lambda), B(\lambda)]$ describes the sensitivity of the sensors while ω is the visible spectrum.

The illuminant $\mathbf{e}$ to be estimated, is expressed as:

$$\mathbf{e} = (R_e, G_e, B_e) = \int_{\omega} e(\lambda)c(\lambda)d\lambda \tag{9}$$

According to the *Grey-World* assumption of Buchsbaum [44] the average reflectance of the scene is achromatic (gray). This hypothesis is mathematically defined as follows:

$$\frac{\int s(\lambda, x)dx}{\int dx} = \kappa \tag{10}$$

where κ is the constant assumed to have the value 0.5.

Next, by replacing s in equation 8 with 10, the following expression is obtained:

$$\frac{\int f(x)dx}{\int dx} = \frac{1}{\int dx}\int\int_{\omega} e(\lambda)s(\lambda, x)c(\lambda)d\lambda dx \quad \Longleftrightarrow \quad \frac{\int f(x)dx}{\int dx} = \kappa\int_{\omega} e(\lambda)c(\lambda)d\lambda \tag{11}$$

As shown in *shades-of-gray* [47] and *grey-edges* [51], white balance can be defined based on Minkowski norm of images. *Grey-World* [44] algorithm estimates the illumination, by stating that the average color of the entire image raised to a power n is achromatic (gray).

$$\left[\frac{\int f^n dx}{\int dx}\right]^{\frac{1}{n}} = \kappa\mathbf{e} = \kappa\left(R_e, G_e, B_e\right) \tag{12}$$

As demonstrated, *shades-of-gray* is part of the same family with *Grey World* [44] and *max-RGB* [45] techniques. This is easy to demonstrate by setting norm $n = 1$ and $n \to \infty$ respectively in the equation 12. For *shades-of-gray* approach n can take any number between 1 and ∞ (default value is set to $n = 6$). When $n = 1$, all the components from the scene contribute uniformly to the average. When n is increased, the impact of the components is directly proportional with their intensity.

Since the first input I_1 shows good visibility in non-hazy regions and discards the color shifting, the second input I_2 is generated from this I_1 in order to enhance the contrast in those regions that suffer due to the airlight influence.

For the **second input** we searched for a relatively complementary processing technique, capable to enhance those regions that present low contrast. Considering the airlight factor from the optical model (eq. 7)(that is both additive and multiplicative with the transmission), and since the haze is dominant in the hazy images, it is expected that the hazy regions would have a great influence over the average of the image. Moreover, due to the fact that the airlight influence increases linearly with the distance, the luminance of these regions is assumed to amplify with the distance. In practice, based on these observations, second input I_2 is obtained automatically by subtracting the average luminance value of the entire image I from the original image I. This operation has the effect of amplifying the visibility in regions degraded by haze, but yields some degradation in the rest of the image (the effect of this conversion is shown in figure 54). A similar effect may be obtained by general contrast enhancing operators (e.g. gamma correction, histogram stretching) that also amplify the visibility in the hazy parts, while destroying the details in the rest of the image. However, this degradation is detected and solved in our fusion approach by defining proper weight maps (please refer to the next subsection and figure 54).

Mathematically, the second input computed for each pixel x is obtained by applying the following expression:

$$\mathcal{I}_2(x) = \gamma \ (\mathcal{I}(x) - \bar{\mathcal{I}}) \tag{13}$$

where γ is a factor that increases linearly the luminance in the recovered hazy regions (default value is $\gamma = 2.5$). This default value for γ matches for those most cases. However, in our experiments there are few exceptions that are not satisfied. These cases are characterized by the fact that hazy regions are relatively darker

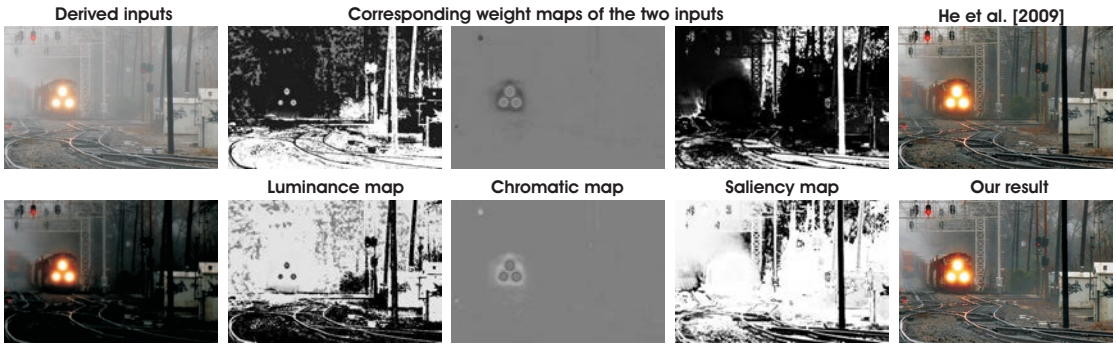

FIGURA 10: Derived inputs and weight maps. In the left side the two derived inputs are shown. In the middle are displayed the three weight maps corresponding to the inputs. Finally, in the right column the results of He et al. [15] and our result are shown.

than non-hazy regions (see figure 14). Since fusion blends selectively (guided by the weight maps) the two inputs, the results of the dehazing operation will suffer also by darker aspect (some algorithms suffer from such problems as can be noticed by a closed inspection of figure 15 of [14], in figure 8 of [17] or in figure 13 of [25]). To generate optimal results for these cases, we derived a general expression for *gamma* ($\gamma = 2(0.5 + \bar{\mathcal{I}})$), that is correlated with the average luminance of the image (we employ this information, since $\bar{\mathcal{I}}$ is a good indicator of the image brightness appearance). The parameter *gamma* has a similar impact as the tone mapping stage of [17] that has been applied on the the haze-free regions, assumed to be in the bottom third part of the original image.

Weight Maps. As can be seen in figures 9 and 54, by applying only these enhancing operations, the derived inputs still suffer from low visibility mainly in those regions with dense haze and low light conditions. The idea that global contrast enhancement techniques are limited to dealing with hazy scenes has been remarked previously by Fattal [13]. This is due to the fact that the optical density of haze varies across the image and affects the values differently at each pixel. Practically, the limitation of the general contrast enhancement operators (e.g. gamma correction, histogram equalization, white balance) is due to the fact that these techniques perform (constantly) the same operation across the entire image. In order to overcome this limitation, we introduce three measures (weight maps). These maps are designed in a per-pixel fashion to better define the spatial relations of degraded regions.

Our weight maps balance the contribution of each input and ensure that regions with high contrast or more saliency from a derived input, receive higher values.

The ***luminance weight map*** measures the visibility of each pixel and assigns high values to regions with good visibility and small values to the rest. Since hazy images present low saturation, an effective way to measure this property is to evaluate the loss of colorfulness. This weight is processed based on the RGB color channel information. We make use of the well known property, that more saturated colors yield higher values in one or two of the color channels. This weight map is simply computed (for each input I_k, with k indexes the derived inputs) as the deviation (for every pixel location) between the R,G and B color channels and the luminance L from the input:

$$\mathcal{W}_L^k(x) = \sqrt{1/3\left[(R^k(x) - L^k(x))^2 + (G^k(x) - L^k(x))^2 + (B^k(x) - L^k(x))^2\right]} \quad (14)$$

Since the luminance L is computed by averaging the RGB channels, this disparity yields higher values for the saturated pixels which are assumed to be part of the initial haze-free regions. On the other hand, because haze produces colorlessness and low contrast, this measure will assign small values (reducing the contribution of these locations to the output) for the hazy but also for the deteriorated regions (e.g. in the second derived input we refer to the regions that have lost their luminance and therefore have a dark appearance). As illustrated in figure 54, the $\mathcal{W}_L$ map is a straightforward, yet effective identification of such regions.

The luminance weight acts as an identifier of the degradation induced in $\mathcal{I}_2$ in the haze-free regions, ensuring a seamless transition between the derived inputs $\mathcal{I}_1$, $\mathcal{I}_2$. On the other hand this map also tends to reduce the global contrast and colorfulness. To overcome these effects, in our fusion framework we define two additional weight maps: a chromatic map (colorfulness) and a saliency map (global contrast).

The ***chromatic weight map*** controls the saturation gain in the output image. This weight map is motivated by the fact that in general humans prefer images characterized by a high level of saturation. Since the color is an inherent indicator of the image quality, often similar color enhancement strategies are also performed in tone mapping.

FIGURA 11: Impact of each weight map to the final result. Results of our fusion technique but using only one weight map at a time.

To obtain this map, for each pixel the distance between its saturation value S and the maximum of the saturation range is computed as following:

$$\mathcal{W}_C^k(x) = exp\left(-\frac{(S^k(x) - S_{max}^k)^2}{2\sigma^2}\right) \tag{15}$$

where k indexes the derived inputs, the default value of the standard deviation is $\sigma = 0.3$ and S_{max} is a constant that depends by the color space employed (in our approach we opted for the HSI color space and higher saturated pixels correspond to $S_{max} = 1$). Therefore, small values are assigned to pixels with reduced saturation while the most saturated pixels get high values. As a result, this map ensures that the initial saturated regions will be better depicted in the final result.

The ***saliency weight map*** identifies the degree of conspicuousness with respect to the neighborhood regions. This perceptual quality measure assesses that a certain object/person stands out from the rest of the image, or from nearby regions. In general, saliency (also referred to as *visual attention* [52, 53]) no matter what the motivation behind it (biologically based [54, 55], computational [53] or combination of both), seeks to estimate the contrast of image regions relative to their surroundings (based on different image features such as intensity, color or orientation).

For this measure, we use the recent saliency algorithm of Achanta et al. [55]. This strategy is inspired by the biological concept of center-surround contrast. The saliency weight at pixel position (x, y) of input I^k is defined as:

$$\mathcal{W}_S^k(x) = \|I_k^{\omega_{hc}}(x) - I_k^{\mu}\| \tag{16}$$

where I_k^{μ} represents the arithmetic mean pixel value of the input I_k (a constant value during the entire process that is computed only once) while $I_k^{\omega_{hc}}$ is the blurred version of the same input that aims to remove high frequency such as noise.

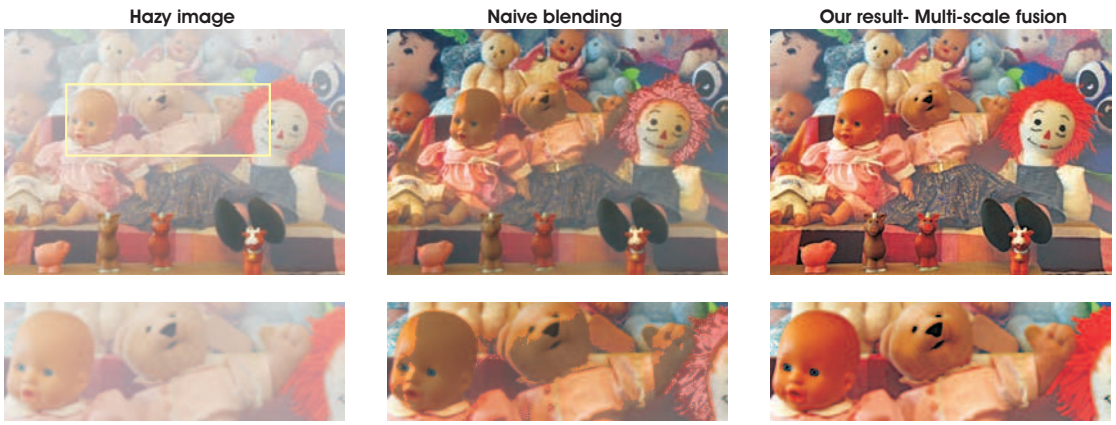

FIGURA 12: The naive blending that directly implements equation 52 introduces halo artifacts, most apparent in locations characterized by strong transitions of the weight maps.

$I_k^{\omega_{hc}}$ is obtained by employing a small 5×5 ($\frac{1}{16}[1, 4, 6, 4, 1]$) separable binomial kernel with the high frequency cut-off value $\omega_{hc} = \pi/2.75$. For small kernels, the binomial kernel is a good approximation of its Gaussian counterpart, but it has the advantage that in this way it can be computed more efficiently. Once the blurred version of the image $I_k^{\omega_{hc}}$ and the arithmetic mean I_k^{μ} are computed, the saliency is obtained in a per pixel fashion. We opted for the approach of Achanta et al. [55] since this technique is computationally efficient. Moreover, it is able to produce maps with well-defined boundaries and uniformly highlighted salient regions (even at high resolution scales). These features of this saliency map, prevent introducing unwanted artifacts in the result image yielded by our fusion technique since neighboring comparable values are assigned similarly on the saliency map. Additionally, the employed map [55] emphasizes large regions and estimates uniform values for the whole salient regions. As a result, the effect of this gain is to enhance the global and local contrast appearance (that corresponds to large scale). As can be observed in figure 11, one of the main effects of this measure is to increase the contrast in highlighted and shadowed regions.

By processing a large and diverse set of degraded images, we observed that the impact of these three measures is, in general, equally important. However, the first measure has the highest impact on the visibility. To better understand the contribution of each of these weight maps, the reader is referred to figure 11 that displays a result processed with our fusion technique, but using only one weight map at a time.

The resulted weights $\mathcal{W}^k$ are obtained by multiplying the processed weight maps

$\mathcal{W}_L^k$, $\mathcal{W}_C^k$, $\mathcal{W}_S^k$. To yield consistent results, we normalize the resulted weight maps ($\bar{\mathcal{W}}^k(x) = \mathcal{W}^k(x) / \sum_k \mathcal{W}^k(x)$). This operation constrains that sum at each pixel location x of the normalized weight maps to equal one.

Multi-scale Fusion. In the fusion process, the inputs are weighted by specific computed maps in order to conserve the most significant detected features. Each pixel x of the output $\mathcal{F}$ is computed by summing the inputs $\mathcal{I}_k$ weighted by corresponding normalized weight maps $\bar{\mathcal{W}}^k$:

$$\mathcal{F}(x) = \sum_k \bar{\mathcal{W}}^k(x) \mathcal{I}_k(x) \tag{17}$$

where $\mathcal{I}_k$ symbolizes the input (k is the index of the inputs) that is weighted by the normalized weight maps $\bar{W}^k$. The normalization of the weights ensures that the intensity scale of the result is maintained in relatively the same scale as the inputs (since the sum of each pixel equals 1, $\sum_k \bar{\mathcal{W}}^k(x) = 1$). The naive solution (please refer to figure 55) that directly implements this equation, introduces strong halos artifacts, mostly in the locations characterized by strong transitions of the weight maps. To prevent such degradation problems, we have opted for the adapted solution that employs a classical multi-scale pyramidal refinement strategy [56]. We also tested several more recent edge preserving techniques (e.g. WLS [57]) but we did not obtain significant improvement. However, recent advanced methods need, in general, to tweak their parameters, as well as being more computationally intensive. In our case, each input $\mathcal{I}_k$, is decomposed into a pyramid by applying Laplacian operator at different scales. Similarly, for each normalized weight map $\bar{\mathcal{W}}^k$, a Gaussian pyramid is computed. Considering that both the Gaussian and Laplacian pyramids have the same number of levels, the mixing between the Laplacian inputs and Gaussian normalized weights is performed at each level independently, yielding the fused pyramid:

$$\mathcal{F}_l(x) = \sum_k G_l \left\{\bar{\mathcal{W}}^k(x)\right\} L_l \left\{\mathcal{I}_k(x)\right\} \tag{18}$$

where l represents the number of the pyramid levels (default value of the number of levels is l=5) and $L\{\mathcal{I}\}$ is the Laplacian version of the input $\mathcal{I}$ while $G\left\{\bar{\mathcal{W}}\right\}$ represents the Gaussian version of the normalized weight map of the $\bar{\mathcal{W}}$.

This step is performed successively for each pyramid layer, in a bottom-up manner. The final haze-free image $\mathcal{J}$ is obtained by summing the contribution of the resulting inputs (levels of pyramid):

$$\mathcal{J}(x) = \sum_{l} \mathcal{F}_l(x) \uparrow^d \tag{19}$$

where $\uparrow^d$ is the upsampling operator with factor $d = 2^{l-1}$. As a default characteristic, in our implementation the contribution of all the three weight maps is equally distributed.

Results and Discussion. To prove the robustness of our method, the new operator has been tested on a large dataset of different natural hazy images. Haze due to dust, smoke and other dry particles reduces visibility for distant regions by causing a distinctive gray hue in the captured images. However, our technique has been successfully tested as well for a slightly different case: foggy scenes (e.g. the first and the third example in figure 13). For our problem, fog has a similar impact as haze, but technically it appears as a dense cloud of water droplets close to the ground when night conditions are clear but cold, and the heat released by the ground is absorbed during the day (please refer to figure 21). We assume that the input hazy/foggy images are color images and the images may contain achromatic objects.

As can be seen in figure 13, but also in figures 15 and 16, our operator is able to yield comparable and even better results with more complex techniques. Compared with the techniques of Tan [14] and Tarel and Hautière [17] our technique is able to better preserve the fine transitions in the hazy regions without introducing unpleasing artifacts. Moreover, the technique of Tan [14] produces results with over-saturated colors. The recent technique of Kratz and Nishino [16] (recently extended in [25]) yields in general aesthetically pleasing results, but may introduce some artifacts in those regions considered to be at infinite depth (e.g. the skyline of the pumpkins field). Similarly, the technique of Fattal [13] performs well, but shows some limitations for situations with dense haze. This is mainly due to the fact that the method of Fattal [13] is basically a statistic interpretation that requires variance to estimate the depth map. On the other hand, our technique yields visually similar results with the technique of He et al. [15]. However, by a closer analysis of the results, a difference can be observed between how colors are restored by the two methods, especially for the distant regions (e.g. in figure 16

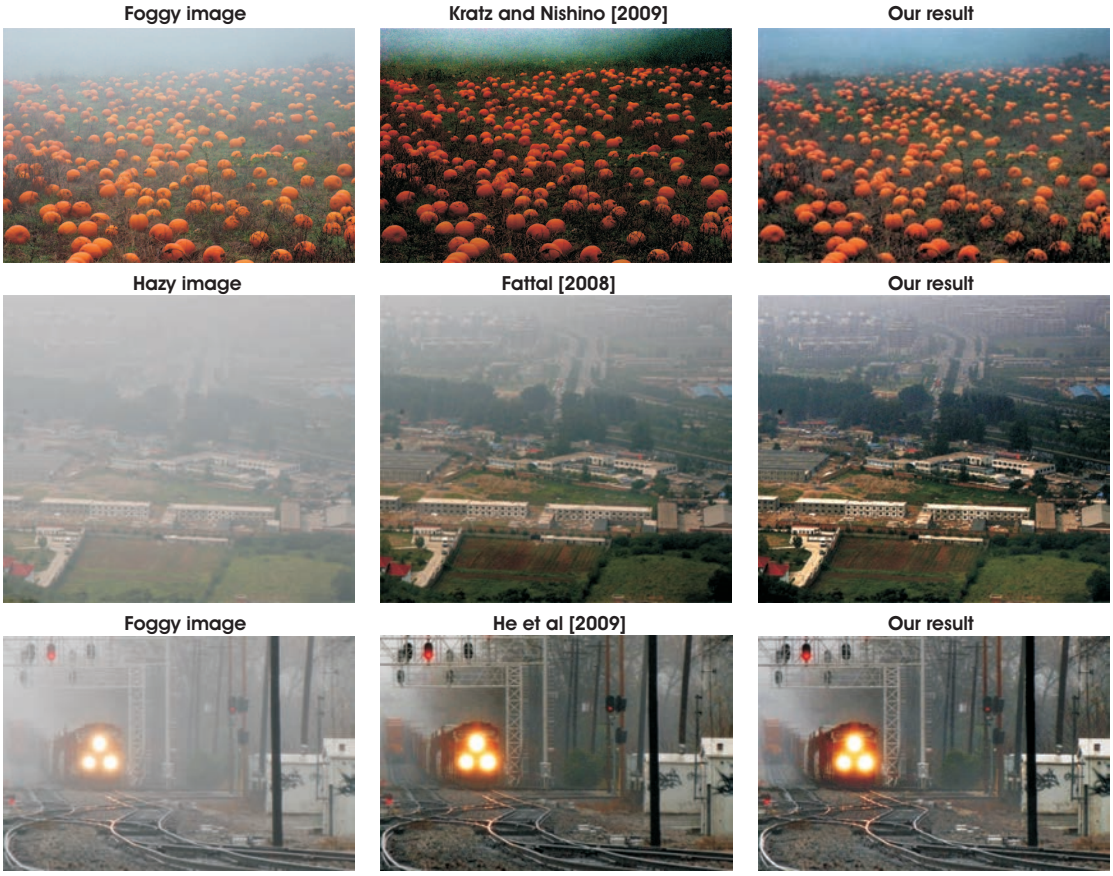

FIGURA 13: Comparison with the single image dehazing techniques of Kratz and Nishino [16], Fattal [13], He et al. [15].

the sky and the clouds have a different hue rendition). Also, the technique of He et al. [15], seems to restore slightly better the fine transitions of the regions closer to the horizon. However, our technique has the advantage of enhancing robustly such degraded images, without estimating the transmission, that needs to be refined by a computationally expensive alpha matting procedure in the approach of He et al. [15].

Figure 14 shows a direct comparison between our result and the output of Schechner et al. [33] that is a multi-image dehazing approach. Nevertheless, as can be seen, our single image dehazing operator is able to produce comparable results with the technique of Fattal [13] but also with the method of Schechner et al. [33]. The technique of Schechner et al. [33] is a polarization-based approach that employs two images - the worst and the best polarization states among the existing image versions. In this case, we processed only one input of those provided by Schechner et al. [33]. Moreover, for this example our technique outperforms unsharp mask and the specialized technique of Tarel and Hautière [17]. Partially,

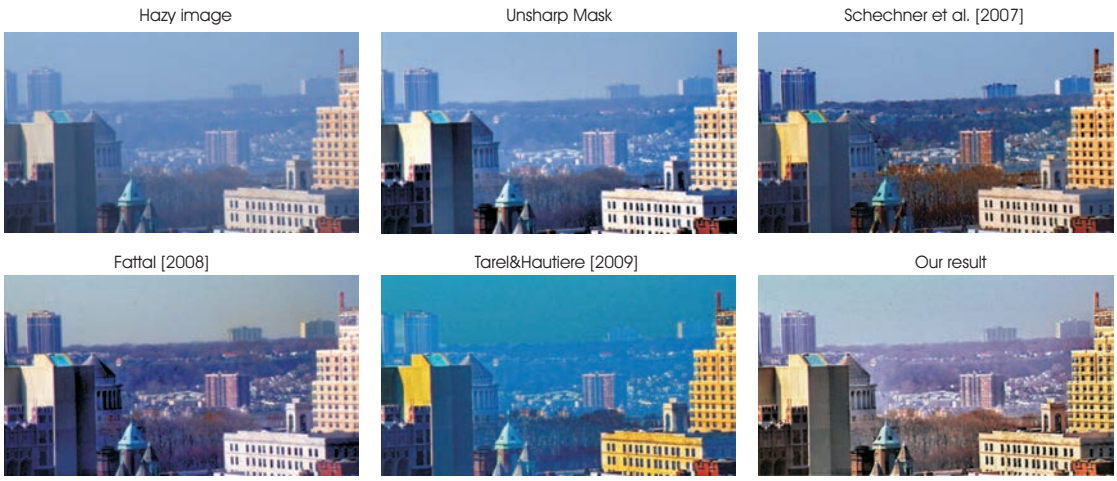

FIGURA 14: A direct comparison between our result and the output of Schechner et al. [33] that is a polarization-based approach that employs two images (the worst and the best polarization states among the existing image versions). For this example our technique yields comparable result with Schechner et al. [33] and Fattal [13] while it outperforms unsharp mask and the specialized technique of Tarel and Hautière [17].

these observations are strengthen by analyzing the gradient distribution of the considered techniques in figures 17, 19.

In figures 15 and 16 are presented direct comparisons against the recent dehazing techniques. In addition to the most representative recent single image dehazing techniques of Tan [14], Fattal [13], He et al. [15], Tarel and Hautière [17], Nishino et al. [25], we analyzed the technique of Kopf et al. [11] that uses a rough 3D map approximation of the scene. Based on these results we perform a quantitative evaluation using the blind measure of Hautière et al. [24].

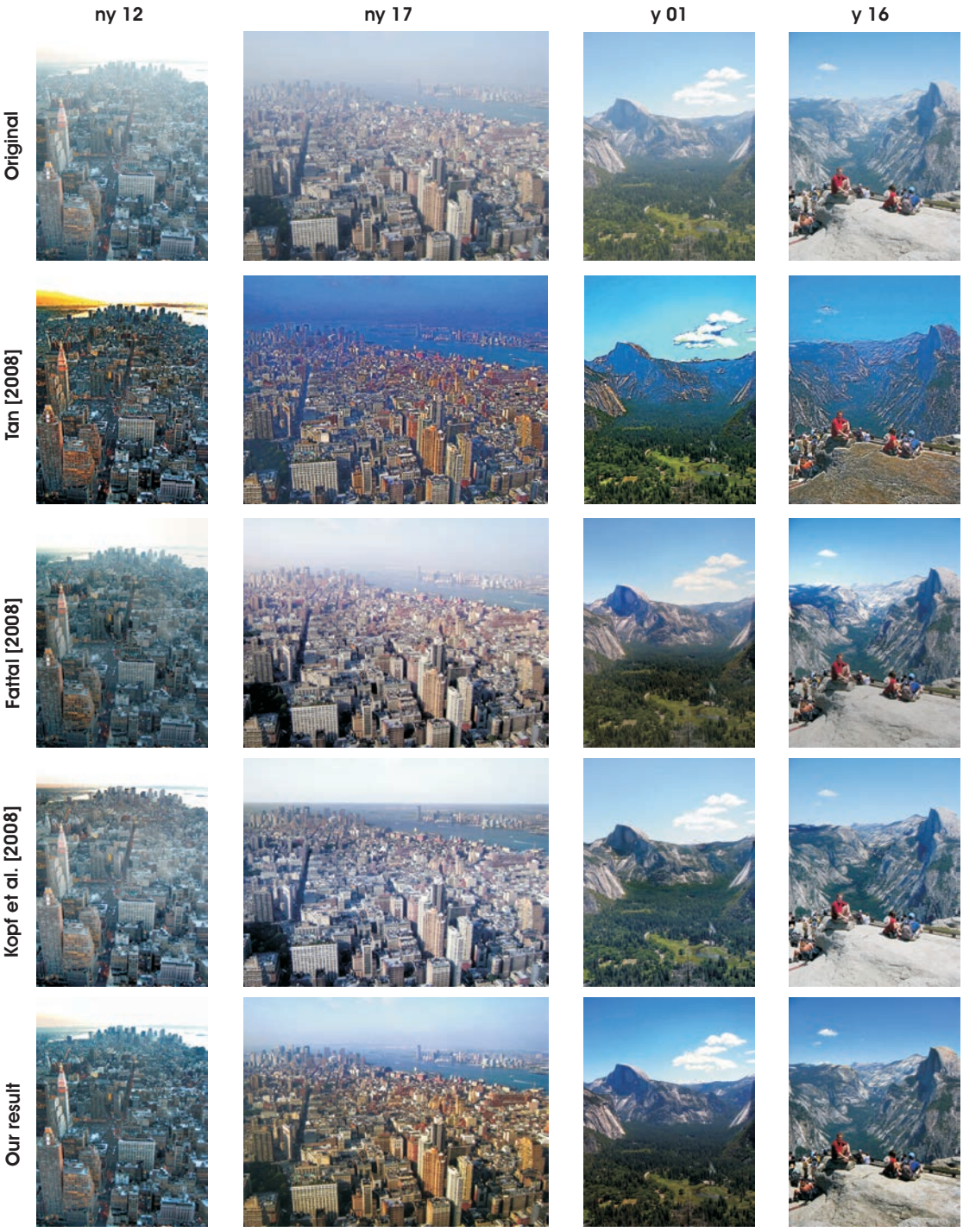

FIGURA 15: Comparison of the recent dehazing techniques. Besides the initial hazy images in this figure are displayed the results of Tan [14], Fattal [13], Kopf et al. [11] and our technique. The reader is asked for a close inspection.

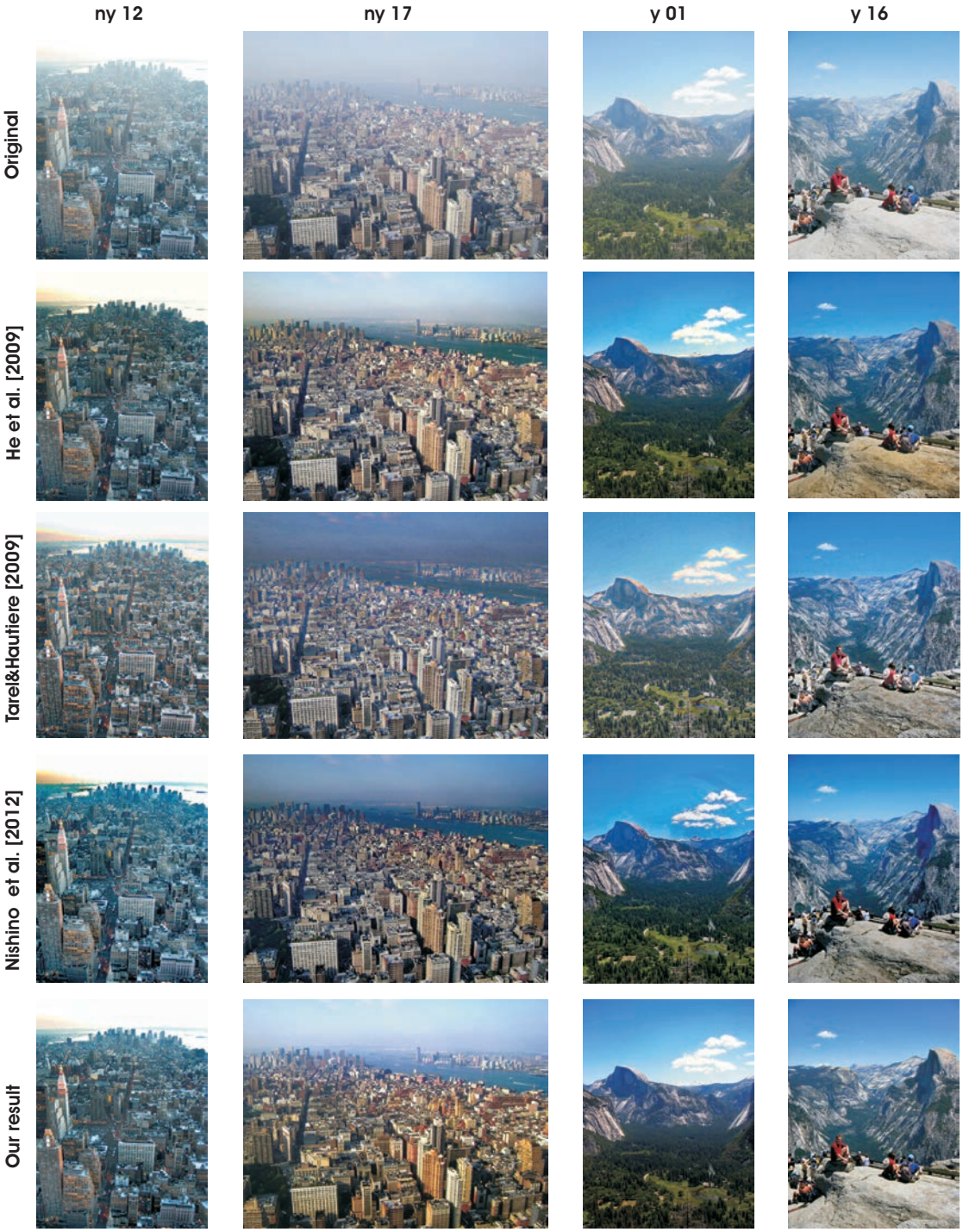

FIGURA 16: Comparison of the recent dehazing techniques. Besides the initial hazy images in this figure are displayed the results of He et al. [15], Tarel and Hautière [17], Nishino et al. [25] and our technique. The reader is asked for a close inspection.

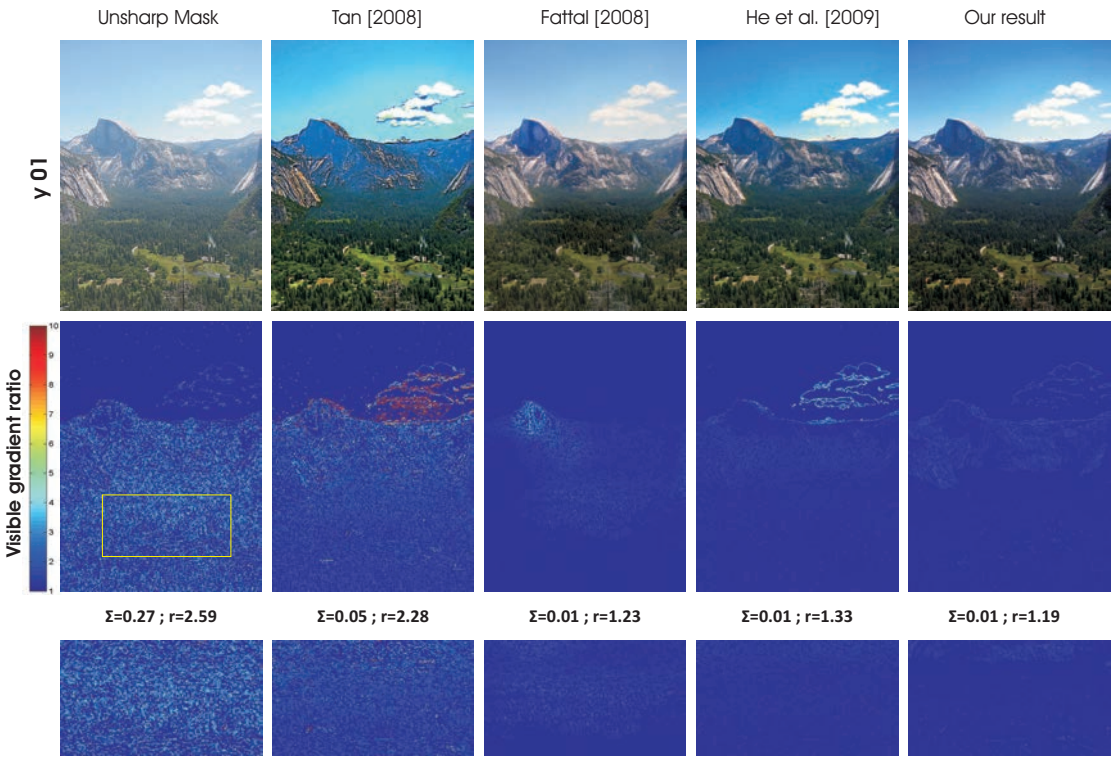

FIGURA 17: Interpretation of the indicator $\bar{r}$ proposed by Hautière et al. [24]. Except *unsharp mask filter* the analyzed methods yield low values of Σ (the percentage of pixels that becomes completely white or black after restoration). In general, methods that increase too much the local contrast are characterized by high values of indicator $\bar{r}$. This is demonstrated by the results shown in this figure and figure 19, supported also by the values shown in table 1. As an example, the method of Tan [14] adds some extra edges in the horizon region. We believe that for a more objective evaluation additional characteristics need to be taken into account (e.g. relation between depth and gradients, the amplitude of gradients, the reversal of the gradients).

Basically, this quality assessment approach consists in computing the ratio between the gradients of the image before and after restoration. This is based on the concept of visibility level, commonly used in lighting engineering. In table 18 we considered four images (named as **ny12**, **ny17**, **y01** and **y17**), where indicator e represents edges newly visible after restoration, indicator $\bar{r}$ represents the mean ratio of the gradients at visible edges, while indicator Σ represents the percentage of pixels which become completely black or completely white after restoration. To compute this indicator we used the parameters used in [17]. Besides the aforementioned dehazing techniques, the values of these indicators when applying just an unsharp filter are shown in the table.

Analyzing the results of table 18, in general, all the considered techniques (including our technique) yield small values of the Σ descriptor (the percentage of

	unsharp mask			Tan			Fattal			Kopf			He			Tarel			Nishino			Ours		
	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$	e	Σ	$\bar{r}$
ny12	-0.09	0.72	2.57	-0.14	0.02	2.34	-0.06	0.086	1.32	0.05	0.00	1.42	0.06	0.0	1.42	0.07	0.0	1.88	-0.01	0.46	1.81	0.02	0.0	1.49
ny17	-0.10	1.28	2.29	-0.06	0.01	2.22	-0.12	0.02	1.56	0.01	0.01	1.62	0.01	0.00	1.65	-0.01	0.0	1.87	-0.07	0.91	1.79	0.12	0.0	1.54
y01	0.04	0.27	2.59	0.08	0.01	2.28	0.04	0.02	1.23	0.09	0.00	1.62	0.08	0.01	1.33	0.02	0.0	2.09	0.11	0.71	1.79	0.07	0.01	1.19
y16	0.09	2.32	1.87	-0.08	0.01	2.08	0.03	0.00	1.27	-0.01	0.00	1.34	0.06	0.00	1.42	-0.01	0.0	2.01	0.01	1.71	1.29	0.18	0.01	1.46

FIGURA 18: Qualitative comparison of the four images (**ny12**, **ny17**, **y01** and **y17**) shown in figures 15 and 16 based on the indicators Σ and $\bar{r}$ of Hautière et al. [24].

pixels which become completely black or completely white after the restoration). On the other hand, indicator e shows that most of the methods depending on the processed image remove some of the visible edges. Interestingly, only our method and He et al. [] technique, are characterized by positive values of the indicator e for the considered images.

Moreover, regarding indicator $\bar{r}$, the measure produces small values of indicator $\bar{r}$ (the ratio of the gradient norms after and before restoration) for our results. Since one may observe that the appearance in our case is globally restored and the value of indicator $\bar{r}$ is kept close to the minimum value ($\bar{r}$=1), one may deduce that in our case the local contrast was restored moderately. This feature is achieved also by the methods of Fattal [13], He et al. [15] and Kopf et al. [11]. On the other hand, the techniques of Tarel and Hautière [17] and Tan [14] increase too strongly the local contrast and as a result these approaches have higher values of indicator $\bar{r}$. Regarding indicator $\bar{r}$, the first group of techniques demonstrates less spurious edges and artifacts. Analyzing images of figure 17 that displays results processed by unsharp mask, Tan [14], Fattal [13], He et al [15] and our approach, one may observe that in general the techniques with low values of indicator $\bar{r}$ show less spurious edges and artifacts.

To the best of our knowledge, the blind measure of Hautière et al. [24] is the only existing method designed to give a quantitative interpretation for dehazing operation. The indicators of this measure are able to reveal only partially the level of restoration and degradation. Obviously, for a more objective evaluation additional characteristics need to be taken into account (e.g. relation between depth and gradients, the amplitude of gradients, the reversal of the gradients). For example, after restoration, gradients that were not visible should be emphasized, and that effect needs to be proportional to the distance from the camera.

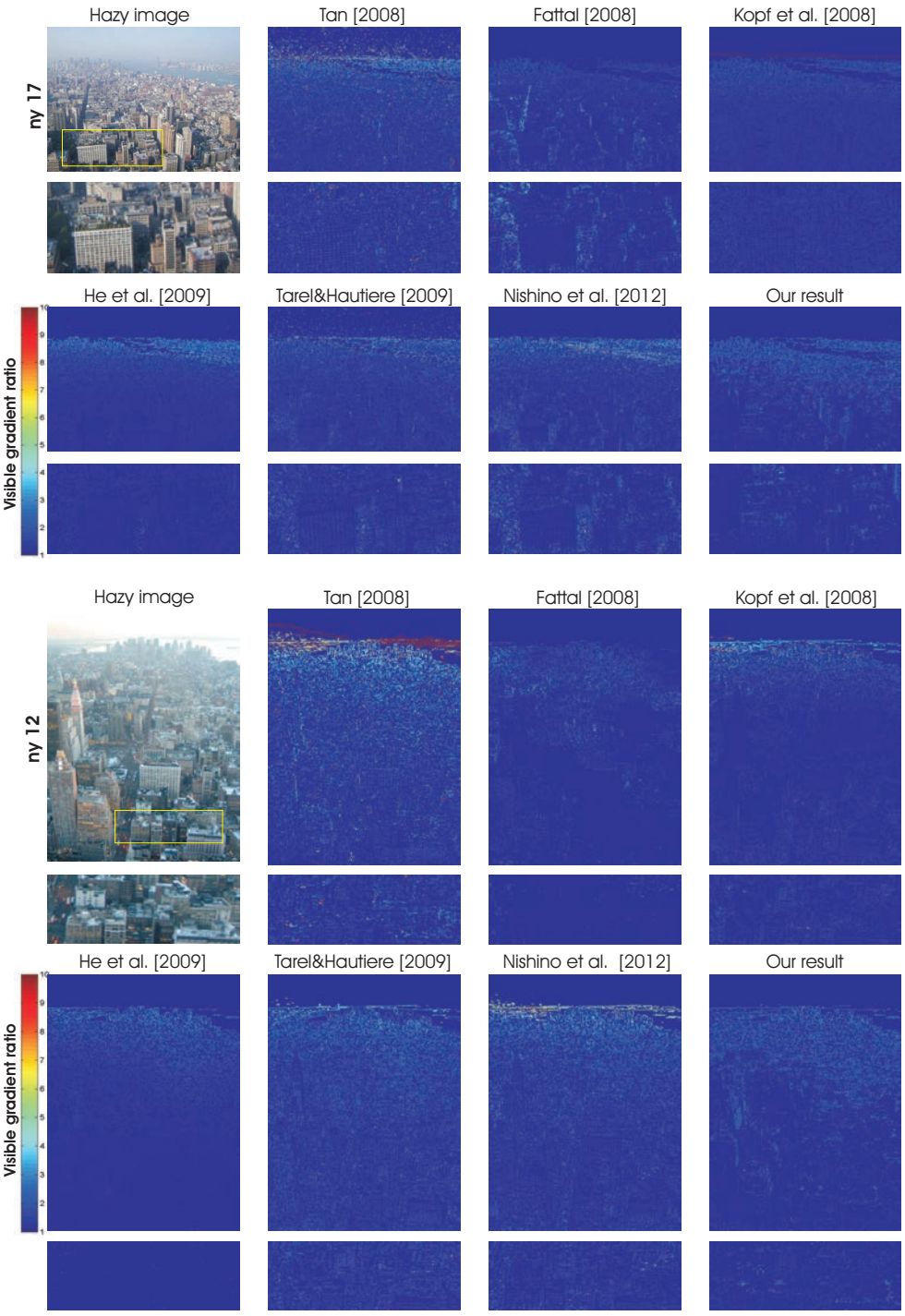

FIGURA 19: Comparison of the analyzed dehazing techniques based on the indicator $\bar{r}$ proposed by Hautière et al. [24]. To better visualize the restoration but also the level artifacts introduced by different techniques a cropped region is shown below.

One would expect to see a stronger *visible gradient ratio* as one goes further away (towards the top of the image). While, the method of Tan [14] yields a too strong emphasis of the gradients, the results of physically-based methods such as Kopf et al. [11], Nishino [25] and also our results demonstrate edge enhancements that are consistent with the depth variation. Interestingly, the results of the method by Fattal [13] are characterized by stronger gradients for near scene points while the results of He et al. [15] present in some cases a very non-linear emphasis of the gradients.

To conclude, by a simple visual inspection of the results in figures 17 and 19, our technique is shown to be less prone to artifacts, yielding results similar to the physically-based techniques of He et al. [15] , Fattal [13], Kopf et al. [11] and Nishino et al. [25]. This is a key advantage of our technique, since we do not explicitly estimate the depth, while our method is much faster and simple than the other approaches that yield similar outputs. Moreover, it can be observed that the techniques of Tan [14] and Tarel and Hautière [17] restore some non-existent gradients.

FIGURA 20: Our technique is limited to deal with homogeneous hazy images.

Furthermore, compared with most of the existing techniques, an important advantage of our strategy is required computation time, since our method is able to processes a 600×800 image in approximately 2-300 ms (20% for derived imputs, 35% for the weight maps while the multi-scale strategy takes approximately 45% of the entire fusion process on an Intel Core i7 CPU, 8GB RAM). In comparison, the method of Tan [14] needs more than 5 minutes per image, He et al. [15] requires 20 seconds, Tarel and Hautière [17] technique takes less than 0.5 seconds, the method of Fattal [13] needs 35 seconds, while the processing times of the method [25] were not reported.

Even though the proposed method performs in general well, as the previous methods, a limitation of our algorithm may be observed for images that are characterized by non-homogeneous haze layers. Figure 20 depicts this issue. As can be seen, the other single image dehazing approaches present serious limitations while

FIGURA 21: More results generated for foggy scenes.

tackling this challenging case (e.g. the technique of Tarel and Hautière [17] yields unpleasing artifacts such as coarse edges and color distortion). Moreover, even though some enhancement may be achieved, our technique is limited to processing color images.

To summarize: in this part we have demonstrated that a fusion-based approach can be used to effectively enhance hazy and foggy images. To the best of our knowledge, this is the first fusion-based strategy that is able to solve such problems using only one degraded image. We have shown that, by choosing appropriate weight maps and inputs, a multi-scale fusion strategy can be used to effectively dehaze images. Our technique has been tested on a large data set of natural hazy images. The method is faster than existing single image dehazing strategies and yields accurate results. In future work we would like to test our method on videos.

II.1.4 Night-Time Dehazing by Fusion

Capturing good quality outdoor images poses interesting challenges since such scenes often suffer from poor visibility introduced by weather conditions such as haze or fog. The process dehazing has been tackled using such information as rough depth [11] of the scene or multiple images [10]. More recently, several techniques [1, 13–17, 58–63], have introduced solutions that do not require any additional information than the single input hazy image.

While the effectiveness of these techniques has been extensively demonstrated on daylight hazy scenes, they suffer from important limitations on night-time hazy scenes. This is mainly due to the multiple light sources that cause a strongly non-uniform illumination of the scene. Night-time dehazing has been addressed only recently [64–66]. Pei and Lee [64] estimate the airlight and the haze thickness by applying a color transfer function before applying the dark channel prior [15, 67] refined iteratively by bilateral filtering as a post-processing step. The method of Zhang et al. [65] estimates non-uniform incident illumination and performs color correction before using the dark channel prior. Li et al.[66] employ an updated optical model by adding the atmospheric point spread function to model the glowing effect. A spatially varying atmospheric light map is used to estimate the transmission map based dark channel prior.

In this section we present a different approach to solving the problem of night-time dehazing. The technique has been published recently by the candidate in [2]. Our approach is the first fusion-based method of restoring hazy night-time images. Image fusion is a well-known concept that has been used for image editing [37], image compositing [39],image dehazing [1], HDR imaging [41], underwater image and video enhancement [4] and image decolorization [68]. The approach described here is built on our previous fusion-based daytime dehazing approach [1] that has been recently extended by Choi et al. in [69].

To deal with the problem of night-time hazy scenes (refer to Fig. 22), we propose a novel way to compute the airlight component while accounting for the non-uniform illumination presents in nighttime scenes. Unlike the well-known dark-channel strategy [67] that estimates a constant atmospheric light over the entire image, we compute this value locally, on patches of varying sizes. This is found to succeed since under night-time conditions, the lighting results from multiple artificial sources, and is thus intrinsically non-uniform. In practice, the

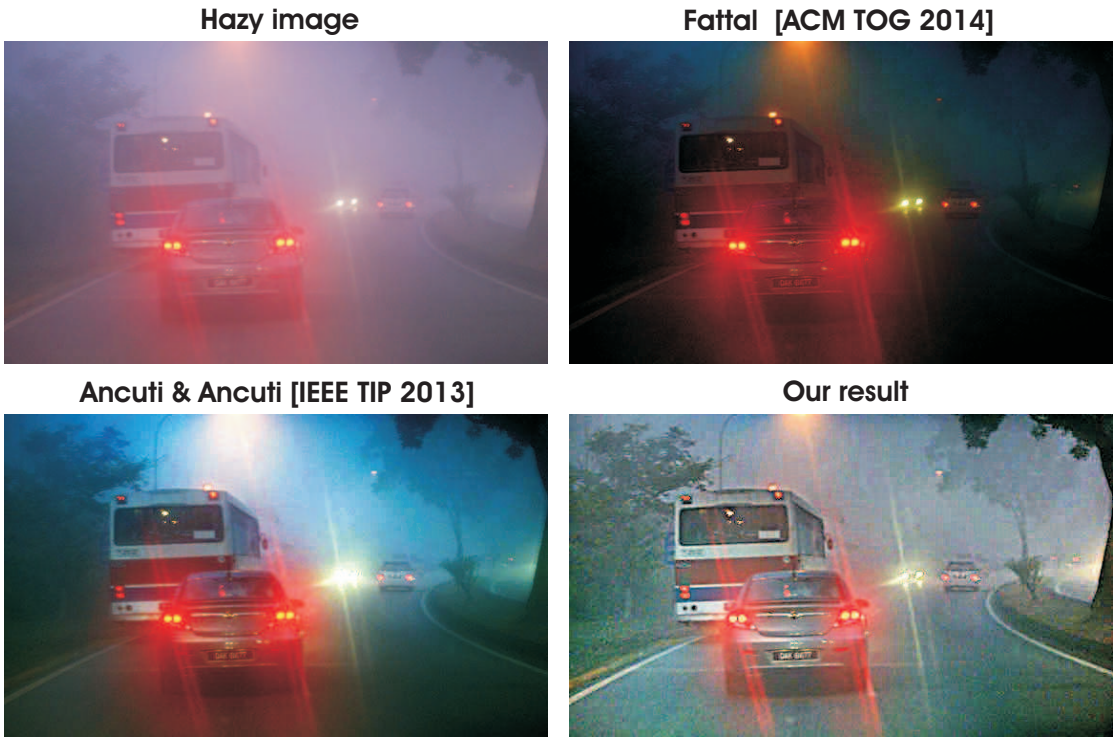

FIGURA 22: *Night-time scene capture is a challenging task under difficult weather conditions and recent single-image dehazing techniques [1, 60] suffer from important limitations when applied to such images.*

local atmospheric light causes the color observed in hazy pixels, which are the brightest pixels of local dark channel patches. Selecting the size of the patches is non-trivial since small patches are desirable to achieve fine spatial adaptation to the atmospheric light, it might also lead to poor light estimates and reduced chance of capturing hazy pixels. For this reason, we deploy multiple patch sizes, each generating one input to the multi-scale fusion process. Our fusion approach is accomplished in three main steps. First, based on our airlight estimation using different sizes of the patches we derive the first two inputs of the fusion approach. To reduce the glowing effect and emphasize the finest details of the scene, the third input is defined to be the Laplacian of the original image. In the second step, the important features of these derived inputs are filtered based on several quality weight maps (local contrast, saturation and saliency). Finally the derived inputs and the normalized weight maps are blended in a multi-scale fashion using a Laplacian pyramid decomposition of the inputs and a Gaussian pyramid of the normalized weights.

The experimental section describes testing of our technique on a large set of different night-time hazy scenes. The results demonstrate the efficacy of our approach as compared to recent competitive techniques both in terms of computational efficiency and quality of the outputs.

Airlight Estimation On Night-time Hazy Scenes. Next part explains how the atmospheric factors affecting the image formation process can be estimated from night-time images formed in the presence of non-uniform (artificial) lightning.

Due to atmospheric particles that absorb and scatter light, only a fraction of the radiance from a scene point reaches the observer. Koschmieder's model [18] is a relevant description of atmospheric effects caused by weather on the observer. In short, it states that the light intensity $\mathcal{I}$ at each image coordinate x is the result of two main additive components - *direct transmission* $\mathcal{D}$ and *airlight* $\mathcal{A}$:

$$\mathcal{I}(x) = \mathcal{D}(x) + \mathcal{A}(x) = \mathcal{J}(x)\ T(x) + A_\infty\ [1 - T(x)] \tag{20}$$

where $\mathcal{J}$ is the scene radiance or haze-free image (estimated, which would reach the observer unaltered in absence of atmospheric effects), T is the transmittivity along the cone of vision and A_∞ is the atmospheric intensity, resulting from environmental illumination.

In Koschmieder's model, the transmission map $T(x)$ is directly related to the depth of the scene. For homogeneous medium, $T(x) = e^{(-\beta\ d(x))}$ where β is the medium attenuation coefficient due to scattering, and d is the distance between the observer and the considered surface.

Following [67], and adopt the well-known dark channel (DC) prior to estimate $T(x)$ without resorting to depth estimation. The DC prior assumes that natural objects have a weak reflectance in one of the color channels (the direct radiance is small, or dark, in at least one of the R, G, B color channels [8]), while the atmospheric intensity conveys all colors (the haze looks grey or white, i.e. all components in A_∞ are significant). Hence, assuming that A_∞ is known (we discuss estimation of it later), then $T(x)$ can be directly estimated from the weakest color (relative to atmospheric color) over a neighborhood of x. Formally, under the assumption that $\min_{y\in\Omega(x)} (\min_{c\in r,g,b} \mathcal{J}^c/{A_\infty}^c) = 0$, Koschmieder's model states that:

$$T(x) = 1 - \min_{y\in\Omega(x)} \left(\min_{c\in r,g,b} \mathcal{I}^c/{A_\infty}^c \right) \tag{21}$$

where A_∞^c is the component of the atmospheric light associated with color c, and $\Omega(x)$ represents a local patch centered at x.

Atmospheric intensity estimation

Early methods estimated the atmospheric intensity as the color vector having the highest intensity pixel [14]. This choice was motivated by the white appearance of haze in day-time scenes. Such estimate could however fail, e.g. on a white object instead of a hazy pixel. To circumvent this problem, the authors of [67] proposed to estimate the atmospheric intensity using the most haze-opaque pixels. These are defined as the ones having the brightest dark channel, i.e as the ones maximizing $I_{DC}(x) = \min_{y\in\Omega(x)}(\min_{c\in r,g,b}\mathcal{I}^c(y))$, where r, g, b denote the R,G,B color channels.

This estimator works well on day-time scenes, but suffers from two weaknesses when applied to night scenes (see Fig. 23). First, it globally estimates the atmospheric intensity, over the entire picture, whereas the night-scenes are characterized by artificial and spatially non-uniform environmental illumination. Second, by maximizing the minimum over the set of color channels, their method finds those locations taking large values in all channels. It thus implicitly assume that the atmospheric intensity is reasonably white, which is the case in day-time scenes, but is not necessarily true of night-scenes which can present strongly colored lighting.

FIGURA 23: **Rough dehazing of night-time scenes.** *Designed for day-time dehazing, the well-known dark channel [67] has important limitations on night scenes because it uniformly estimates the airlight constant. As may be observed, our patch-based estimate (also not refined) of the airlight component is more appropriate for the night-time hazy scenes. In particular, color and details that are close to light sources are better enhanced. By simply replacing these estimates in the model equation results in some rough dehazing outputs.*

To address those two limitations, we propose (i) to estimate the atmospheric intensity locally, within spatial neighborhoods $\Psi(x)$ around each coordinate x, and (ii) to compute each component of the atmospheric light independently. Formally, we define the local atmospheric intensity $A^c_{L\infty}(x)$ to be:

$$A^c_{L\infty}(x) = \max_{y\in\Psi(x)}\left[\min_{z\in\Omega(y)}\left(\mathcal{I}^c(z)\right)\right] = \max_{y\in\Psi(x)}\left[I^c_{MIN}(z)\right] \tag{22}$$

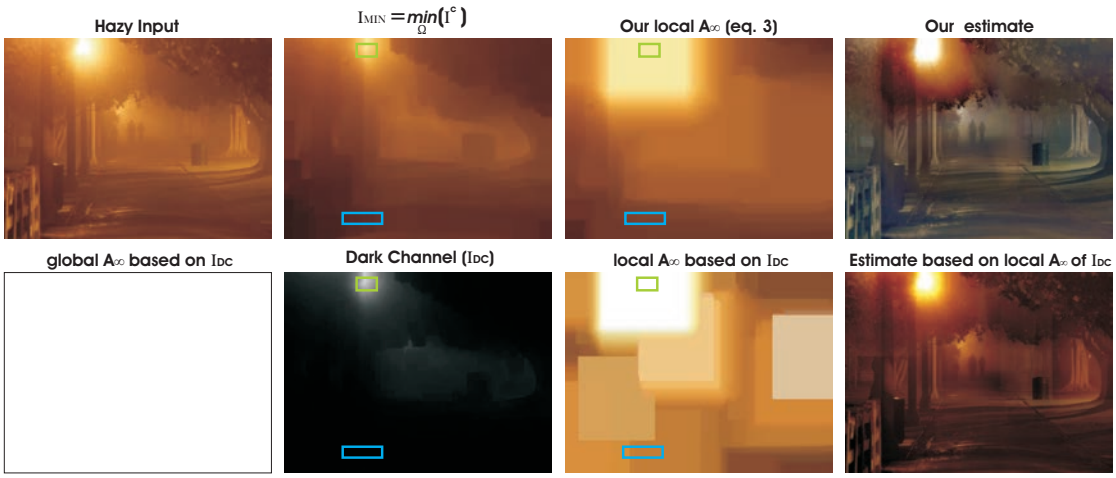

FIGURA 24: **Local airlight estimation.** *A global estimate of the A_∞ based on dark channel results in a white airlight. As a consequence, the dark channel might become very small (see blue rectangle in second image of bottom row), which means $1-T(x) \approx 0$ in Equation21 and no airlight influence in Equation 20. In contrast, our local airlight estimate on the same blue rectangle results in a colored atmospheric light, which in turns results in a non-unity transmission and a non-zero atmospheric light influence.*

This formulation is again motivated by Koschmieder's model, where scene radiance $\mathcal{J}^c(x)$ can be written as $\rho^c(x).A^c_{L\infty}(x)$, with $\rho^c(x)$ denoting the normalized radiation coefficient. Maximizing independently on each color, as done in Equation 22, is equivalent to finding the coordinate y around which the normalized radiation coefficient is maximum for color c. In particular, when an object reflects all incident color c, this coefficient is close to one at every coordinate z within neighborhood of y, and $\mathcal{I}^c(z) \approx A^c_{L\infty}(z).T(z) + A^c_{L\infty}(z).(1-T(z)) = A^c_{L\infty}(z)$.

Obviously, each potential light source has an influence that goes beyond the size of the patch Ω used to validate the lighting relevance. Here, all results have been generated using patches Ψ twice the size of Ω.

Figure 24 compares the atmospheric intensity estimated by global and local strategies. Local estimation appears to capture the major changes arising from environmental illumination, while the global approach does not. More importantly, the bottom, rightmost pictures reveal the benefit of computing each atmospheric intensity component independently, compared to searching for the location in $\Psi(x)$ maximizing the minimum over the 3 color channels, as a straightforward locally adaptive extension of [67] would do. We also observe that the image reconstructed estimated by our local method of the atmospheric intensity is of better appearance (both in color and details) those resulting from local estimation obtained based on joint processing of the color channels.

Finally, it is worth noting that, when $\Psi(x)$ is defined to cover the entire image, our method reduces to a global estimator. Interestingly, in this case, Fig. 25 reveals that the global estimate of $A^c_{L\infty}$ derived by maximizing $\min_{y\in\Omega(x)}(\mathcal{I}^c(y))$ over the entire image is quite similar to the one proposed in [67] for day-time scenes. Hence, our proposed estimator may be regarded to be a night-friendly generalization of the concepts introduced in [67].

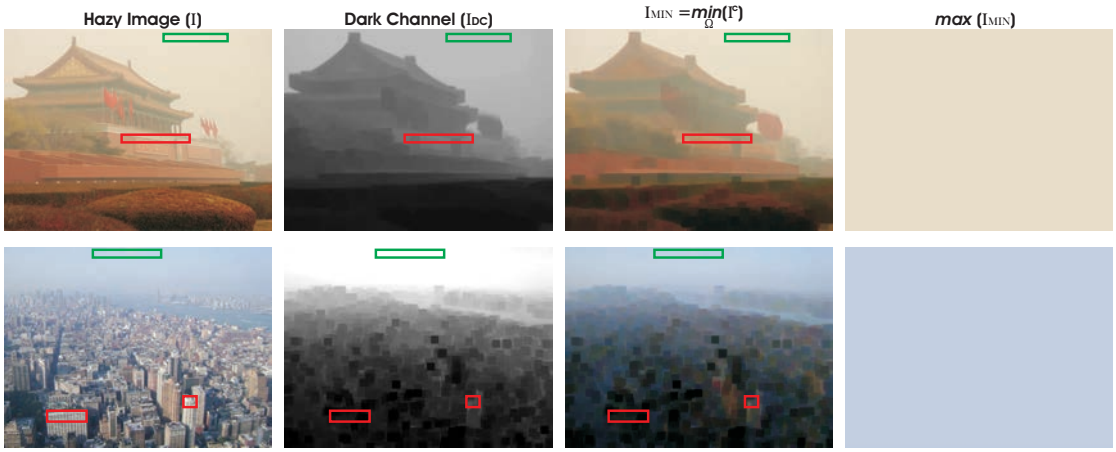

FIGURA 25: **Airlight estimation**. *Our global airlight estimate $A^c_{L\infty}$, derived by maximizing $\min_{y\in\Omega(x)}(\mathcal{I}^c(y))$ over the entire image, appears to be quite similar to the atmospheric intensity estimated by He et al. [67], from the brightest region of the dark channel (depicted by green rectangles). The red rectangles show that, in daytime scenes, the two approaches equally reject the high image intensity locations that are not relevant regarding airlight estimation. He et al. [67] method rejects them because the prior is dark in those regions, our strategy because I_{MIN} gets darker than the initial image in those regions that are not subject to intense airlight illumination. Since I_{MIN} does not make any implicit assumption about the whiteness of the atmospheric illumination, it is more general than [67], especially in presence of artificial colored lighting.*

Fusion Process. While a significant degree of picture enhancement is obtained using the above described local estimation procedure in the optical model, important artifacts arise at and around patch transitions, where color shifting and glowing defects are visible. Moreover, as detailed below, the choice of patch size could result in poor quality output images owing non-uniformity of the airlight in night-time scenes. To solve this problem we propose a multi-scale fusion approach which can effectively and seamlessly enhance hazy night-time images.

Derived Inputs. Our fusion technique is a single image-based approach that uses several inputs from the original hazy image. In the first stage we employ the strategy previously described that locally estimates the airlight values. An important problem that arises is how to choose the optimal size of the patch in order to be able to deal with the case where multiple light sources are contained

within the same patch. Selecting a size that is too small risks improper haze removal, effects of diffusely distributed light from a source may not be included in that patch.

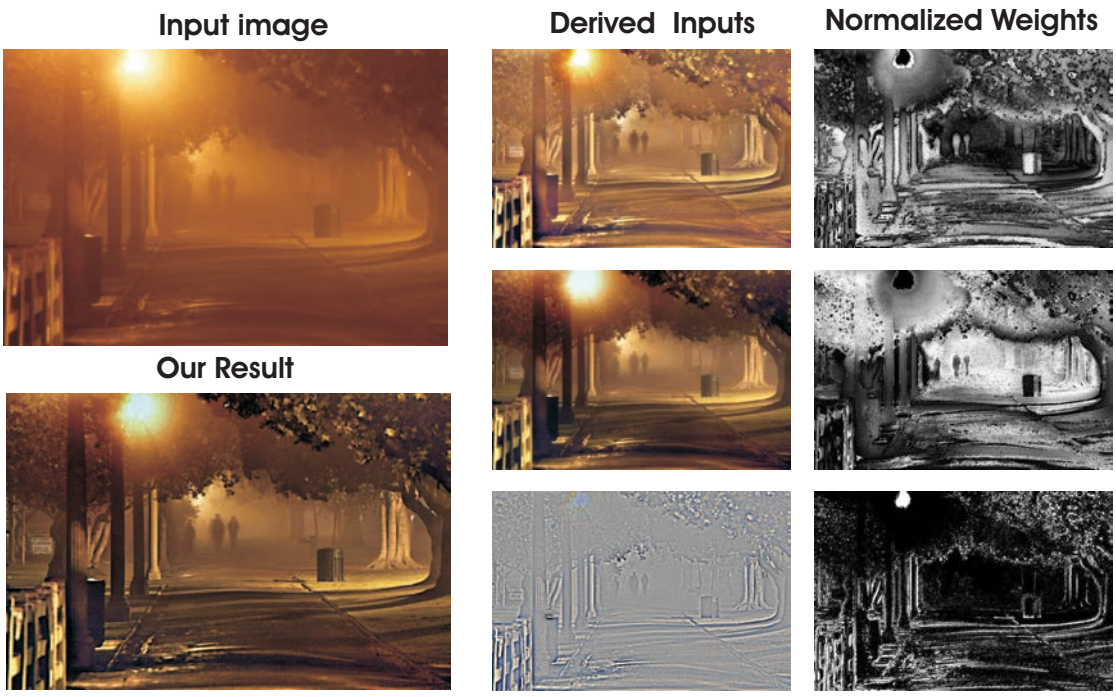

FIGURA 26: *Derived inputs and corresponding normalized weight maps.*

By choosing a too large patch size the haze is better removed, but as may be seen in Fig. 26 the color might be shifted, the influence of the airlight might not be entirely removed and some details may remain poorly restored.

Based on this observations, we derive a **first** and a **second input** based on our airlight estimation using different patch sizes. The **first input** is computed using a small patch size (e.g. 20×20 for an image of size 800×600), thereby preventing estimation of the airlight from multiple light sources. However this input is characterized by an important loss of global contrast and chroma. We solve this limitation by computing a **second input** using larger patches (e.g. 80×80 for an image of size 800×600). This derived input considerably improves the global contrast, since it removes a significant fraction of the airlight. The transitions between neighboring patches is smoothed using a simple gaussian filter. When more than one light source is included in the region of interest, a winner take it all procedure decides which light source estimates the airlight.

However, as shown in Fig. 26, glowing effects are still visible in the derived inputs. To reduce such undesired effects we derive a **third input** which is the discrete

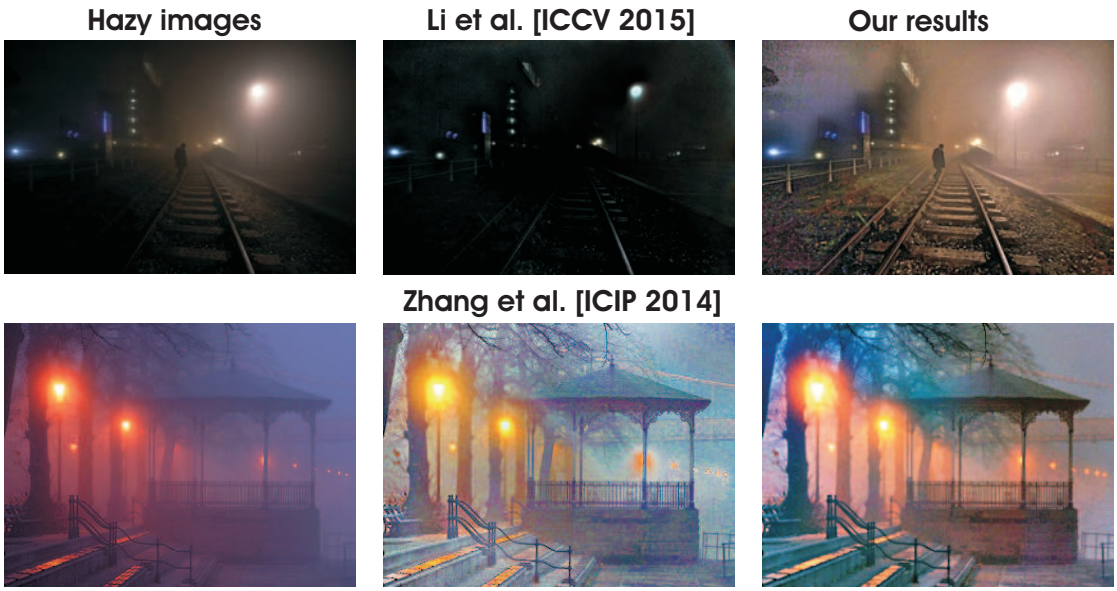

FIGURA 27: *Comparative results with the recent night-time dehazing technique of Zhang et al. [65] and Li et al. [66].* **Please refer to supplementary material for more comparative results.**.

Laplacian of the original image. This input makes it possible to enhance the finest details that are transferred to the fused output.

Definition of the Weight Maps. Inspired by our previous fusion dehazing approach [1], we derive three weight maps to ensure that regions of high contrast or of high saliency will receive greater emphasis in the fusion process.

Local contrast weight estimates the amount of local variation of each input and is computed by applying a Laplacian filter to the luminance of each processed image. This indicator has been used in applications such as tone mapping [41] and assigns high values to edges and texture variations.

Saturation weight map controls the saturation gain in the output image. This factor is motivated by the fact that humans generally prefer images characterized by a high level of saturation. This measure is computed as the standard deviation across channels at each coordinate.

Saliency weight map highlights the most conspicuous regions of an image compared with their surroundings. Using the well-known saliency technique of Achanta et al. [55] this weight map is computed as a difference between a Gaussian smoothed version of the input and its mean value.

Multi-scale Fusion. The main goal of the fusion process is to produce a better output image by effectively blending inputs that are guided by specific weight maps designed in order to preserve the most significant features of the inputs.

The simplest way is to directly combine the inputs and weight maps as $\mathcal{R}_{NF}(x) = \sum_k \bar{\mathcal{W}}^k(x)\mathcal{I}_k(x)$ (where $\mathcal{I}_k$ represents the k^{th} input weighted by the normalized weight maps $\bar{W}^k$). However, this naive fusion strategy has been shown to cause annoying halo artifacts, mostly at locations with strong transitions in the weight maps. Such unpleasing artifacts can be overcame by using a multi-scale Laplacian decomposition, a well-known concept dating to Burt and Adelson [56].

Similarly to other single-image dehazing approaches [1, 69], each input $\mathcal{I}_k$, is decomposed into a Laplacian pyramid while the normalized weight maps $\bar{\mathcal{W}}^k$ are decomposed using a Gaussian pyramid. Using the same number of levels, the Gaussian and Laplacian pyramids are independently fused at each level:

$$\mathcal{R}_l(x) = \sum_k G_l\left\{\bar{\mathcal{W}}^k(x)\right\} L_l\left\{\mathcal{I}_k(x)\right\} \tag{23}$$

where l represents the number of the pyramid levels, $L\{\mathcal{I}\}$ denotes the Laplacian of the input $\mathcal{I}$, and $G\left\{\bar{\mathcal{W}}\right\}$ is the Gaussian-smoothed normalized weight map $\bar{\mathcal{W}}$.

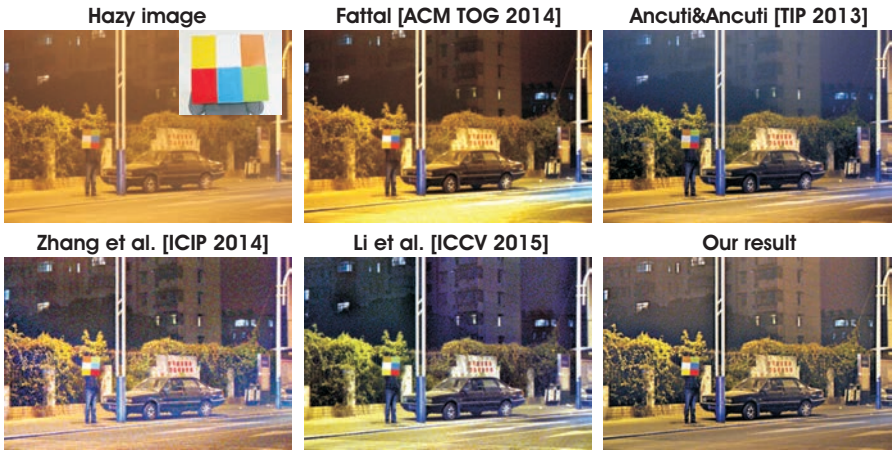

FIGURA 28: **Comparative results.** *The night-time hazy image with color palette (left side, top row) is enhanced by several dehazing techniques. See Table 1 for the PSNR values.*

The fused result $\mathcal{R}$ is processed by summing the contributions from all the computed levels of the pyramid:

$$\mathcal{R}(x) = \sum_l \mathcal{R}_l(x) \uparrow^d \tag{24}$$

where $\uparrow^d$ is the upsampling operator with factor $d = 2^{l-1}$.

Results and Discussion. We extensively tested our approach using the recent image dataset introduced in [66] that contains various quality and formats of images taken of night-time scenes. We compared our method with the recent night-time dehazing techniques of Zhang et al. [65] and Li et al. [66] and also with the day-time dehazing methods of Ancuti and Ancuti [1] and Fattal [60]. For all the results we used the original code provided by the authors on their webpages.

While Fig. 22 demonstrates limitations of day-time dehazing techniques when applied to night-time hazy images, Fig. 27 directly compares our approach with the recent specialized techniques of Li et al. [66] and Zhang et al. [65]. The method of Li et al. [66] tends to darken the original image and to over-amplify colors in some regions. On the other hand, as compared with our approach, the strategy of Zhang et al. [65] is less robust to the the glowing effect and may introduce color artifacts.

Moreover, our approach has the advantage of simplicity and computational efficiency. Our unoptimized Matlab implementation processes an 800×600 image in less than 4 seconds. The method of Li et al. [66] computes results on a similar image in more than 30 seconds while the method of Zhang et al. [65] requires a similar computation as He et al. [15] (approx. 20 seconds per image).

	yellow	white	brown	red	blue	green	**average**
Fattal	21.10	23.94	15.43	20.71	15.12	15.77	18.68
Ancuti	17.75	15.82	13.49	19.16	14.01	17.01	16.20
Zhang et al.	21.20	23.21	**21.30**	20.10	15.38	12.66	18.98
Li et al.	19.80	23.21	16.94	**23.38**	**17.69**	**21.09**	20.35
Our method	**27.33**	**30.04**	18.58	23.21	17.59	17.66	**22.40**

TABELA 1: *Evaluation of the results in Fig. 28 based on the PSNR values computed as an average on RGB components for each of the 6 colors of the reference palette.*

We also performed a quantitative evaluation using the pair of images provided by Zhang et al. [65]. On the left side of the top row of Fig. 28 is shown the reference color palette and the night-time hazy image containing the same palette. We processed this input image using several different dehazing techniques [1, 60, 65, 66] and computed the PSNR values for each of the 6 colors (shown in Table 1). As can be seen, our approach generally performs better in terms of PSNR compared with the other techniques.

II.1.5 D-Hazy: A dataset to evaluate quantitatively dehazing algorithms

In this section we describe a novel dataset that allows to quantitatively evaluate the existing dehazing techniques. The technique has been published recently by the candidate in [3].

Image dehazing, a typical image enhancement technique studied extensively in the recent years, aims to recover the original light intensity of a hazy scene. While earlier dehazing approaches employ additional information such as multiple images [10] or a rough estimate of the depth [11], recent techniques have tackled this problem by using only the information of a single hazy input image [1, 13–17, 58–60, 62, 63, 70]. The existing techniques restore the latent image assuming the physical model of Koschmieder [18]. Since dehazing problem is mathematically ill-posed there are various strategies to estimate the two unknowns: the airlight constant and the transmission map. Fattal [13] employs a graphical model that solves the ambiguity of airlight color assuming that image shading and scene transmission are locally uncorrelated. Tan's method [14] maximizes local contrast while constraining the image intensity to be less than the global atmospheric light value. He et al. [15, 67] introduce a powerful approach built on the statistical observation of the dark channel, that allows a rough estimation of the transmission map, further refined by an alpha matting strategy [35]. Tarel and Hautière [17] introduce a filtering strategy assuming that the depth-map must be smooth except along edges with large depth jumps. Kratz and Nishino [16] propose a Bayesian probabilistic method that jointly estimates the scene albedo and depth from a single degraded image by fully leveraging their latent statistical structures. Ancuti et al. [59] describe an enhancing technique built on a fast identification of hazy regions based on the *semi-inverse* of the image. Ancuti and Ancuti [1] introduce a multi-scale fusion procedure that restore such hazy image by defining proper inputs and weight maps. The method has been extended recently by Choi et al. [69]. Meng et al. [71] propose a regularization approach based on a novel boundary constraint applied on the transmission map. Fattal [60] presents a method inspired from color-lines, a generic regularity in natural images. Tang et al. [63] describe a framework that learns a set of feature for image dehazing.

There have been a few attempts to quantitatively evaluate dehazing methods. All of them have been defined as non-reference image quality assessment (NR-IQA)

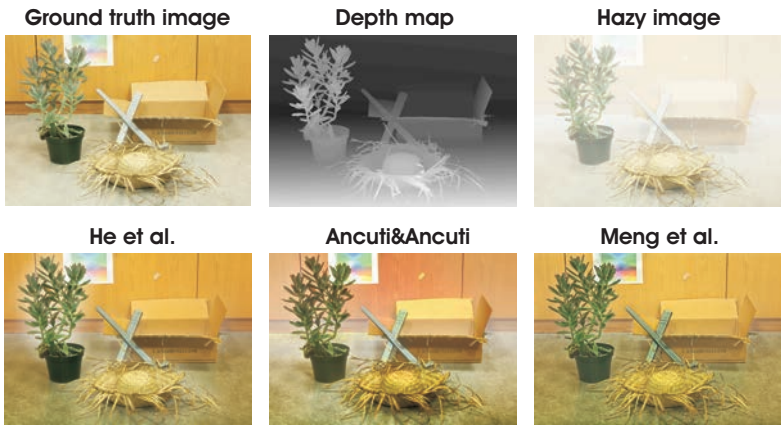

FIGURA 29: *D-HAZY dataset provides ground truth images and the corresponding hazy image derived from the depth map (known). In the bottom row are shown results yielded by several recent dehazing techniques [1, 67, 71].*

strategies. Hautiere et al. [24] propose a blind measure based on the ratio between the gradient of the visible edges between the hazy image and the restored version of it. Chen et al. [72] introduce a general framework for quality assessment of different enhancement algorithms, including dehazing methods. Their evaluation was based on a preliminary subjective assessment of a dataset which contains source images in bad visibility and their enhanced images processed by different enhancement algorithms. Moreover, general non reference image quality assessment (NR-IQA) strategies [73–75] have not been designed and tested for image dehazing.

However, none of these quality assessment approaches have been commonly accepted and as a consequence a reliable data set for dehazing problem is extremely important. Unlike other image enhancing problems for dehazing task capturing a valid ground truth image is not trivial. The procedure to record both the reference (haze-free) and the hazy image in the same illumination condition is generally intractable. The FRIDA dataset [76] designed for Advanced Driver Assistance Systems (ADAS) is a synthetic image database (computer graphics generated scenes). It contains 66 roads synthesized scenes and besides the reduced level of generality and complexity of the scenes as a computer-generated dataset some parameter settings are not valid for real scenarios.

Our dataset contains 1400+ images of real complex scenes. In order to generate the hazy images we use the extended *Middleburry*[1] dataset that contains high quality real scenes and corresponding depth map. Moreover, we improve our dataset

[1] http://vision.middlebury.edu/stereo/data/scenes2014/

using recent *NYU-Depth V2*[2], a large dataset that includes various indoor scenes with RGB and depth maps captured by a Microsoft Kinect sensor. Employing the Koschmieder's physical model [18] of light transmission in hazy scenes, assuming uniform atmospheric intensity and uniform haze density, and using the reference depth map available from the dataset, we are able to synthesize haze in the considered scenes. Even if the strict validity of the Koschmieder model is arguable in arbitrary illumination and haze density conditions, it is revelvant to synthetise hazy images based on this model because it is at the core of all modern dehazing techniques. Hence, all those methods should provide good results when the model is valid. As an important and surprising contribution, our work however reveals that, even for hazy images that perfectly fit the model, none of the existing dehazing technique is able to accurately reconstruct the original image from its hazy version. This observation has been derived from a comprehensive evaluation of several state of the arts dehazing approaches based on SSIM [77] and CIEDE2000 [78] measures computed between the reference (haze-free image) and the restored results produced by different dehazing techniques.

From Depth to Hazy Scenes. A hazy medium is characterized by small particles that respond to changes in relative humidity acting as small droplets nuclei when the humidity is higher than a certain level. In such medium, the light that is passing through it is attenuated along its original course and is distributed to other directions. Mathematically this process is expressed by the the image formation model of Koschmieder's model [18] that is widely accepted by all the recent dehazing approaches. Based on this model, due to the atmospheric particles that absorb and scatter light, only a certain percentage of the reflected light reaches the observer. The light intensity $\mathcal{I}$ of each pixel coordinate x, that passes a hazy medium, is the result of two main additive components - *direct attenuation* $\mathcal{D}$ and *airlight* $\mathcal{A}$:

$$\mathcal{I}(x) = \mathcal{D}(x) + \mathcal{A}(x) = \mathcal{J}(x)\ T(x) + A_{\infty}\ [1 - T(x)] \tag{25}$$

where $\mathcal{J}$ is the scene radiance of a clear medium (haze-free image), T is the *transmission* along the cone of vision and A_{∞} is the atmospheric light (a color constant that is computed globally for the day-time dehazing).

[2] http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

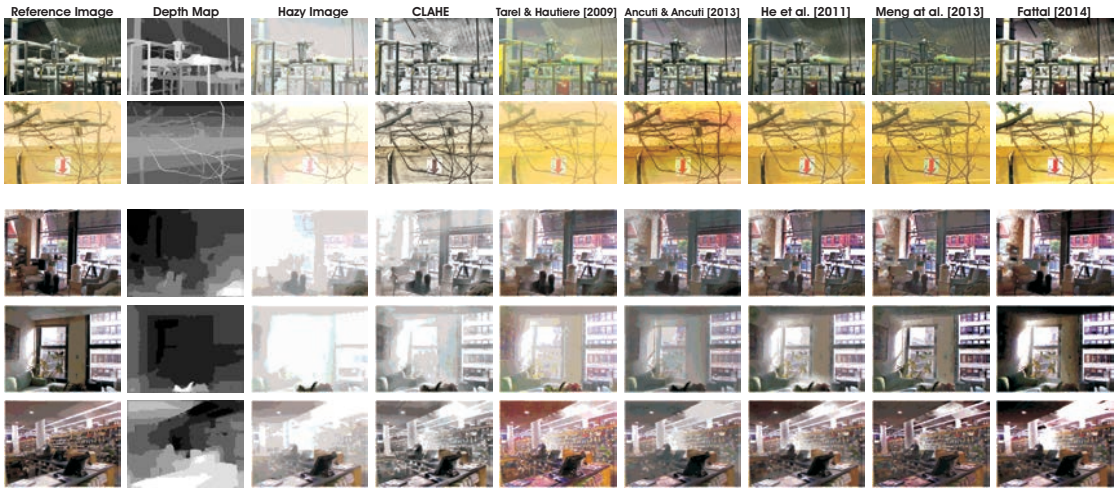

FIGURA 30: ***Comparative results.*** *The first two rows show the ground truth, depth map, hazy images and results derived from the Middleburry dataset while the last three rows show results derived from NYU-Depth dataset.*

The airlight component is linearly correlated with the distance between the observer and the target object of the scene. The first component of the model, *direct attenuation* $\mathcal{D}$, describes how the scene radiance is attenuated with the distance. The second one, the *airlight* component $\mathcal{A}$ represents the principal source of the additive color shifting and is expressed as:

$$\mathcal{A}(x) = A_{\infty} \ \ [1 - T\,(x)] \tag{26}$$

where T is the transmission and represents the relative fraction of light able to cross the hazy medium between the observer and scene surface, without being scattered.

Basically, the transmission map T is directly related with the depth of the scene and considering a homogeneous medium this value is expressed as:

$$T\,(x) = e^{[-\beta \ \ d(x)]} \tag{27}$$

where β is the medium attenuation (extinction) coefficient due to the light scattering, while d represents the distance between the observer and the considered surface.

Using Depth to Synthesize Hazy Scenes. Depth is a key parameter in Equation 27. In general, the existing datasets, are relatively limited in resolution, , realism and accuracy of depth ground truth. To overcome these limitations the recent work of Scharstein et al. [79] has introduced a novel dataset to evaluate

stereo algorithms. The dataset is generated using structured light [80]. It represents an extension of the well-known Middlebury dataset and contains 23 images (6-megapixel) of indoor scenes with subpixel-accurate depth ground truth.

To deal withthe the lack of ground truth in occluded regions, for the images of the Middlebury dataset we have employed the recent weighted median filtering strategy introduced by Ma et al. [81].

Additionally, for a more comprehensive evaluation we have also considered the recent NYU-Depth V2 data set [82]. This data set includes various indoor scenes with RGB and depth maps captured by Microsoft Kinect sensor. Missing depth values have been filled in using the colorization scheme of Levin et al. [83]. While this dataset is not as accurate as the Middlebury dataset it has the advantage to be much larger (1449, 640×480 images), with various scenes.

Based on the optical model described above, we synthesize the hazy scenes using the reference (haze-free) image and its corresponding depth map. First, for each image, the transmission map is estimated based on equation 27 using the depth d and the medium attenuation coefficient β. β is set by default to 1, which corresponds to moderate and homogenous haze. Additionally, for the atmospheric light constant A_∞ we assume a pure white value [1 1 1], and generate the hazy images based on equations 20 and 27 as:

$$\mathcal{I}(x) = \mathcal{J}(x) \ e^{[-\beta \ d(x)]} + A_\infty \ \left[1 - e^{[-\beta \ d(x)]}\right] \tag{28}$$

Evaluated Techniques. In this study, using our new dataset D-HAZY, we evaluate perform an comprehensive validation comparing several state of the art single image dehazing techniques and one well-known enhancing method. For all tested algorithms we use the original implementation provided by the authors. In the following we briefly describe these techniques.

1. Tarel and Hautiere [17] introduce one of the first single dehazing image approach. The method restores the visibility of hazy images based on a filtering strategy. They assumes that the white-balance is performed as a pre-processing step and estimates the transmission as a percentage of the difference between the local average of image and the local standard deviation of it. The transmission is refined based on an extended version of the median filter. The method has the

	CLAHE		Tarel		Ancuti&Ancuti		He et al.		Meng et al.		Fattal	
	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000
Adirondack	0.728	11.257	0.851	15.195	0.890	10.820	0.858	10.775	0.882	11.106	0.753	16.057
Backpack	0.610	14.896	0.874	11.721	0.877	11.739	0.916	10.013	0.890	10.629	0.869	13.300
Bicycle1	0.768	18.067	0.876	8.843	0.899	12.669	0.882	15.296	0.766	23.259	0.809	16.448
Cable	0.499	25.054	0.649	26.469	0.617	24.249	0.710	16.379	0.668	18.895	0.743	13.684
Classroom1	0.609	11.205	0.799	23.315	0.886	11.839	0.911	6.315	0.872	9.834	0.870	20.742
Couch	0.613	11.519	0.716	25.640	0.854	13.262	0.908	6.736	0.862	10.775	0.779	22.941
Flowers	0.703	16.910	0.792	15.861	0.816	14.618	0.876	8.646	0.829	14.013	0.876	8.703
Jadeplant	0.549	26.364	0.687	20.187	0.632	28.415	0.689	19.223	0.699	21.418	0.631	32.390
Mask	0.682	14.781	0.875	11.754	0.860	11.603	0.888	9.703	0.828	14.875	0.886	11.823
Motorcycle	0.761	13.026	0.760	18.585	0.825	14.486	0.816	13.706	0.805	14.150	0.760	17.858
Piano	0.699	9.757	0.790	20.301	0.839	10.336	0.888	6.177	0.855	8.861	0.794	12.588
Pipes	0.628	16.898	0.678	24.285	0.715	19.193	0.779	9.643	0.751	11.541	0.681	15.463
Playroom	0.672	11.952	0.792	18.367	0.831	13.233	0.878	7.430	0.822	12.070	0.802	16.091
Playtable	0.717	9.418	0.831	19.060	0.842	9.118	0.907	8.028	0.865	14.119	0.813	11.519
Recycle	0.688	13.400	0.895	11.890	0.898	12.474	0.876	14.964	0.871	14.846	0.727	20.655
Shelves	0.754	7.681	0.883	15.202	0.929	7.181	0.890	11.649	0.889	14.413	0.871	12.400
Shopvac	0.669	17.659	0.726	29.251	0.788	19.268	0.827	16.287	0.821	17.569	0.758	32.504
Sticks	0.715	20.073	0.884	8.489	0.852	15.214	0.948	7.213	0.953	8.196	0.921	10.865
Storage	0.692	13.335	0.853	16.636	0.806	14.782	0.883	10.194	0.870	12.086	0.819	16.851
Sword1	0.607	20.070	0.862	12.097	0.835	14.809	0.910	9.660	0.892	10.436	0.865	13.822
Sword2	0.638	12.184	0.872	13.485	0.887	12.051	0.899	12.458	0.849	15.494	0.758	37.965
Umbrella	0.574	20.371	0.818	15.130	0.837	15.933	0.838	20.229	0.798	21.489	0.659	33.149
Vintage	0.711	14.907	0.858	8.236	0.856	14.612	0.917	10.049	0.780	21.796	0.858	15.794
Average	0.665	15.252	0.810	16.956	0.829	14.431	0.865	11.338	0.831	14.429	0.796	18.418

FIGURA 31: ***Quantitative evaluation.*** *For each image derived from the Middleburry dataset we compute the SSIM and CIEDE2000 indexes between the ground truth images and the enhanced results of the evaluated techniques.*

advantage to be fast working for both color and grayscale images.

2. He et al. [15, 67] proposes a novel prior the dark-channel that is derived from the dark object of Chavez [8]. They explore the statistics that in most of the local regions which do not cover the sky, some pixels in general are characterized by very low values in at least one color channel. These filtered pixels per patch are used to estimate the haze transmission that is refined by an alpha matting strategy. In our evaluation we employ the dark channel prior refined based on the guided filter [84].

3. Meng et al. [71] introduce a regularization strategy that explores effectively the boundary constraint on the transmission map. The boundary constraint of the transmission is an extension of the well-known dark channel prior [15]. The transmission is refined by an optimization problem using the boundary constraint combined with a weighted L1-norm based contextual regularization.

4. Ancuti and Ancuti [1] describe an effective a multi-scale fusion strategy for single-image dehazing. They derive two inputs, the first one is processed by white balancing the original hazy image while the second input is obtained by subtracting for each pixel the average luminance value of the entire image. Their important features are filtered by computing several measures (weight maps) that are blend in a multi-scale fusion strategy.

5. Fattal [60] introduces an approach that explores the color lines, pixels of small image patches typically exhibit a one-dimensional distribution in RGB color space. A first estimate of the transmission map is obtained by computing the lines offset from the origin. The final transmission map is produced based on a Markov random field that refines the noise and other artifacts due to the scattering.

6. CLAHE (Contrast-limiting adaptive histogram equalization) [85] is a well-known enhancing technique that restores the contrast of the images. Desighned original for medical imaging CLAHE extends the adaptive histogram equalization (AHE) by applying a contrast limiting procedure. CLAHE splits the images into contextual regions and employs the histogram equalization to each of the region. To generate the CLAHE results we used the original implementation of Matlab2014b.

Validation. Qualitatively, as expected, CLAHE [85] yields the less visually compelling results (see Fig. 30). While the method of Tarel and Hautiere [17] has the advantage to be computationally efficient, the results produced by this method look over-saturated with unpleasing halo artifacts. Guided by several perceptual measures the method of Ancuti and Ancuti [1] mitigates the introduction of structural artifacts due to the multi-scale fusion strategy. Despite of its euristic built-in concept, the method of He et al. [15] appears to perform generally better than the other approaches both in color and structure restoration. However, it can be observed that this approach yields over-corrected results in white and gray regions where no color channel is dominant. Since it also builds on the dark channel prior, it is not surprising to observe that the approach of Meng et al. [71] yields similar results as He et al. [15], with slightly reduced artifacts, thanks to high-order filtering of the transmission map. Whilst being quite effective on some images, the technique of Fattal [60] regularly over-saturates some regions due to the color-lines prior, and sometimes introduces unpleasing structural artifacts around the edges.

Quantitatively, to validate the different techniques described previously, we compare their outcome with the ground truth haze-free images provided of the D-HAZY dataset. Since PSNR has been proven to not be very effective in predicting human visual response to image quality [77], we compute the well-known structural similarity (SSIM) index [86] that compares local patterns of pixel intensities

that have been normalized for luminance and contrast. SSIM index yields decimal values between -1 and 1, with maximum value 1 for two identical images.

Moreover, because in image dehazing restoration of the color is crucial and cannot be evaluated reliably by SSIM we employ an additional evaluation metric. The difference between two colors is a metric of high interest in color science. While the earlier measures (e.g. CIE76 and CIE94) shown important limitations to resolve the perceptual uniformity issue, CIE introduced CIEDE2000 [78, 87] which defines a more complex, yet most accurate color difference algorithm. CIEDE2000 yields values in the range [0,100] with smaller values indicating better color preservation, and values less than 1 corresponding to visually imperceptible differences.

Table 31 presents a detailed validation for the 23 Middleburry dataset images and table 2 shows the average values of SSIM and CIEDE2000 measured over the 1449 NYU-Depth images.

From these tables, we conclude that the method of He et al. [15] performs the best in average. A second group of methods including Meng et al. [71], Ancuti and Ancuti [1] and Fattal [60], perform relatively well both in terms of structure and color restoration.

	CLAHE	**Tarel**	**Ancuti**	**He**	**Meng**	**Fattal**
SSIM	0.622	0.719	0.771	0.811	0.773	0.747
CIEDE2000	18.054	17.742	14.136	11.029	12.216	14.656

TABELA 2: *Quantitative evaluation of the 1449 images generated based on the NYU-Depth dataset. In this table are shown the average values of the SSIM and CIEDE2000 indexes over the entire dataset (1449 images).*

In general, all the tested methods introduce structural distortions such as halo artifacts close to the edges, that are amplified in the faraway regions. Moreover, due to the poor estimation of the airlight and transmission map from the hazy image, some color distortions may create some unnatural appearance of the restored images. In summary, there is not a single technique that performs the best for all images. The relatively low values of SSIM and CIEDE2000 measures prove once again the difficulty of single image dehazing task and the fact there is still much room for improvement.

II.2 Enhancing Underwater Images and Videos

This section describes a novel strategy to enhance underwater videos and images. Built on the fusion principles, our strategy derives the inputs and the weight measures only from the degraded version of the image. In order to overcome the limitations of the underwater medium we define two inputs that represent color corrected and contrast enhanced versions of the original underwater image/frame, but also four weight maps that aim to increase the visibility of the distant objects degraded due to the medium scattering and absorption. Our strategy is a single image approach that does not require specialized hardware or knowledge about the underwater conditions or scene structure. Our fusion framework also supports temporal coherence between adjacent frames by performing an effective edge preserving noise reduction strategy. The enhanced images and videos are characterized by reduced noise level, better exposedness of the dark regions, improved global contrast while the finest details and edges are enhanced significantly. In addition, the utility of our enhancing technique is proved for several challenging applications.

Underwater imaging is challenging due to the physical properties existing in such environments. Different from common images, underwater images suffer from poor visibility due to the attenuation of the propagated light. The light is attenuated exponentially with the distance and depth mainly due to absorption and scattering effects. The absorption substantially reduces the light energy while the scattering causes changes in the light direction. The random attenuation of the light is the main cause of the foggy appearance while the the fraction of the light scattered back from the medium along the sight considerably degrades the scene contrast. These properties of the underwater medium yields scenes characterized by poor contrast where distant objects appear misty. Practically, in common sea water, the objects at a distance of more than 10 meters are almost indistinguishable while the colors are faded since their characteristic wavelengths are cut according to the water depth.

There have been several attempts to restore and enhance the visibility of such degraded images. Mainly, the problem can be tackled by using multiple images [10], specialized hardware [88] and by exploiting polarization filters [89]. Despite their effectiveness to restore underwater images, these strategies have demonstrated several important issues that reduce their practical applicability. First, the hardware

solutions (e.g. laser range-gated technology and synchronous scanning) are relatively expensive and complex. The multiple-image solutions require several images of the same scene taken in different environment conditions. Similarly, polarization methods process several images that have different degrees of polarization. While this is relatively feasible for outdoor hazy and foggy images, for the underwater case, the setup of the camera might be troublesome. In addition, these methods (except the hardware solutions) are not able to deal with dynamic scenes, thus being impractical for videos.

In this section, we describe a novel approach that is able to enhance underwater images based on a single image, as well as videos of dynamic scenes. Our approach, first published in [4] is built on the fusion principle that has shown utility in several applications such as image compositing [39], multispectral video enhancement [90], defogging [91] and HDR imaging [41]. In contrast to these methods, our fusion-based approach does not require multiple images, deriving the inputs and the weights only from the original degraded image. We aim for a straightforward and computationally inexpensive that is able to perform relatively fast on common hardware. Since the degradation process of underwater scenes is both multiplicative and additive [92] traditional enhancing techniques like white balance, color correction, histogram equalization shown strong limitations for such a task. Instead of directly filtering the input image, we developed a fusion-based scheme driven by the intrinsic properties of the original image (these properties are represented by the weight maps). The success of the fusion techniques is highly dependent on the choice of the inputs and the weights and therefore we investigate a set of operators in order to overcome limitations specific to underwater environments. As a result, in our framework the degraded image is firstly white balanced in order to remove the color casts while producing a natural appearance of the sub-sea images. This partially restored version is then further enhanced by suppressing some of the undesired noise. The second input is derived from this filtered version in order to render the details in the entire intensity range. Our fusion-based enhancement process is driven by several weight maps. The weight maps of our algorithm assess several image qualities that specify the spatial pixel relationships. These weights assign higher values to pixels to properly depict the desired image qualities. Finally, our process is designed in a multi-resolution fashion that is robust to artifacts. Different than most of the existing techniques, our framework can deal with dynamic scenes. To preserve temporal coherence in videos we apply temporal bilateral filtering [90] between adjacent frames.

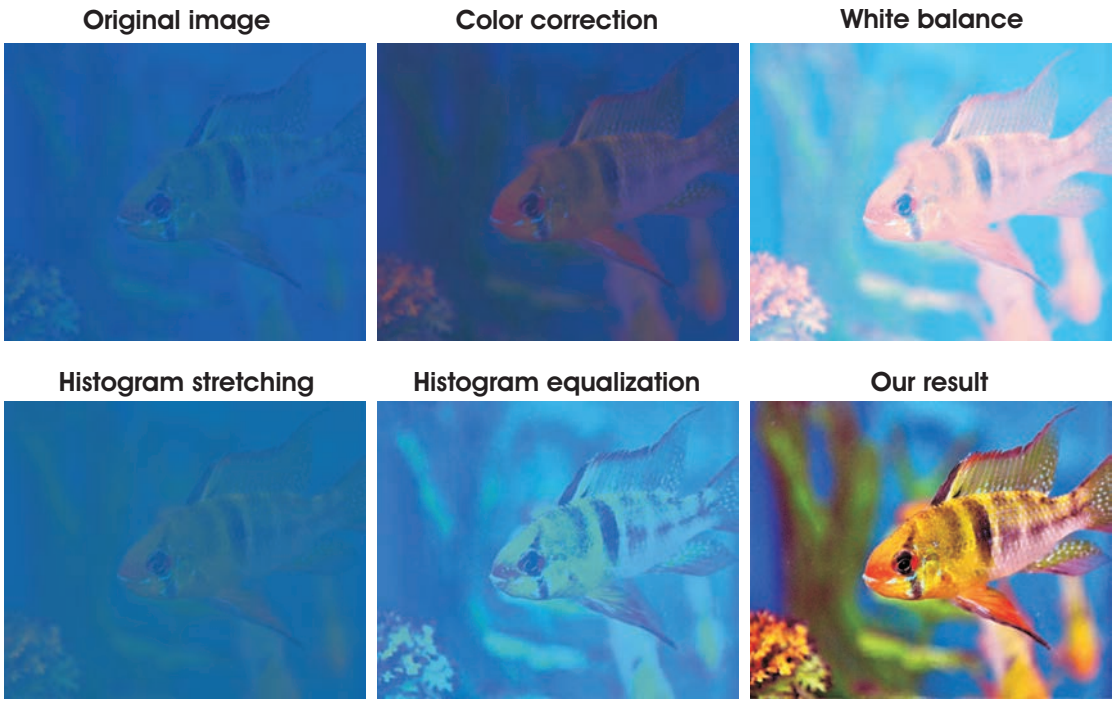

FIGURA 32: Traditional enhancing techniques that are found in commercial tools presents limitations when dealing with underwater images.

Contributions. Our technique is characterized by the following main contributions:

1. A straightforward fusion-based framework that effectively blends different well known filters in order to enhance underwater images based on a single input.
2. Our strategy is able to enhance underwater videos of dynamic scenes. Until now, this was demonstrated only using hardware-based solutions.
3. A robust white balancing techniques specialized for underwater scenes that was validated based on an extensive study.
4. We demonstrate that the simple Laplacian pyramid yields effective results comparable with the recent edge preserving filters such as WLS [57].
5. To the best of our knowledge we are the first that demonstrate utility of an underwater enhancement/restoration technique for several complex applications such as segmentation, image matching by local feature points and image dehazing.

Related Work. An important class of techniques that restore underwater images is represented by specialized hardware solutions [88, 93, 94]. For instance, the divergent-beam underwater Lidar imaging (UWLI) system [88] uses an optical/laser-sensing technique to capture turbid underwater images. Unfortunately, these complex systems are very expensive and power consuming.

Traditional enhancing techniques that are found in commercial tools such as color correction, histogram equalization and stretching, linear mapping and white balancing presents limitations when dealing with subsea images since they simply process pixel values, ignoring the scene content (please refer to Figure 32 and supplementary material). Based on these classical techniques, several specialized underwater enhancing approaches have been introduced in the literature [95–97]. Similarly to traditional enhancing techniques, these approaches are only successful when are applied to relatively well illuminated scenes and in general they introduce strong halos and to distort the colors. We take a different approach that effectively blends several filters in order to solve more challenging and diverse underwater scenarios. Our strategy is designed to follow the spatial relations of the pixels by defining proper weights. We show that our technique is robust to restore a wide range of underwater images (e.g. different cameras, depths, light conditions) with high accuracy being less prone to artifacts and recovering important faded features and edges.

Due to some similarities of light propagation in these environments, the underwater restoration task could also be related to the problem of restoring images degraded by fog and haze. Recently, several single image dehazing techniques have been introduced [13, 15–17]. These dehazing techniques assume the Koschmieder's theory [18] on the apparent luminance of objects observed against background sky on the horizon. Obviously, the underwater imaging is more challenging due to the complexity of light scaterring and lack of scene illumination. Since dehazing could be seen as a simplified underwater problem, we show that our technique is suitable as well for the task of restoring hazy images.

Other descaterring approaches employ multiple images [10, 98] or a rough approximation of the scene model [11]. Narasimhan and Nayar [10] exploited changes in intensities of scene points under different weather in order to detect depth discontinuities in the scene. *Deep Photo* system [11] is able to restore images by employing the existing georeferenced digital terrain and urban 3D models. Since this additional information (images and depth approximation) is generally not available, these methods are relatively impractical for common users.

Polarization methods are probably the most successful underwater restoration methods. These approaches are based on different assumptions using several images of the same scene that have been taken employing different degrees of polarization obtained by rotating a polarizing filter attached to the camera. Chenault

and Pezzaniti [99] assumes that the natural illumination scattered toward the observer is unpolarized while the radiance of the target object is polarized. Other approaches [89, 100] assume the degree of polarization of the target objects to be negligible. Even though in some cases these methods may better recover the distant regions, the polarization techniques were not proved for videos and are therefore limited when dealing with dynamic scenes.

II.2.1 Underwater Enhancing by Fusion

Our work proposes an alternative single image-based solution built on the multi-scale fusion principles. We aim for a simple and fast approach that is able to increase the visibility of a wide variation of underwater videos and images. Even though we do not explicitly follow specialized optical models (e.g. McGlamery [101]), our framework blends specific inputs and weights carefully chosen in order to overcome the limitation of such environments. For the most of the processed images shown in this section the back-scattering component (yielded in general due to the artificial light that hits the water particles and then is reflected back to the camera) has a reduced influence. This is generally valid for underwater scenes decently illuminated by natural light. However, even when artificial illumination is needed, the influence of this component can be easily diminished by modifying the angle of the source light [93].

Our enhancing strategy consists of three main steps: inputs assignment (derivation of the inputs from the original underwater image), defining weight measures and multi-scale fusion of the inputs and weight measures.

Inputs of the Fusion Process. When applying a fusion algorithm the key to obtain good visibility of the final result is represented by the well tailored inputs and weights. Different than most of the existing fusion methods (however, none of them designed to deal with underwater scenes), our fusion technique processes only a single degraded image. The general idea of image fusion is that the processed result, combines several input images by preserving only the most significant features of them. Thus, results obtained by a fusion-based approach fulfills the depiction expectation when each part of the result presents an appropriate appearance in at least one of the input images. In our single-based image approach two inputs of the fusion process are derived from the original degraded image.

Our enhancing solution does not search to derive the inputs based on the physical model of the scene, since the existing models are quite complex to be tackled. Instead, we aim for a fast and simple technique that works generally. The first derived input is represented by the color corrected version of the image while the second is computed as a contrast enhanced version of the underwater image after a noise reduction operation is performed (see figure 33).

White Balancing of the Inputs. White balancing is an important processing step that aims to enhance the image appearance by discarding unwanted color casts, due to various illuminants. In water deeper than 30 ft, white balancing suffers from noticeable effects since the absorbed colors are difficult to be restored. Additionally, underwater scenes present significant lack of contrast due to the poor light propagation in this type of medium.

Considering the large availability of white balancing methods [102] we have searched for a proper solution to our problem. In the following are briefly revised several important approaches that we have analyzed (more in-depth details are found on [102]).

The Finlayson's approach Shades-of-Grey [103] computes the illumination of the scene for each channel by using the Minkowski p -norm. For $p = 1$, this expression is a particular case of the Gray-World [102] while for $p = \infty$ it approximates the White-Patch hypothesis [102]. Despite of its simplicity, the low-level approach of Finlayson and Trezzi [103] has shown to yield comparative results to those of more complex white balance algorithms such as the recent method of [50] that relies on on natural image statistics. The Grey-Edge hypothesis of Weijer and Gevers [104], similarly with Shades-of-Grey [103] can also be formulated by extending the p-th Minkowski form.

In our experiments, we have noticed that solutions derived from the White-Patch algorithm [102] generally fail since the underwater images contain only reduced regions with specular reflection. Additionally, the solution of Gray-Edge algorithm [104] performs poorly in such cases, mainly due to the fact that underwater images are characterized by low contrast and less visible edges than natural images. However, we found that the most appropriate strategy is the Gray-World approach of Buchsbaum et al. [102]. One common problem noticed during our tests for most of the white-balancing techniques (observed either for the entire image or only for

small regions of the image) is the color-deviation [102] that appears when the illumination is poorly estimated. For instance in the underwater images, where the appearance is overall blue, the parts that are miss-balanced will show reddish appearance(that corresponds to the opponent color of the illumination).

Our approach minimizes this effect of color shifting for the entire scene. Related to these previous approaches, our solution is similar to the Shades-of-Grey [103] but much computationally effective. We found to be more robust to increase the average value estimated with a percentage λ instead to variate the norm value of p.

As a result, in our framework, the illumination is estimated by the value μ_I that is computed from the average of the scene μ_{ref} and adjusted by the parameter λ:

$$\mu_I = 0.5 + \lambda\mu_{ref} \tag{29}$$

The average color μ_{ref} is used to estimate the illuminant color (a common solution derived from Gray-World [102]) and can be obtained based on Minkowski norm when $p = 1$. Furthermore, to assign parameter λ we analyze the density and the distribution on the color histogram. Consequently, we set a higher value for λ when the detected set of colors is small. The value that variates in the range $[0, 0.5]$ of λ decreases inverse-proportionally with the number of colors. In general, we have observed that a default value of 0.2 yields visually pleasing results (since most of the processed underwater images present a relative uniform color distribution). Despite of its simplicity, our white balance strategy is able to remove effectively the color cast but also to recover the white and gray shades of the image, while producing a natural appearance of the output. Our method overcomes the standard Gray-World as well as the other considered techniques (please refer to the supplementary material for an extensive study of different white balance techniques applied for underwater images).

Practically, the ***first input*** of the fusion process is computed based on this straightforward white balancing operation. Nevertheless, white balancing solely is not able to solve the problem of visibility, and therefore we derive an additional input (described in the next subsection) in order to enhance the contrast of the degraded image.

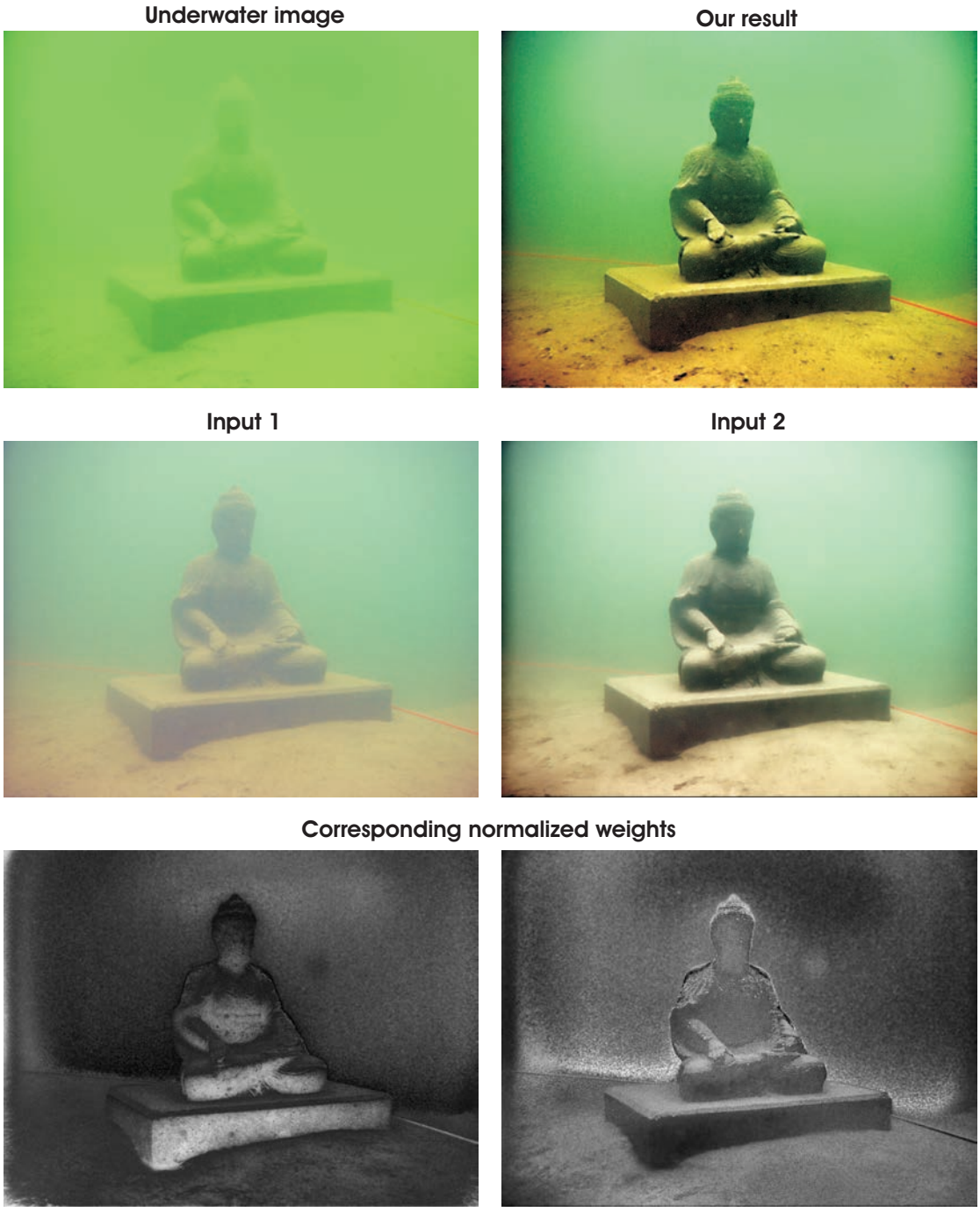

FIGURA 33: Top line: degraded image and our result; Middle line: the two inputs derived from the original image required by our fusion approach; Bottom line: the corresponding normalized weight maps $\bar{W}$.

Temporal Coherent Noise Reduction. Due to the impurities and the special illumination conditions, underwater images are noisy. Removing noise while preserving edges of an input image enhances the sharpness and may be accomplished by different strategies such as median filtering, anisotropic diffusion and bilateral filtering. However, for videos this task is more challenging since both spatial and temporal coherence need to be taken into account. The bilateral filter [105, 106]

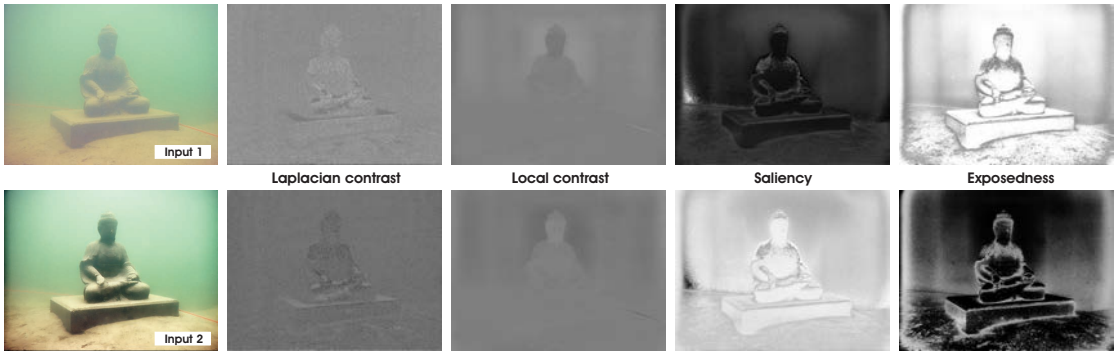

FIGURA 34: The two inputs derived from the original image presented in previous figure and the corresponding normalized weight maps.

is one of the common solutions being an non-iterative edge-preserving smoothing filter that has proven usefull for several problems such as tone mapping, mesh smoothing and dual photography enhancement. By considering the domain Ω of the spatial filter kernel f (Gaussian with standard deviation σ_f), the bilateral filter blends the center pixel s of the kernel with the neighboring pixels p that are similar to s:

$$J_s = \frac{1}{k(s)} \sum_{p \in \Omega} f(p - s, \sigma_f) g(D(p, s), \sigma_g) I_p \tag{30}$$

where $D(p, s) = I_p - I_s$ is the difference in intensities, the normalization $k(s) = \sum_{p \in \Omega} f(p - s, \sigma_f) g(D(p, s), \sigma_g)$,
g is the range kernel that is a Gaussian with standard deviation σ_g that penalizes pixels across edges that have large intensity differences.

However, the bilateral filter does not guarantee the preservation of the temporal coherence for videos. Even though, in the absence of motion, a simple average of all pixels at each coordinate through time would represent a decent solution, for real dynamic scenes this naive strategy yields undesirable ghosting artifacts. Inspired by Bennet et al. [90] where the solution was used in context of multispectral video fusion, we employ a *temporal bilateral filter* strategy on the white balanced version of the frames that aims to reduce noise and smoothing frames while preserving temporal coherence. Choosing an appropriate value of σ_g to simultaneously deal with noise while still preserving edges is difficult. High values of σ_g would yield halos while small values of σ_g are not able to reduce sufficiently undesired noise. Instead of just comparing the intensities as $D(p, s) = I_p - I_s$, we compute the sum of squared differences (SSD) between small spatial neighborhood Ψ around s and p weighted by a Gaussian $\Gamma(x, y)$:

$$D(p, s) = \sum_{x}^{\Psi} \sum_{y}^{\Psi} \Gamma(x, y)(I_p - I_s)^2 \tag{31}$$

Typically, the size of neighborhood Ψ is 3×3 or 5×5. This simple approach significantly reduces the ambiguity between noise and edges since the larger neighborhood Ψ reduces the impact of single-pixel temporal noise.

In our fusion framework, the ***second input*** is computed from the noise-free and color corrected version of the original image. This input is designed in order to reduce the degradation due to volume scattering. To achieve an optimal contrast level of the image, the second input is obtained by applying the classical contrast local adaptive histogram equalization [107]. To generate the second derived image common global operators can be applied as well. Since these are defined as some parametric curve, they need to be either specified by the user or to be estimated from the input image. Commonly, the improvements obtained by these operators in different regions are done at the expense of the remaining regions. We opted for the local adaptive histogram since it works in a fully automated manner while the level of distortion is minor. This technique expands the contrast of the feature of interest in order to simultaneously occupy a larger portion of the intensity range than the initial image. The enhancement is obtained since the contrast between adjacent structures is maximally portrayed. To compute this input several more complex methods, such as the gradient domains or gamma correction multi-scale Retinex (MSR) [108], may be used as well.

Weights of the Fusion Process. The design of the weight measures needs to consider the desired appearance of the restored output. We argue that image restoration is tightly correlated with the color appearance, and as a result the measurable values such as salient features, local and global contrast or exposedness are difficult to integrate by naive per pixel blending, without risking to introduce artifacts. Higher values of the weight determines that a pixel is advantaged to appear in the final image (see figure 34).

Laplacian contrast weight (W_L) deals with global contrast by applying a Laplacian filter on each input luminance channel and computing the absolute value of the filter result. This straightforward indicator was used in different applications such as tone mapping [41] and extending depth of field [56] since it assigns high values to edges and texture. For the underwater restoration task, however, this weight is not sufficient to recover the contrast, mainly because it can not distinguish between a ramp and flat regions. To handle this problem, we searched for an additional contrast measurement that independently assess the local distribution.

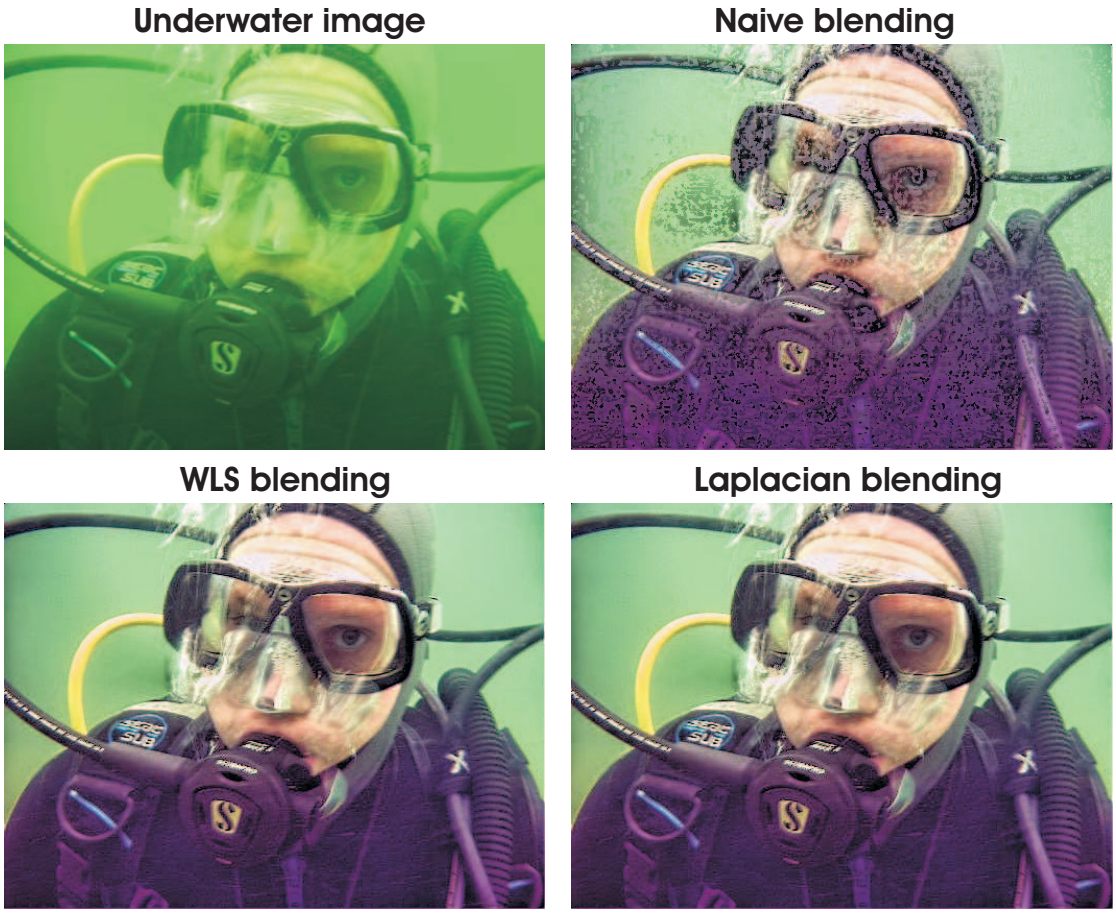

FIGURA 35: ***Blending strategies***. Considering the underwater image (top-left) by directly performing the naive blending (equation 49) yields unpleasing artifacts (top-right). On the other hand, by employing the multi-scale approaches based on WLS filter [Farbman et al 2008] (bottom-left) and Laplacian pyramid yield significant improvements. As may be observed, the difference between WLS and Laplacian pyramid is negligible.

Local contrast weight (W_{LC}) comprises the relation between each pixel and its neighborhoods average. The impact of this measure is to strengthen the local contrast appearance since it advantages the transitions mainly in the highlighted and shadowed parts of the second input. The (W_{LC}) is computed as the standard deviation between pixel luminance level and the local average of its surrounding region:

$$W_{LC}(x,y) = \left\| I^k - I^k_{\omega_{hc}} \right\| \tag{32}$$

where I^k represents the luminance channel of the input and the $I^k_{\omega_{hc}}$ represents the low-passed version of it. The filtered version $I^k_{\omega_{hc}}$ is obtained by employing a small 5×5 ($\frac{1}{16}[1,4,6,4,1]$) separable binomial kernel with the high frequency cut-off value $\omega_{hc} = \pi/2.75$. For small kernels the binomial kernel is a good approximation of its Gaussian counterpart, and it can be computed more effectively.

Saliency weight (W_S) aims to emphasize the discriminating objects that lose their prominence in the underwater scene. To measure this quality, we have employed the saliency algorithm of Achanta et al. [55]. This computationally efficient saliency algorithm is straightforward to be implemented being inspired by the biological concept of center-surround contrast. However, the saliency map tends to favor highlighted areas. To increase the accuracy of results, we introduce the exposedness map to protect the mid tones that might be altered in some specific cases.

Exposedness weight (W_E) evaluates how well a pixel is exposed. This assessed quality provides an estimator to preserve a constant appearance of the local contrast, that ideally is neither exaggerated nor understated. Commonly, the pixels tend to have a higher exposed appearance when their normalized values are close to the average value of 0.5. This weight map is expressed as a Gaussian-modeled distance to the average normalized range value (0.5):

$$W_E(x,y) = exp\left(-\frac{(I^k(x,y)-0.5)^2}{2\sigma^2}\right) \tag{33}$$

where $I^k(x,y)$ represents the value of the pixel location (x,y) of the input image I^k, while the standard deviation is set to $\sigma = 0.25$. This map will assign higher values to those tones with a distance close to zero, while pixels that are characterized by larger distances, are associated with the over- and under- exposed regions. In consequence, this weight tempers the result of the saliency map and produces a well preserved appearance of the fused image.

To yield consistent results, we employ the normalized weight values $\bar{W}$ (for an input k the normalized weight is computed as $\bar{W}^k = W^k / \sum_{k=1}^{K} W^k$), by constraining that the sum at each pixel location of the weight maps W equals one (the normalized weights of corresponding weights are shown at the bottom of figure 33).

Multi-scale Fusion Process. The enhanced image version $\mathcal{R}(x,y)$ is obtained by fusing the defined inputs with the weight measures at every pixel location (x,y):

$$\mathcal{R}(x,y) = \sum_{k=1}^{K} \bar{W}^k(x,y) I^k(x,y) \tag{34}$$

where I^k symbolizes the input (k is the index of the inputs - $K = 2$ in our case) that is weighted by the normalized weight maps $\bar{W}^k$. The normalized weights $\bar{W}$

are obtained by normalizing over all k weight maps W in order that the value of each pixel (x, y) to be constrained by unity value ($\sum \bar{W}^k = 1$).

As can be seen in figure 35 the naive approach to directly fuse (to apply directly equation 49) the inputs and the weights introduces undesirable halos. A common solution to overcome this limitation is to employ multi-scale linear [56, 109] or non-linear filters [57, 106]. The class of non-linear

filters are more complex and has shown to add only insignificant improvement for our task (applying WLS [57] yields minor improvements compared with Laplacian pyramid as depicted in figure 35). Since it is straightforward to implement and computationally efficient, in our experiments the classical multi-scale Laplacian pyramid decomposition [56] has been embraced. In this linear decomposition, every input image is represented as a sum of patterns computed at different scales based on the Laplacian operator. The inputs are convolved by a Gaussian kernel, yielding a low pass filtered versions of the original. In order to control the cut-off frequency, the standard deviation is increased monotonically. To obtain the different levels of the pyramid, initially we need to compute the difference between the original image and the low pass filtered image. From there on, the process is iterated by computing the difference between two adjacent levels of the Gaussian pyramid. The resulting representation, the Laplacian pyramid, is a set of quasi-bandpass versions of the image.

In our case, each input is decomposed into a pyramid by applying the Laplacian operator to different scales. Similarly, for each normalized weight map $\bar{W}$ a Gaussian pyramid is computed. Considering that both the Gaussian and Laplacian pyramids have the same number of levels, the mixing between the Laplacian inputs and Gaussian normalized weights is performed at each level independently yielding the fused pyramid:

$$\mathcal{R}^l(x, y) = \sum_{k=1}^{K} G^l \left\{ \bar{W}^k(x, y) \right\} L^l \left\{ I^k(x, y) \right\} \tag{35}$$

where l represents the number of the pyramid levels (typically the number of levels is 5), $L\{I\}$ is the Laplacian version of the input I, and $G\left\{\bar{W}\right\}$ represents the Gaussian version of the normalized weight map $\bar{W}$. This step is performed successively for each pyramid layer, in a bottom-up manner. The restored output is obtained by summing the fused contribution of all inputs.

The Laplacian multi-scale strategy performs relatively fast representing a good trade-off between speed and accuracy. By independently employing a fusion process at every scale level the potential artifacts due to the sharp transitions of the weight maps are minimized. Multi-scale fusion is motivated by the human visual system that is primarily sensitive to local contrast changes such as edges and corners.

Results and Discussion. The proposed strategy was tested for real underwater videos and images taken from different available amateur photographer collections. As a result, images and videos have been captured using various cameras and setups. However, an important observation is that we process only 8-bit data format even though many professional cameras have the option to shoot in the RAW mode that usually stores the unprocessed data of the camera's sensor in 12-bit format.

Our technique is computationally effective taking approximately 2 seconds (Matlab code) for a 800×600 frame but we believe that an optimized implementation could run real-time on common hardware. The reader is referred to the supplementary material for additional results (images and videos). By a general visual inspection it can be observed that our technique is able to yield accurate results with enhanced global contrast, color and fine details while the temporal coherence of the videos is well preserved.

In figure 36 we compare our results with the polarization technique of Schechner and Averbuch [89] that uses two frames taken with wide-field polarized illumination. By employing our technique on the provided white balanced version of one of their inputs we are able to produce a more pleasing image version.

In Figure 37 we compare our technique with several specialized underwater enhancing techniques. We considered the specialized single underwater image enhancing techniques [95] but also the recent specialized dehazing technique [17]. By a closer inspection (please observe as well the middle row of Figure 37) our result presents less halos and color distortions. For this example to visualize how contrast is modified we employed the IQM metric [110] that was originally developed to evaluate tone mapping operators. This metric utilizes a model of the human visual system being sensitive to three types of structural changes: loss of visible contrast (green), amplification of invisible contrast (blue) and reversal of visible contrast (red). As a general remark, compared with the other considered approaches, the

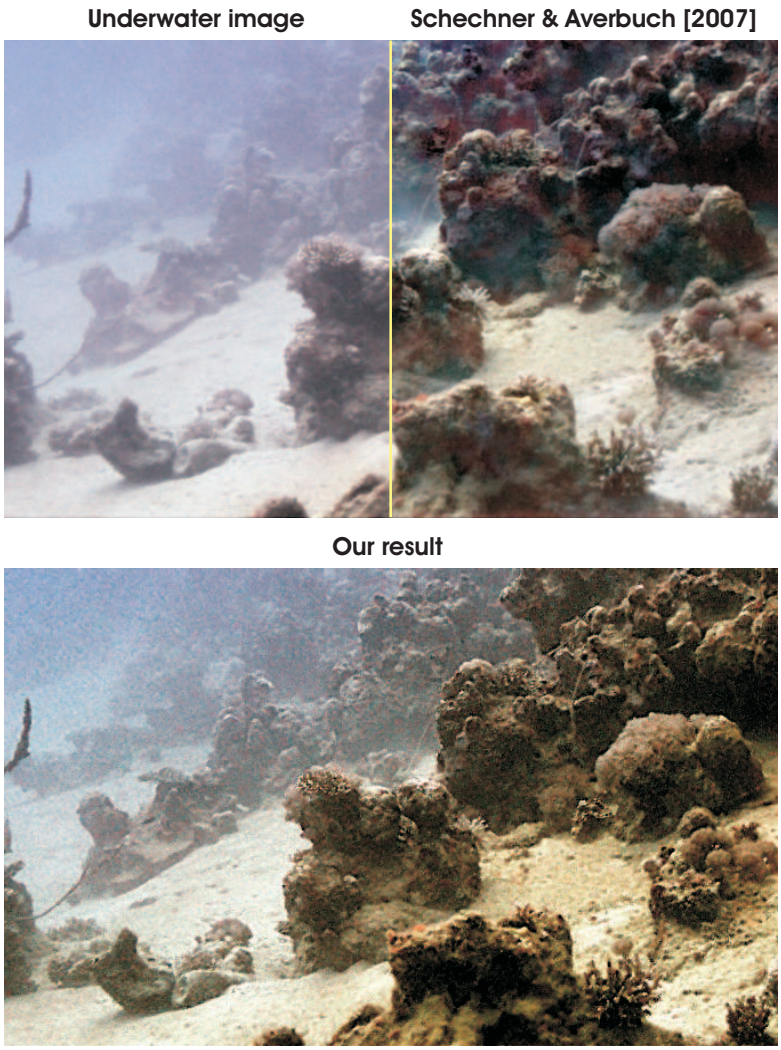

FIGURA 36: ***Comparison with polarization methods of [Schechner and Averbuch 2007].*** We applied our technique on the white balanced version of one of the employed inputs provided by the authors.

most predominant structural change characteristic to our method is the amplification of the contrast (blue) and only very few locations exhibit reverse (red) and loss (green) of the contrast.

Our technique shown limitations when dealing with images of very deep scenes taken with poor strobe and artificial light.In such cases, even some enhancement could be obtained, the bluish appearance however still remains. Moreover, when the illumination is poor the very distant parts of the scene cannot be recovered reliably. The restoration of distant objects and regions represents also a general limitation of our approach compared with hardware and polarization-based techniques that in general perform better in such cases due to the additional available information.

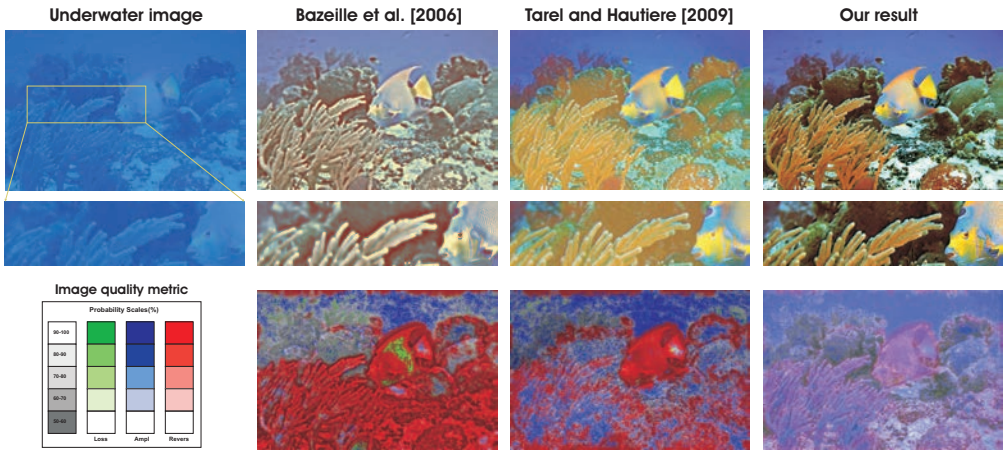

FIGURA 37: ***Comparative results.*** Compared with the outputs of other enhancing methods our result is less prone to halos and color distortions. This also results (bottom row) when applied IQM metric [110]. As may be observed, based on this metric our approach mainly amplifies the contrast (blue).

Applications. We found our technique suitable for several other applications that are briefly described in the following section. More results are included as well in the supplementary material.

Matching images by local feature points is a fundamental task of many computer vision applications. We employ the SIFT [111] operator for an initial pair of underwater images and as well for the restored versions of the images (see figure 38). We use the original implementation of SIFT applied exactly in the same way in both cases. For the initial case SIFT filters 3 good matches and one mismatch while the matching of the enhanced image versions yields 43 valid matches and no mismatches. These promising achievements demonstrate that our technique does not introduce artifacts but mainly restores both global contrast and local features of underwater images.

FIGURA 38: ***Local feature points matching.*** Compared with the initial images (top) by applying standard SIFT on our enhanced versions (bottom) the matching result is improved considerable.

Segmentation aims to divide images into disjoint and homogeneous regions with respect to some characteristics (e.g. texture, color). In this work we employ the $GAC++$ [112] that represents a state-of-the-art geodesic active contours method (variational PDE). Figure 39 proves that by processing underwater images with our approach the segmentation result is more consistent while the filtered boundaries are perceptually more accurate. This task demonstrates that our technique does not introduce halos close to object boundaries.

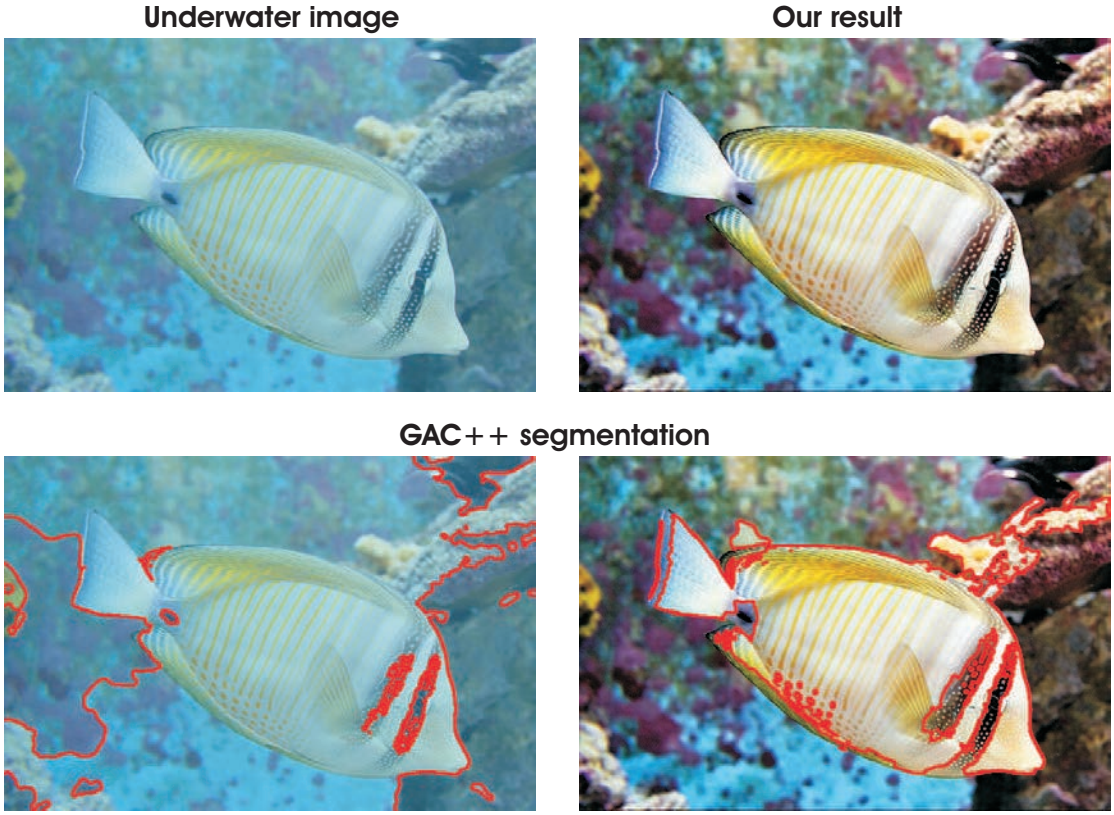

FIGURA 39: ***Image segmentation.*** Processing underwater images with our method, the segmentation result is more consistent while the filtered boundaries are perceptually more accurate.

Image dehazing [13, 59] is the process of removing the haze and fog effects from the spoilt images. Because of similarities between hazy and underwater environments due to the light scattering process, we found our strategy appropriate for this challenging task. However, as explained previously, since the underwater light propagation is more complex we believe that image dehazing could be seen as a subclass of the underwater image restoration problem. Comparative results with state-of-the-art single image dehazing techniques [13, 16] are shown in Figure 40.

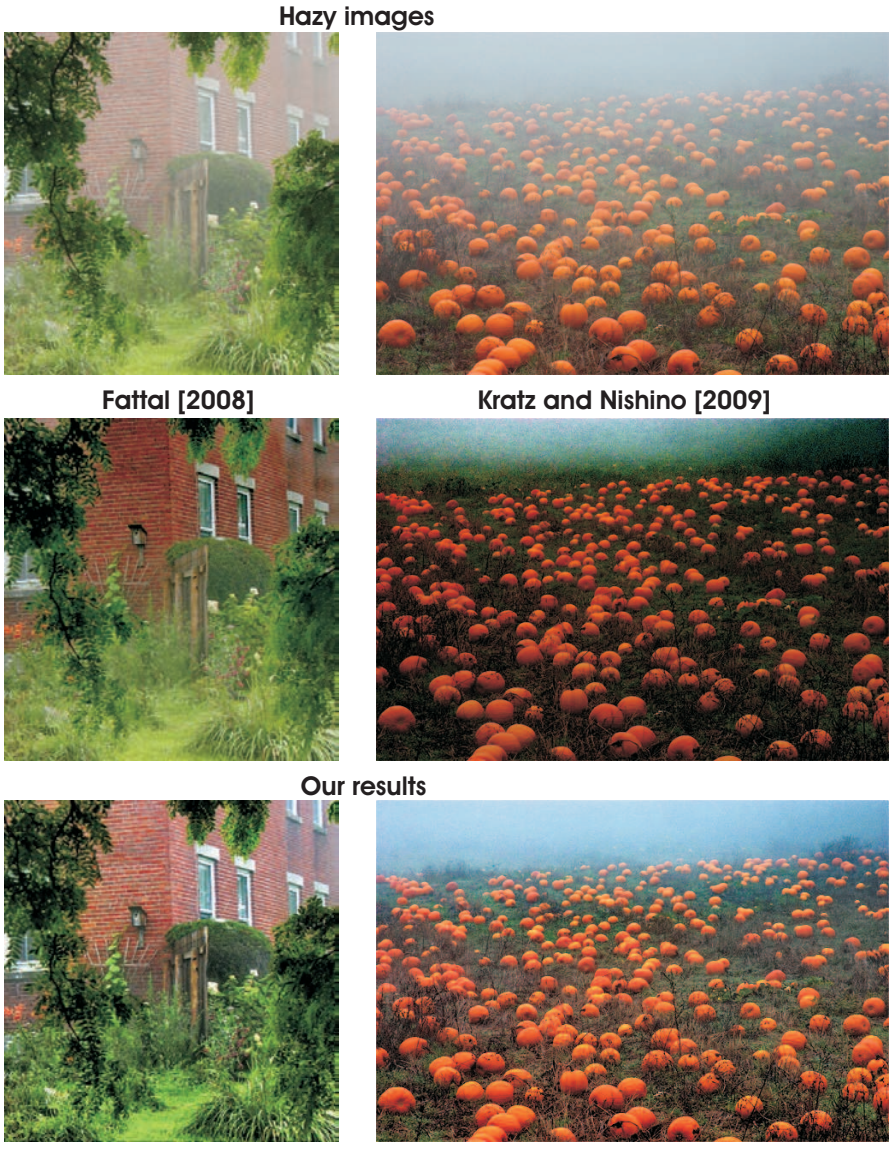

FIGURA 40: ***Single image dehazing.*** Our approach is able to dehaze images comparable to and even more accurate than the specialized single dehazing techniques.

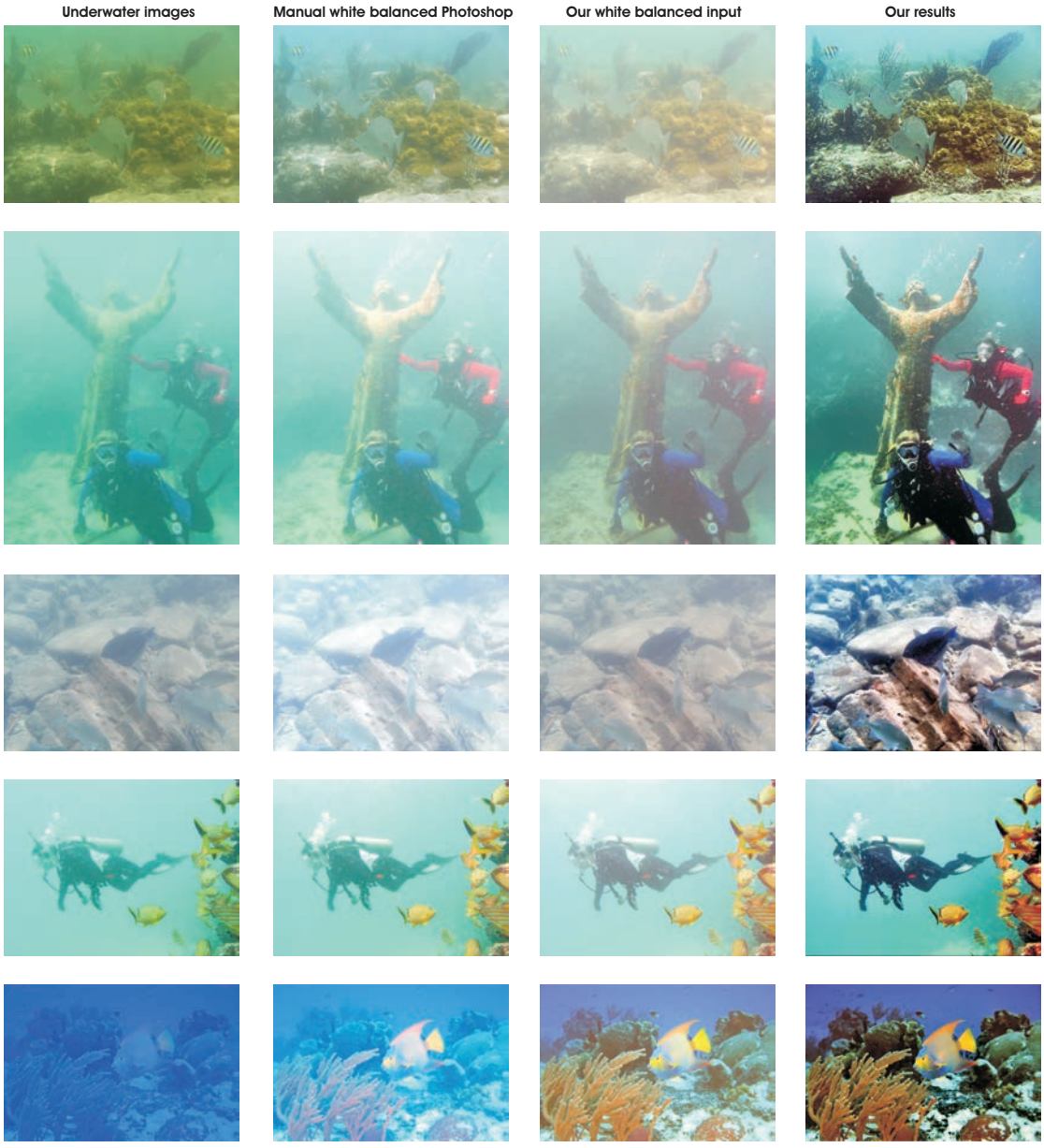

FIGURA 41: From left to right: original underwater images, white balanced results obtained by manually selecting the whitest point in Photoshop, white balance used to generate our algorithm inputs and our results.

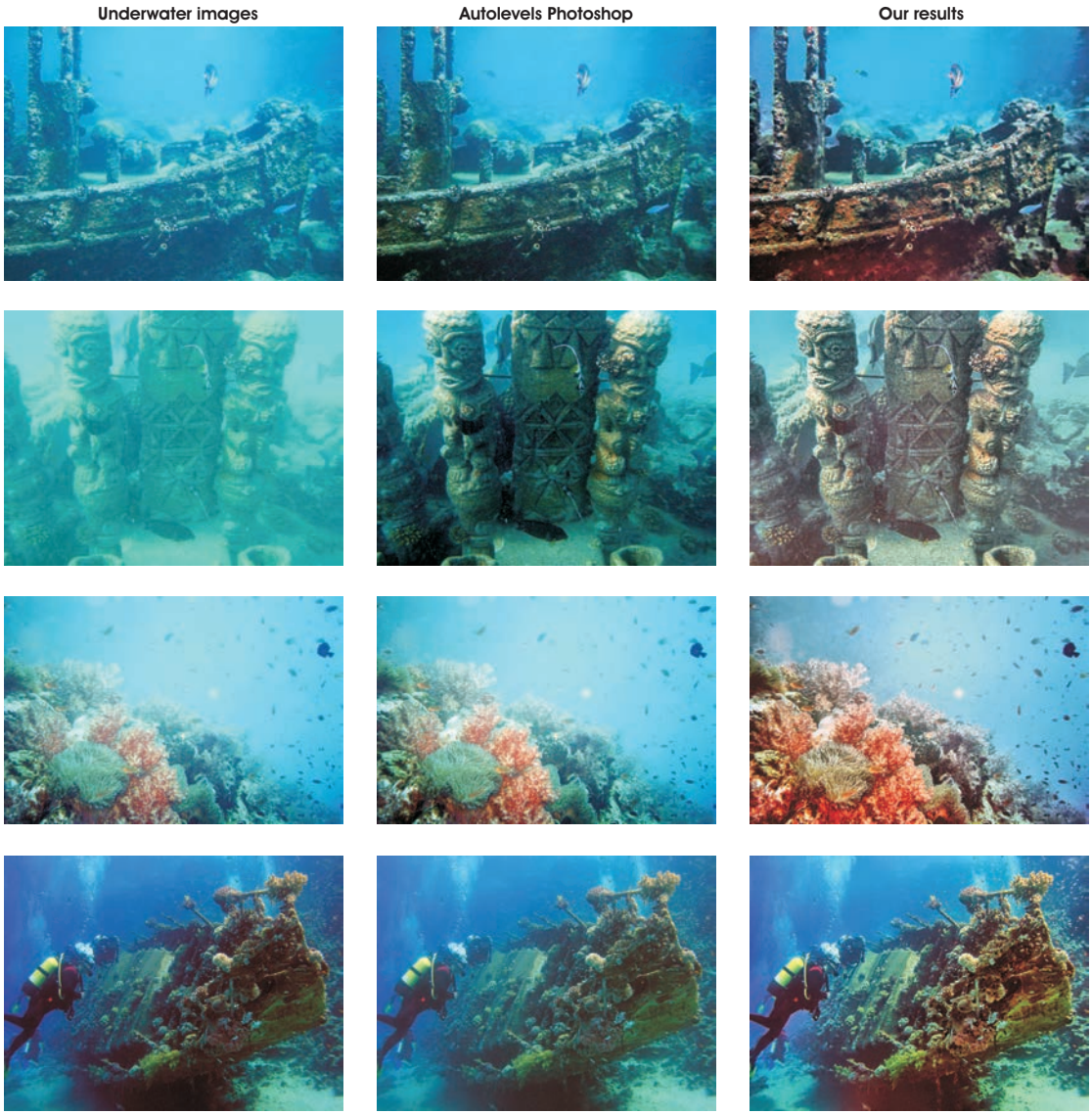

FIGURA 42: From left to right: original underwater images, images enhanced using Autolevel operation of Adobe Photoshop and our results.

FIGURA 43: ***Image dehazing.*** Comparison against state-of-the-art single image dehazing methods (except for ***Deep photo*** - the method of Kopf et al. [2008] uses additionally an approximated 3D model of the scene).

FIGURA 44: ***Matching by local feature points.*** For the first pair of images by applying standard SIFT operator on the initial images only 3 good matches (blue-yellow crosses) and one mismatch (red cross) are found. On the other hand, applying SIFT in exactly the same way on the restored versions of the images yielded by our technique, yields 43 valid matches (blue-yellow crosses) and no mismatches. For the second pair of images, SIFT applied on the original underwater images filters 4 good matches while applied on the preprocessed images the same operator yields 14 valid matches.

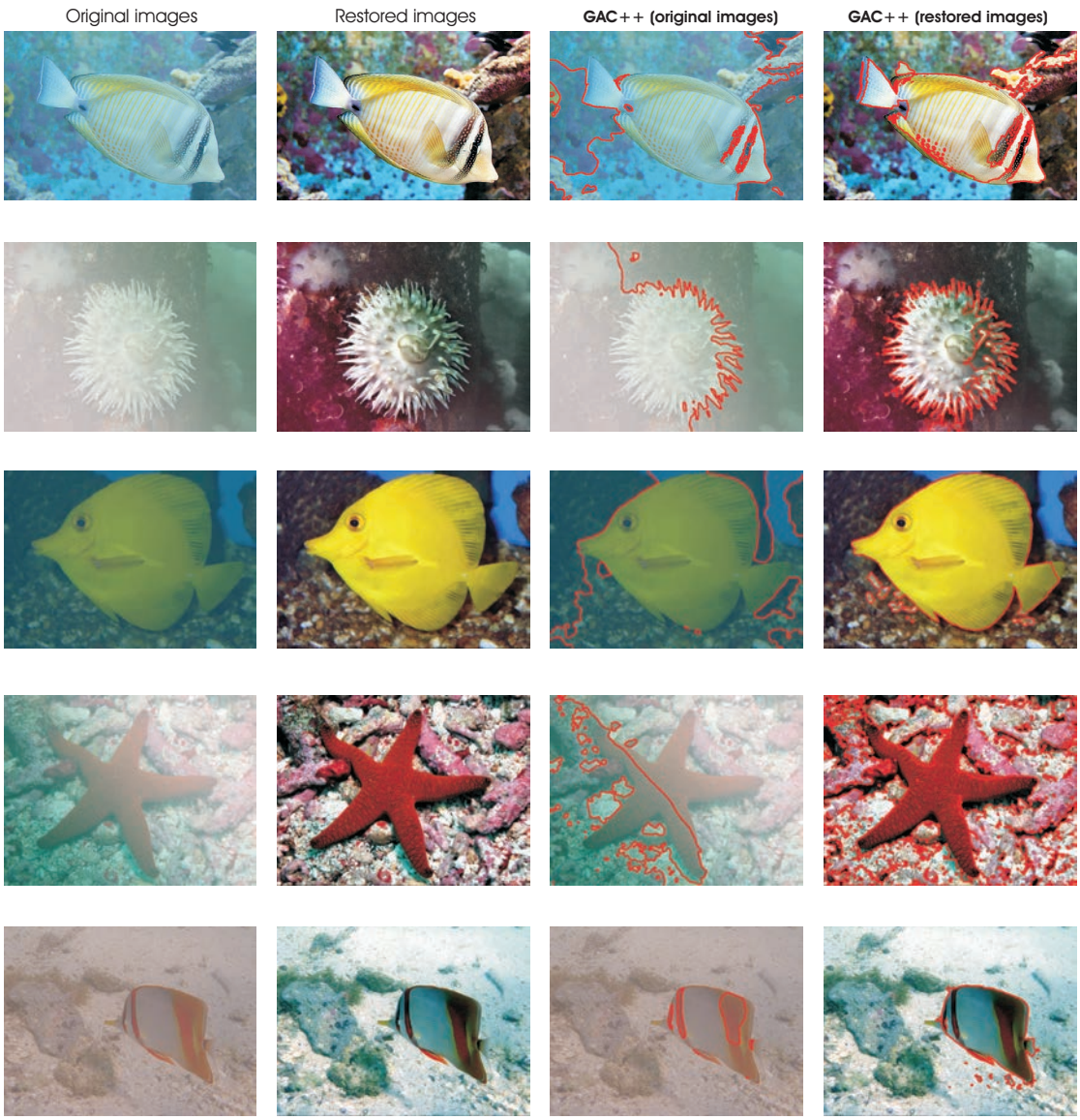

FIGURA 45: ***Image segmentation.*** We employed the geodesic active contours segmentation method (GAC++) for both original and preprocessed images. As can be observed by preprocessing underwater images with our approach the segmentation is more consistent while the filtered boundaries are perceptually more accurate.

II.2.2 Multi-scale Underwater Descattering

In this section we introduce a novel approach based on local estimation of the back-scattering influence. The technique has been published recently by the candidate in [5]. Underwater imaging is required in many applications [93] such as control of underwater vehicles [113], marine biology research [114], inspection of the underwater infrastructure [115] and archeology [116]. However, as compared with computer vision and image processing applications in the surface environment, image analysis underwater is a much more difficult problem, awing to the dense and strongly non-uniform medium where light scatters, i.e. is forced to deviate from its straight trajectory. The poor visual quality of underwater images is mainly due to the attenuation and back-scattering of illumination sources. Back-scattering refers to the diffuse reflection of light, in the direction from which it emanated.

Early underwater imaging techniques employed specialized hardware [88] and multiple images polarized over diverse angles [117], resulting in either expensive or impractical acquisition systems. Recently, inspired by outdoor dehazing [1, 2, 14, 15, 59, 69, 70], several single-image based underwater image enhancement solutions [4, 62, 118–121] have been introduced. Chiang and Chen [119] first segment the foreground of the scene based on a depth estimate resulting from the Dark Channel Prior (DCP) [15, 67], then perform color correction based on the amount of attenuation expected for each light wavelength. Galdran et al. [121] introduce the Red Channel to recover colors associated with short wavelengths in underwater. Ancuti et al. [4] derives two color corrected inputs and merge them using a multi-scale fusion technique [56]. While the technique proposed in this part is also based on a multi-scale fusion strategy, here, we derive three distinct inputs that are robust in the presence of highly non-uniform illumination of the scenes (see Fig. 46).

Despite these recent efforts, existing single-image underwater techniques exhibit significant limitations in the presence of turbulent water and/or artificial ambient illumination (see Fig. 22). This is mainly due to poor estimation of the back-scattered light, which is generally assumed to be uniform over the entire image. A unique global value of back-scattered light is only valid in relatively simple underwater scenes having nearly uniform illumination, as is encountered in most outdoor hazy scenes.

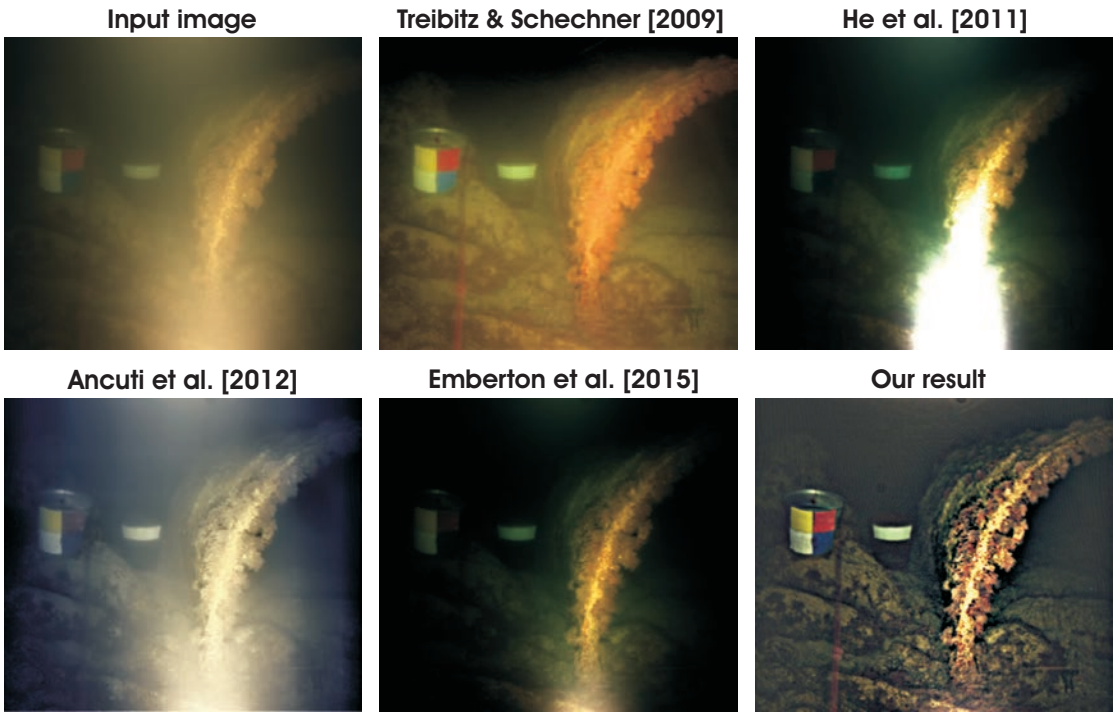

FIGURA 46: ***Underwater scene restoration.*** Special-purpose single-image dehazing method of He et al. [67] and also specialized underwater dehazing methods of Ancuti et al. [4] and Emberton et al. [62] are limited in their ability to recover the visibility of challenging underwater scenes. While the polarization-based technique (uses multiple images) of Treibitz and Schechner [117] is competitive, our approach better restores both color and contrast (local and global) in the underwater image.

Following the optical underwater model [101], we first compute the back-scattered light by searching for the brightest location along each image patch. By simply inverting the optical model using our local estimate of the back-scattered light, we are able to obtain a good degree of visual restoration, even on in extreme underwater scenes. Since the size of the patch should depend on multiple parameters characterizing the captured underwater scene (e.g. the dimensions and colors of objects in the scenes, nature of the ambient light, non-uniform illumination) we use a variety (only two in practice) of patch dimensions, each size supporting the recovery of distinct and complementary features of the scene. Specifically, a first image is derived using a smaller patch to better restore the contrast, while a second image is derived based on a large patch size, which makes it possible to consistently recover regional color. To highlight and enhance the fine details of the initial image, we also compute its discrete Laplacian. These three images derived from the original image, guided by several quality measures, are seamlessly blended using a multi-scale fusion approach [56].

We perform an extensive qualitative and quantitative evaluation against several

existing specialized techniques. Despite its simplicity, our method proves quite robust and produces competitive results on highly opaque and non-uniformly illuminated underwater scenes.

Underwater Image Formation. Based on the well-known optical model of McGlamery [101], in the underwater medium the total radiance of an image $\mathcal{I}$ that reaches the observer is due to three additive components: a *direct component* E_D, a *forward-scattering* component E_{FS} and a *back-scattering* component E_{BS}. These components mainly result from the radiances of objects in the scene and the ambient light.

The *direct component* E_D represents the attenuated version (with distance) of the reflected light, and is expressed at each image coordinate x as:

$$E_D(x) = J(x)e^{-\eta d(x)} = J(x)t(x) \tag{36}$$

where $J(x)$ is the radiance of the object, $d(x)$ is the distance between the observer and the object, and η is the attenuation coefficient. The exponential term $e^{-\eta d(x)}$ is also known as the transmission t through the underwater medium.

Forward-scattering E_{FS}, is the deflection of a portion of the incident light. In general, it is associated with a small fraction of the overall image degradation process.

Back-scattering, also known as the *veiling light* [100], is the principal cause for the loss of contrast and the color shifting of underwater images. For reasonable distances (between 3-10 m) this component may be expressed as [100]:

$$E_{BS}(x) = B_\infty(1 - e^{-\eta d(x)}) \tag{37}$$

where B_∞ is a scalar known as the *back-scattered light* or the *water background* [122]. Assuming homogeneous lighting along the line of sight, this component may be regarded as originating from an equivalent sources at infinity [100].

Incorporating these additive components and ignoring the forward scattering component, the simplified underwater optical model employed in most existing descattering techniques, becomes:

$$\begin{aligned} \mathcal{I}(x) &= J(x)e^{-\eta d(x)} + B_\infty(1 - e^{-\eta d(x)}) \\ &= J(x)t(x) + B_\infty(1 - t(x)) \end{aligned} \tag{38}$$

Since the underwater camera model (38) has a similar form as the optical model of Koschmieder [18], used to characterize the propagation of light in the atmosphere, many recent approaches have proposed to restore underwater images based on the extension of popular and effective outdoor dehazing methods. For instance, Chiang and Chen [119] estimate the rough depth of the underwater scene based on the Dark Channel Prior (DCP) [67], then adjust the bluish tone based on a wavelength compensation strategy. Similarly, Galdran et al. [121] propose a variation of DCP, in which the so-called Red Channel is used to recover colors associated with short wavelengths underwater.

However, these outdoor dehazing-derived techniques appear to be successful mostly on less challenging underwater scenes. They generally assume shallow underwater scenes with relatively transparent water and effective ambient illumination. These cases, which are indeed similar to outdoor hazy scenes, represent only a fraction of the underwater imaging problem. At greater depths and under artificial illumination, the visibility degradation is more critical and dehazing approaches suffer from important limitations when restoring contrast and color (see Fig. 22). We explain in the rest of this part how (38) can be inverted in the case of arbitrary underwater scenes.

Estimation of Transmission and Back-scattering. There are two unknowns in the image acquisition model defined by equation (38): the transmission map $t(x)$ and the back-scattered light B_∞.

Following recent techniques, we approximate the transmission $t(x)$ based on the Dark Channel Prior (DCP) of He et al. [67]. This prior assumes that natural objects have a weak reflectance in one of the color channels. In other words, the direct radiance is small, or dark, in at least one of the R,G,B color channels. Given this assumption, the transmission map, $t(x)$, can be estimated from the weakest color over a neighborhood of x. Formally, the DCP asserts that $\min_{y\in\Omega(x)} (\min_{c\in r,g,b} \mathcal{J}^c/{B_\infty}^c) = 0$. Hence, the optical model (38) yields:

$$t(x) = 1 - \min_{y\in\Omega(x)} \left(\min_{c\in r,g,b} \mathcal{I}^c(x)/{B_\infty}^c \right) \tag{39}$$

where $\Omega(x)$ represents a local patch centered at x.

In practice, $t(x)$ is reasonably well approximated by replacing the back-scatteredlight B_∞ with the maximal color intensity vector $[1, 1, 1]$, so that:

$$t(x) \approx \quad 1 - \min_{y \in \Omega(x)} \left(\min_{c \in r,g,b} \mathcal{I}^c(x) \right) \tag{40}$$

In the literature [67], the term $\min_{y \in \Omega(x)} (\min_{c \in r,g,b} \mathcal{I}^c(x))$ is referred to as the dark channel image, and is denoted $\mathcal{I}_{\mathcal{DC}}(x)$.

Global Back-scattering Estimation. The other unknown that is required to invert the optical underwater acquisition model is the back-scattered light B_∞, which represents the light that is scattered back by floating particles. This is the main cause of contrast and color degradation in underwater images.

Most existing underwater descattering strategies compute a global value of the back-scattered light over the entire image. Similar to what is done in outdoor dehazing [13, 14, 67], this value is usually determined in the brightest region of the dark channel image [120, 121]. This is because, when $I_{DC}(x) \to 1$, equation (40) implies that $t(x) \to 0$, and the optical model then states that $I(x) \to B_\infty$.

Formally, following He et al. [67], the back-scattered light can be estimated as:

$$\begin{aligned} &B_\infty = \mathcal{I}(y*), \ with \\ &y* = \underset{y | \mathcal{I}_{\mathcal{DC}}(y) > \mathcal{I}_{\mathcal{DC}}{}^{99.9}}{\boldsymbol{arg\,max}} \left(\mathcal{I}^r(y) + \mathcal{I}^g(y) + \mathcal{I}^b(y) \right) \end{aligned} \tag{41}$$

In this equation, $y*$ denotes the location of the brightest pixel among those pixels whose dark channel value lies above the 99.9^{th} percentile $\mathcal{I}_{\mathcal{DC}}{}^{99.9}$, while r, g, b refer to the red, green and blue color components, respectively. The above equation can be written on a color component basis, as:

$$B^c_\infty \simeq \max_{y | \mathcal{I}_{\mathcal{DC}}(y) > \mathcal{I}_{\mathcal{DC}}{}^{99.9}} \mathcal{I}^c(y), \tag{42}$$

where c corresponds to each one of the color channels ($c = [r, g, b]$) and $\mathcal{I}_{\mathcal{DC}}{}^{99.9}$ denotes the 99.9^{th} percentile of the dark channel over the entire image.

To achieve robustness to specular reflections and glowing effects that tend to mislead the color estimator defined by (41) and (42) in underwater mediums [123], we introduce an alternative definition of the global back-scattered light estimator that finds and uses the minimum of $\mathcal{I}^c$ over a small neighborhood around y.

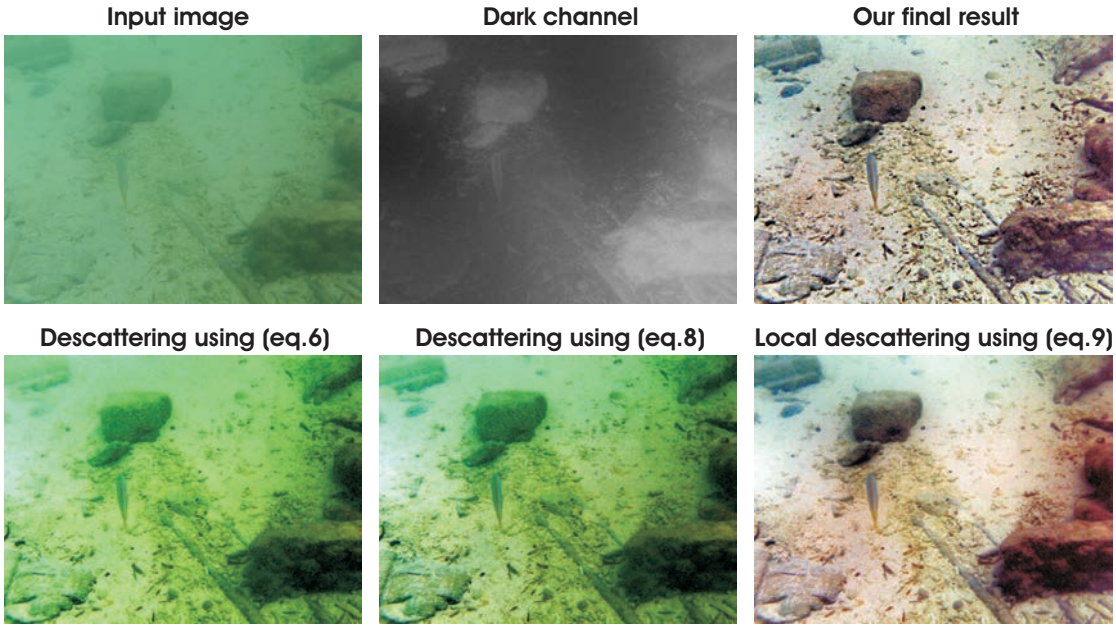

FIGURA 47: ***Global vs. Local descattering in uniform underwater medium.*** Compared with global estimates of the back-scattering, the strategy based on our patch-based estimation of back-scattering allows for better restoration of visibility, even in a uniform underwater medium.

Formally, we define:

$$B^c_{G_\infty} = \max_{y \in M^I_{DC}} \left(\min_{z \in \Omega(y)} \mathcal{I}^c(z) \right) \tag{43}$$

where $\Omega(y)$ is a neighborhood around y, and M^I_{DC} is the set of locations in the image support $\mathcal{I}$ where the dark channel reaches its global maximum value ${\mathcal{I}_{\mathcal{DC}}}^{max}$, i.e. $M^I_{DC} = \{y | \mathcal{I}_{\mathcal{DC}}(y) = \mathcal{I}^{max}_{DC}\}$. Hence, instead of keeping an arbitrary 99.9^{th} percentile, we retain the entire set of coordinates where $\mathcal{I}_{\mathcal{DC}}$ is maximum. Since the dark channel $\mathcal{I}_{\mathcal{DC}}(x)$ is defined based on minimization over a patch, M^I_{DC} always include multiple locations. Our validations reveal that using M^I_{DC} (instead of the 99.9^{th} percentile), does not negatively impact the global estimation. Moreover, it makes the expression easier to generalize to a local estimator, as explained in the next sub-section. Before moving to our proposed local estimator, we first point out the equivalence and the limitations of the two global estimators defined by (41) and (43). As shown in Fig. 47, expressions (41) and (43) yield comparable results when inverting the optical model (38), and provide satisfying approximations in an underwater medium with relatively uniform illumination. However, as illustrated in Fig. 48, both global estimators fail in the presence of non-uniform illumination.

Local Back-scattering Estimation. The problem of non-uniform back-scattering becomes especially challenging at greater depths, where artificial illumination is

FIGURA 48: ***Global vs. Local descattering in non-uniform underwater medium.*** While global estimation of the back-scattering performs poorly in such challenging underwater scenes, the same restoration strategy yields consistent improvement by using our local estimation of back-scattering.

required. Local estimation of the back-scattered light has been recently considered for underwater dehazing by Emberton et al. [62]. They designed a hierarchical rank-based method, using a set of features to find those image regions which are the most haze-opaque. Here, we introduce an alternative approach to be able to adapt locally to the back-scattered light. Our proposal achieves visually improved descattering, as attested by the results presented in Fig. 50 and Fig. 52. Starting from equation (43), we propose to compute the back-scattered light at location x as:

$$B^c_{L_\infty}(x) = \max_{y \in M^{\Psi(x)}_{DC}} \left(\min_{z \in \Omega(y)} \mathcal{I}^c(z) \right) \tag{44}$$

where $\Psi(x)$ is a square patch centered at x. As for the global estimator, $M^{\Psi(x)}_{DC}$ denotes the set of positions in $\Psi(x)$ where the dark channel is maximum, i.e. reaches the value of its local maximum over $\Psi(x)$. In practice, the patch Ψ is typically larger than Ω. We use a default value of 2 for the ratio between the sizes of the Ψ and Ω patches. A higher ratio is recommended for processing underwater scenes with relatively uniform illumination and less turbidity. As shown in Figs. 47 and 48, the contrast and color are better recovered when inverting the underwater acquisition model based on our proposed local back-scattering estimator as compared to inverting based on the global method.

Descattering by Multi-scale Fusion. A critical issue related to equation (44) lies in the selection of the patch size Ω. Using a large patch tends to reduce the impact of specular reflections and glowing effects in underwater mediums, but also tends to preserve the undesirable hazy appearance with reduced improvement of the contrast as compared to using a smaller patch (see Fig. 49).

Inspired by our previous approach [4], of keeping the best from among small and large patches, we propose a multi-scale fusion approach to blend the images obtained with both a large and a small patch size. Image fusion is a well-known process that aims to optimize the appearance of a reconstructed scene by effectively blending the information of multiple inputs. Multi-scale fusion based on the

Laplacian pyramid [56] has been shown to be effective for various computational imaging applications such as extended depth-of-field [36], image editing [37], image compositing [39], HDR imaging [41], image decolorization [6, 68] and single image dehazing [1, 58]. The fusion process is typically guided by several spatial weight maps that capture, the contribution of each input image to the final output. These weight maps typically assess contrast, saturation and saliency.

Our fusion technique is a single image-based approach that derives several inputs from an original underwater input. First, we derive inputs based on our patch-based estimation of the back-scattering influence, as discussed in the previous section. We observed that choosing different patch sizes yields good recovery either of contrast (small patch) or color (large patch). Since we cannot simultaneously obtain good restoration of both of these important factors using a single patch size, we derive two inputs generated using two different patch sizes.

Using a large patch size better restores color while choosing a small patch size is more effective in restoration of the visibility (global contrast).

As a result, to remove most of the hazy appearance of the underwater images we derive a **first input** based on our local back-scattering estimation (44) computed using a small patch size (e.g. 20×20 for an image of size 800×600). Moreover, to better restore color we derive, using the same approach, a **second input** using a larger patch size (e.g. 60×60 for an image of size 800×600). Our experiments have revealed that considering additional patch sizes does not deliver any added value to the reconstruction.

However, as can be seen in Fig. 49, those two inputs fail to capture the finest detail. To also transfer this important information to the final result we derive a **third input** which is the discrete Laplacian of the original image.

Inspired by our previous fusion underwater approach [4], we derive three weight maps. This ensures that locations of high contrast or high saliency will receive greater emphasis in the fusion process.

Local contrast weight detects the degree of local variation of each derived input. Estimated similarly as in [1, 41] it assigns high values to edges and texture variations, by computing the L_1 norm of Laplacian filter applied to the luminance channels of each input .

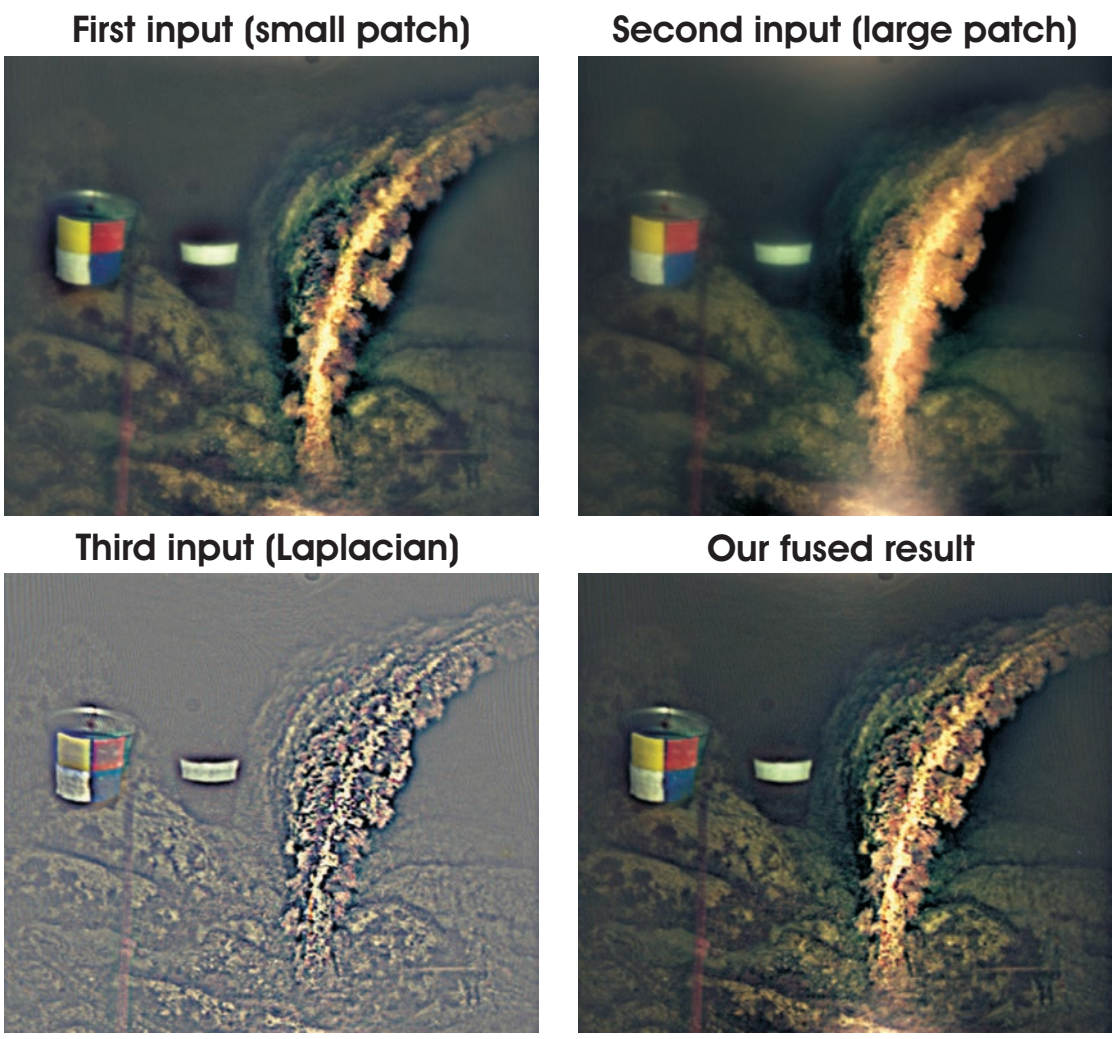

FIGURA 49: ***Derived Inputs.*** Our single image-based fusion algorithm considers three derived inputs. The first one estimates back-scattering on a small patch and removes most of the hazy appearance, the second estimates back-scattering on a large patch and primarily restores the color, and the third computes the Laplacian of the original to preserve fine details.

Saturation weight map is motivated by the fact that humans generally prefer images containing a high level of color saturation. This measure is computed as the standard deviation across color channels at each pixel coordinate.

Saliency weight map advantages the most conspicuous regions of an image. This weight is computed using the saliency technique of Achanta et al. [55].

Each of those weight maps is scaled so that its range lies between 0 and 1. A pixelwise product of those three normalized maps is then used to derive a single map.

Since naive fusion implementation (directly blending the inputs and weight maps) causes unpleasing halo artifacts [4], we instead employ the multi-scale Laplacian

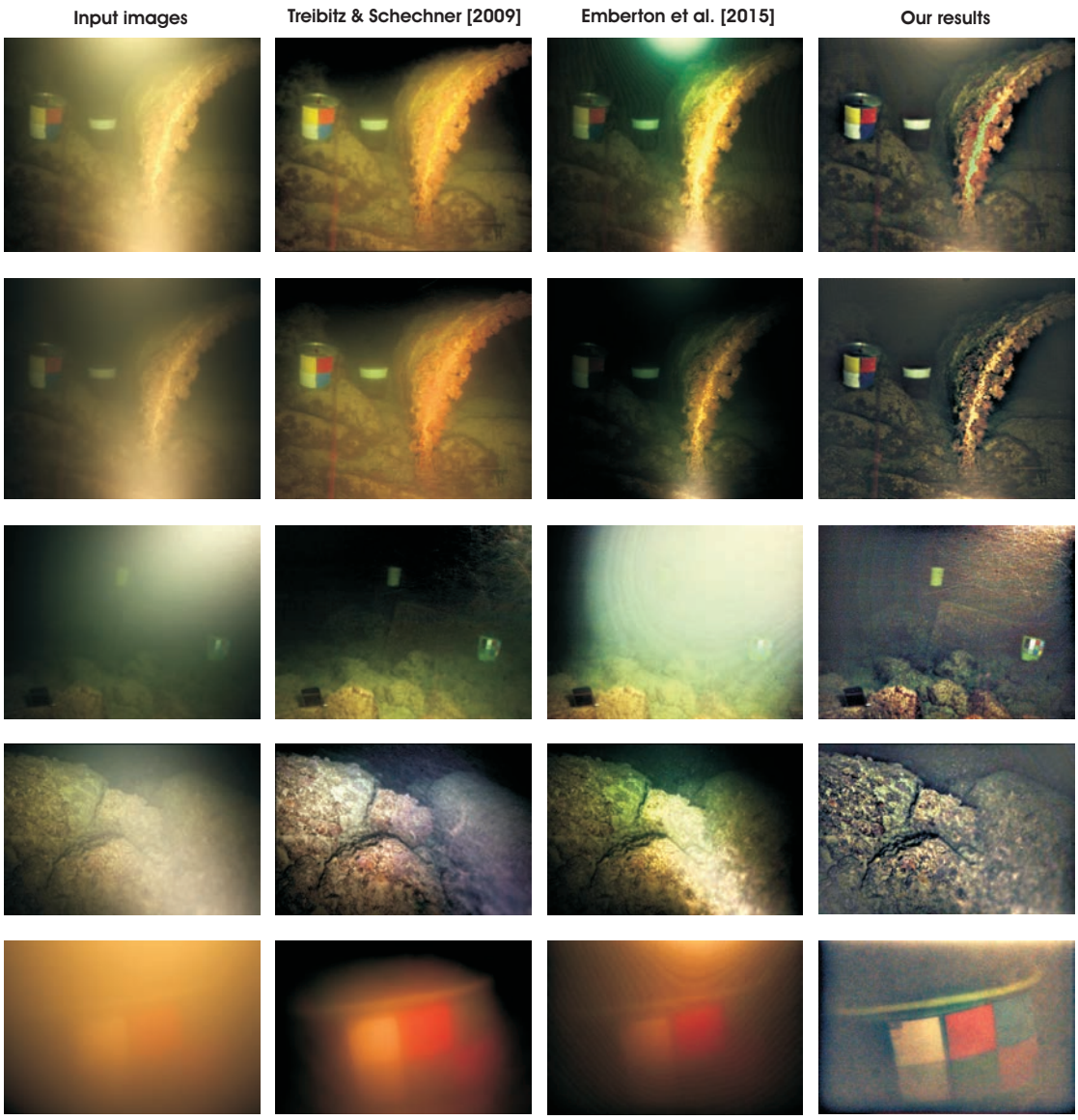

FIGURA 50: Comparative results on underwater scenes with non-uniform lighting.

decomposition [56]. Using the same number of levels, the Gaussian and Laplacian pyramids are independently fused at each level:

$$\mathcal{R}_l(x) = \sum_k G_l\left\{\bar{\mathcal{W}}^k(x)\right\} L_l\left\{\mathcal{I}_k(x)\right\} \tag{45}$$

where l is the level of the pyramid, k is the input index, $L\{\mathcal{I}\}$ denotes the Laplacian of the input $\mathcal{I}$, and $G\left\{\bar{\mathcal{W}}\right\}$ is the Gaussian-smoothed normalized weight map $\bar{\mathcal{W}}$. Here, the normalization ensures that the sum of weights over the three inputs is equal to unity at each pixel. The final fused result $\mathcal{R}$ is processed by summing the contributions from all the computed levels of the pyramid.

Results and Discussion. Our approach has been extensively tested on a large

	Carlevaris-B. et al.		He et al.		Drews-Jr et al.		Galdran et al.		Emberton et al.		Our method	
	VM	PCQI	VM	PCQI	VM	PCQI	VM	PCQI	VM	PCQI	VM	PCQI
Shipwreck	53.23	0.32	39.29	1.01	41.44	0.65	48.85	0.93	44.64	0.94	**49.29**	**1.10**
Fish	52.79	**1.19**	41.57	1.02	41.06	0.86	42.97	0.84	53.11	1.15	**57.39**	1.15
Reef1	76.64	1.05	69.78	1.00	80.74	1.04	72.31	0.79	**79.45**	**1.08**	74.62	1.03
Reef2	73.33	0.53	51.97	0.71	55.54	0.43	49.93	0.77	**58.63**	0.54	55.68	**1.01**
Reef3	**104.98**	0.90	63.90	0.98	83.24	0.73	71.07	0.88	89.93	0.88	79.24	**1.26**
Galdran1	70.08	0.68	69.12	1.06	71.38	0.75	69.71	0.51	**93.35**	1.15	84.71	**1.18**
Galdran9	43.73	1.10	27.91	0.98	30.64	0.86	41.81	1.16	36.13	1.14	**48.96**	**1.22**
Ancuti1	26.86	1.04	20.91	0.86	19.84	0.91	27.99	0.96	23.79	1.04	**52.65**	**1.25**
Ancuti2	13.21	0.62	17.32	0.64	17.63	0.48	24.93	0.60	16.66	0.60	**62.54**	**1.47**
Ancuti3	44.44	1.01	36.72	1.07	43.49	0.97	48.56	1.02	43.85	1.13	**64.14**	**1.25**
Average	55.93	0.84	43.85	0.93	48.50	0.77	49.81	0.85	53.95	0.96	**62.92**	**1.19**

FIGURA 51: ***Quantitative assessment of the results based on the VM and PCQI indices. Several associated images are shown in Fig. 52. From left to right we compare with the methods of Carlevaris-Bianco et al. [118], He et al. [67], Drews-Jr et al. [120], Galdran et al. [121] and Emberton et al. [62].***

number of underwater images captured in various environments. We compared our results with those produced by recent special-purpose underwater enhancment methods [62, 118, 120, 121] and also with the seminal dehazing techniques of He et al. [67].

Fig. 50 presents several images taken from the work of Tali and Schechner [117]. These images are artificially illuminated and, due to multiple scattering process [117], the light is non-uniformly spread over the entire scene. Even if Emberton et al. [62] estimates back-scattering influence locally, it fails to restore both the contrast and the color of the scene. As shown also in Fig. 22, outdoor dehazing techniques [67] exhibit similar limitations. On the other hand, our approach performs considerably better than the polarization approach of Treibitz and Schechner [117] that employs two images (taken with different states of the light-source polarizer).

Fig. 52 completes the comparative analysis by presenting the results obtained when restoring underwater scenes with relatively uniform illumination. In this particular case, existing underwater approaches [62, 118, 120, 121] generally lead to satisfying results. For this less challenging case, we performed a quantitative evaluation. We considered the same set of 10 underwater images used by Emberton et al. [62] to assess the various underwater techniques [4, 119–121].

As quantitative assessment metrics, we employed two contrast-based measurements: the hazy visibility metric (VM) [124] also used in [62], and the recent patch-based contrast quality index (PCQI) of Wang et al. [125]. Table 51 shows that our approach generally achieves the best results based on the VM and PCQI

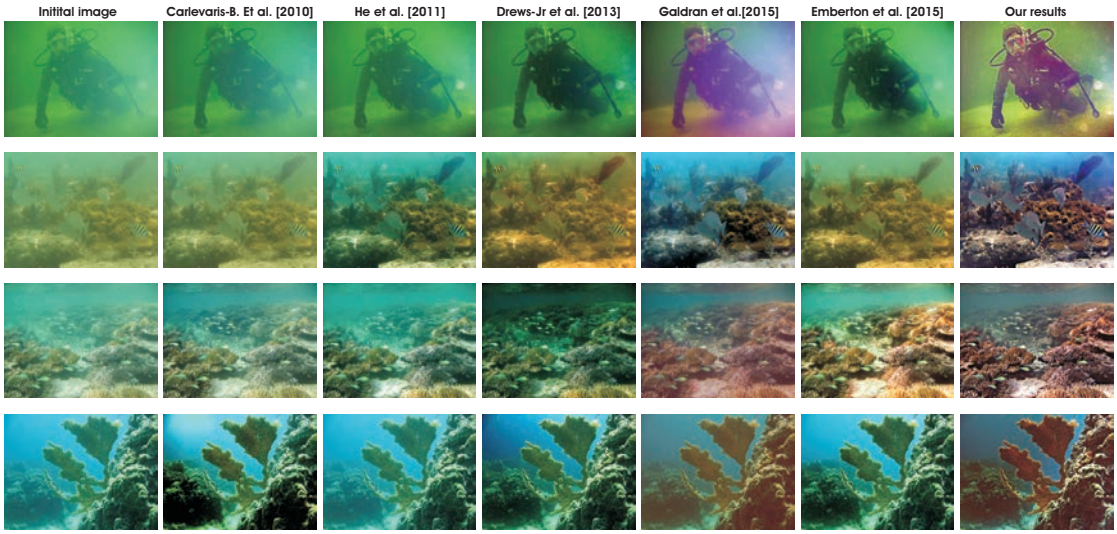

FIGURA 52: ***Comparative results on underwater scenes with uniform lightning***. From a set of ten underwater images used in [62] and evaluated in Table 51, the comparative results for images referred as **Ancuti2**, **Ancuti3**, **Galdran1** and **Reef1** are shown.

indexes. This conclusion is confirmed by careful inspection of the images presented in Fig. 52.

In summary, while existing underwater approaches are generally competitive on underwater scenes with reasonable and uniform illumination, our fusion technique demonstrates significant improvements when restoring visibility on more challenging artificially illuminated underwater scenes.

II.3 Laplacian-Guided Image Decolorization

The decolorization (color-to-grayscale) process converts a three-channel color image into a single channel image version. Many image processing and pattern recognition applications operate on a single image channel (grayscale) and therefore the loss of information produced by the decolorization operator has to be minimized. Moreover, different monochrome devices and black-and-white printing require perceptually plausible decolorised (grayscale) images.

Traditionally, the standard decolorization conversion is simply employed as the luminance channel of different color spaces (e.g. $CIEL^*a^*b^*$, $YCbCr$, HSL/HSV) [126]. However, this simple global mapping disregards important chromatic information and therefore in many cases the output grayscale images do not preserve the original appearance of the color while important transitions and details are often faded and misidentified (please observe the L^* channel shown in figure53). This limitation has been addressed by various techniques [127–134]. Roughly, the existing techniques can be divided in two main classes: local [68, 129, 135] and global [127, 128, 130, 132, 136, 137] mappings. While local mapping techniques preserve better the local features, global mapping techniques are able to map better over the entire image the same color to the same gray level (crucial in such transformations).

In this part we introduce a novel local grayscale mapping strategy that filters the features and transitions based on the Laplacian information. The work has been published recently by the candidate in [6]. Our grayscale transformation, designed in RGB color space, takes as individual inputs the three color channels (R, G, B). As can be observed in Fig. 53 although the edges are not visible on the luminance channel L^* (due to linear mapping strategy) they are relatively well defined on at least one of the RGB color channels. Since we aim to preserve most of the visible edges of the original color image, our algorithm is guided by two weight maps that transfer in the final result the most significant information of each derived input (RGB color channels). The first weight map, defined based on the Laplacian operator, filters the local contrast map assigning high values to local transitions of the image. The second weight map is designed to emphasize not only the most significant values, but also the regions characterized by high contrast in RGB color space. In order to minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion, using a Laplacian pyramid decomposition of

the inputs combined with the Gaussian pyramid of normalized weights. Multi-scale fusion is a well-known concept that has been demonstrated to be less prone to introduce artifacts for many enhancing applications [1, 4, 39, 41, 58, 68, 69].

The presented work is inspired by our previous fusion approach [68] and the local mapping method of the method of Smith et al. [129] that has been classified in the Cadik's study [138] as the most perceptually accurate. However, due to the unsharp mask-related strategy, Smith et al. [129] method is prone to introduce discontinuities along edges. In comparison, our multi-scale fusion algorithm shown to perform faster yielding less local artifacts (please refer to Fig.53 and Fig.57 but also to the supplementary material). On the other hand, unlike our former strategy [68] the proposed technique is a simplified one since we use only 3 inputs (the approach of [68] employs 4 inputs that include also an input that aims to conserve the contrast based on Helmholtz-Kohlrausch effect). Moreover, here we use only 2 (different than in [68]) weights compared with 3 weights used in [68]. As a result, the fusion strategy presented in this part is less complex and therefore more computationally effective, yielding results with less local distortions as demonstrated in the validation section.

Moreover, compared with the method of Ancuti et al. [68] that employs color information, the presented method uses instead local contrast guidance for estimating the weight maps. This is motivated by simultaneous contrast experienced by color perception. Simultaneous contrast is more intense when two color are complementary e.g. red-green, blue-yellow. For instance analyzing the yellow-blue pair in Fig. 53, it can be observed that on the blue channel there is a high contrast between the yellow and blue regions.

Our comprehensive experiments demonstrate that our technique is able to yield more consistent results compared to the state-of-the-art local mapping techniques and comparable with the more complex global approaches. Moreover, our method is straightforward to implement and computationally effective. It demonstrates consistency over varying palettes being able to maintain temporal coherence of videos.

As a second main contribution, we perform a comprehensive qualitative but also a quantitative evaluation of the state-of-the-art decolorization operators. We statistically interpret a simple blind contrast assessment that measures the saturated

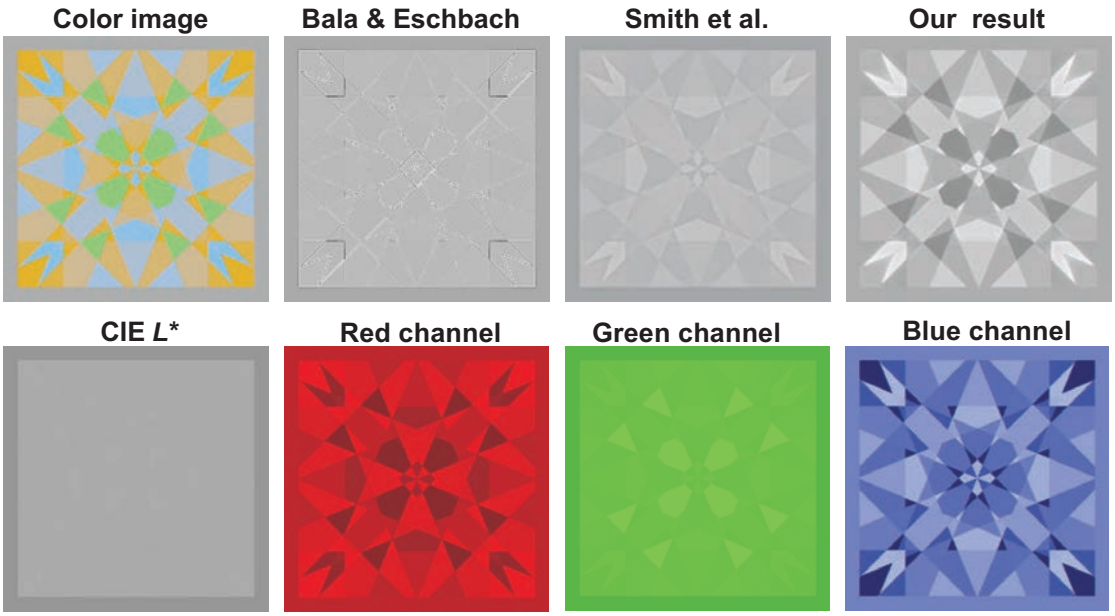

FIGURA 53: *On the top line, from left to right are shown the initial color image, the luminance channel L^* (standard conversion) in the $CIEL^*a^*b^*$ color space and our result. On the bottom line from left to right are presented the R,G,B color channels. Please observe the visibility level of edges and local contrast of the red and blue channels. Our technique does not act as a simple channel selector, since the final result contains both local information of the blue and red channels (notice the center of the image).*

pixels between the initial and the decolorised image versions. The results on natural and synthetic images demonstrate the utility and the robustness of the novel decolorization technique.

II.3.1 Our Decolorization Technique

In practice, the standard decolorization approach employs a global linear mapping such as:

$$L = w_R R + w_G G + w_B B \tag{46}$$

where the weights w_R, w_G, w_B are scalar parameters that represent the contribution ratio of each channel to the final grayscale result. For preserving the range consistency of the luminance result, the weights are normalized such as the sum of all weights (for each pixel (x,y) of the image) to sum one ($w_R(x,y) + w_G(x,y) + w_B(x,y) = 1$). Obviously, this linear blending does not account for image characteristics such as local contrast and differences in color-to-gray mapping, as shown in Fig. 8.

On the other hand, non-linear mapping techniques are built based on some local information represented by specific weight maps. These weights are computed

independently for each pixel. Unlike linear mapping, the weights (weight maps) are vectors. While the non-linear methods guided by appropriate weight maps are more effective to represent in the same time the local and global appearance, compared with linear-mapping, they are prone to introduce additional unpleasing local artifacts (e.g. please notice the results of Fig.8 and Fig.57). These visual degradations are introduced mainly close to the locations where the transitions (high frequencies) of estimated weights do not match the transitions of the original image.

To overcome these limitations, we introduce a non-linear mapping algorithm built on the fusion principle. Deriving several inputs and weight maps (vectors) from the original color image, the inputs are mixed into a single output, by preserving only the most significant features of the derived inputs. As already mentioned, our fusion technique defines three inputs derived from the original color image (the R,G and B color channels). As will be detailed below, our weight maps are designed in order to minimize both the local and global information (some) inevitable loss due to compression process.

Laplacian weight computes the amount of local variation of each derived input. Since Laplacian operator $L\left\{\mathcal{I}(x,y)\right\} = \nabla^2\mathcal{I}(x,y)$ is a simple but robust indicator for local contrast [41], we define the first weight map as:

$$\mathcal{W}_L^k(x,y) = \overline{L\left\{\mathcal{I}^k(x,y)\right\}} + \left|L\left\{\mathcal{I}^k(x,y)\right\}\right| \tag{47}$$

where k is the index of the input channels ($k = 1..3$ since we have three inputs: R, G, B color channels). The first term represents the arithmetic mean of the Laplacian computed for each input channel. It reflects the total amount of important transitions (edges), while the second term (the absolute value of Laplacian) indicates the location of the existing transitions. This weight map assigns high values to the important signal transitions, related in general to edges and texture. Additionally, the first term (arithmetic mean) ensures a consistent temporal coherence of our operator for videos (please refer to Fig. 56). Similar contrast measures built on Laplacian have been used previously but for different applications such as extended depth-of-field [36] and tone mapping [41].

While this measure can be derived from other more complex filters, we opted to prove our concept based on the Laplacian, a fundamental filter. However, employing only this weight, our fusion strategy is not able to render accurately

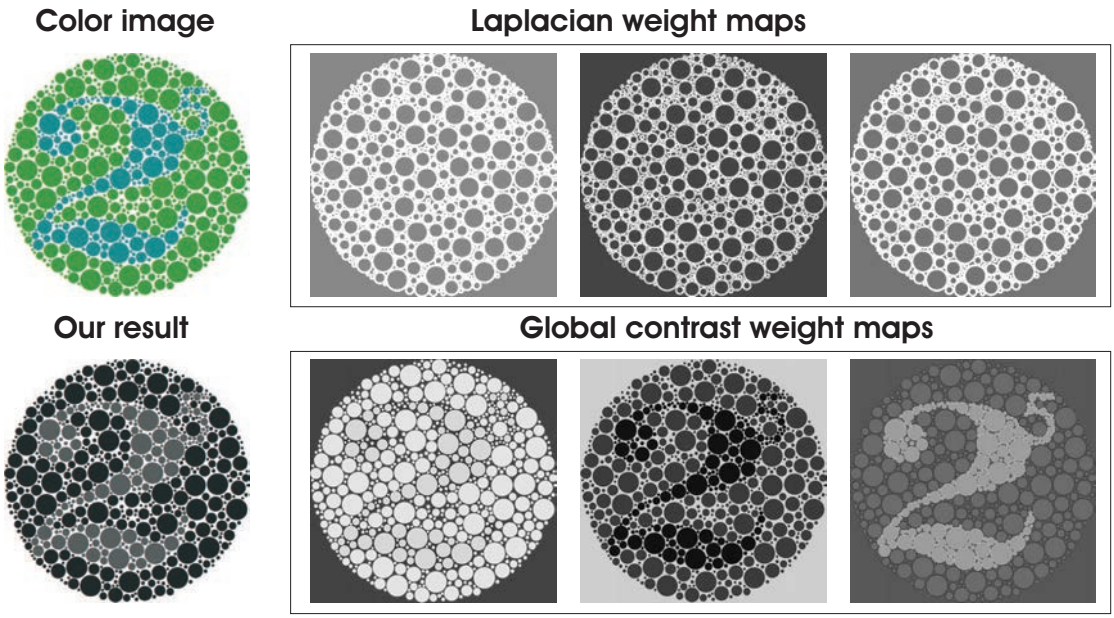

FIGURA 54: ***Weight maps.*** *It is noticeable that the Laplacian weight maps emphasize the local transitions but are not able to conserve the global information. On the other hand, the original global appearance is preserved and transferred in the final result by our second weight map (Global contrast weight).*

the original global appearance of the color input (please refer to Fig.54). To overcome this limitation we define a second measure described next.

Global contrast weight filters and propagates the important global values in each of the input channels. In general, analyzing an image one expects that high contrast regions in rapport with their neighborhoods to be visually highlighted (to appear brighter). Therefore, we searched for a simple but efficient measure to satisfy this perceptual feature of human visual system. Basically, this measure has to assign high values to regions that have higher contrast compared to the estimated average contrast of the scene. Mathematically, this global weight map is expressed by subtracting from each input its arithmetic mean of the Laplacian:

$$\mathcal{W}_G^k(x,y) = \left[\mathcal{I}^k(x,y) - \overline{L\left\{\mathcal{I}^k(x,y)\right\}}\right]^{\delta} \tag{48}$$

Computed on each of the three inputs (R,G,B color channels), this measure assigns small values for those values close to estimated arithmetic mean of Laplacian. The parameter δ has the default value $\delta = 2$ (increasing this value, intensifies the global discriminability in the output result, as well). This weight represents a good tradeoff solution since comparing every region with all the others (as employed by the approach of Gooch et al. [127]) shown to be computationally prohibitive.

Obviously, this weight map is not able to ensure distinct values for adjacent regions that are equally distant compared to the mean value $\overline{L\{\mathcal{I}^k(x,y)\}}$. In such cases our first defined weight map overcomes this issue preserving the original appearance due to the filtered edges.

The three inputs (R, G, B) and the weight maps (Laplacian and Global contrast weight maps) are fused by our strategy as described in the following. Generally, fusion techniques simply define how every pixel of the inputs are mixed guided by the the weight maps. Basically, we aim to preserve in the grayscale output the most relevant pixels of each input filtered by the defined weight maps. This is achieved by simply assigning high values to the locations where the weight maps have high values. Mathematically, the naive result is obtained by the following expression:

$$\mathcal{F}(x,y) = \sum_k \bar{\mathcal{W}}^k(x,y)\mathcal{I}^k(x,y) \tag{49}$$

where the index k counts the number of the inputs, while each pixel (x,y) of $\mathcal{F}$ is obtained by summing the corresponding locations of the inputs $\mathcal{I}^k$ that are balanced by the normalized weight maps $\bar{\mathcal{W}}^k$.

FIGURA 55: ***Naive vs. Multi-scale fusion.*** *As can be seen compared with the multi-scale strategy, the naive fusion implementation (equation 49) yields inconsistent results characterized by unpleasing halos artifacts close to the edges.*

This naive implementation (equation49) has shown to yield inconsistent results characterized by unpleasing halos artifacts close to the edges (see Fig. 55). To overcome this problem we employ the multi-scale pyramidal refinement strategy [40, 56]. As a result, for each of the three inputs, a Laplacian multi-scale pyramid is computed by applying Laplacian operator at different scales. At the same scales, for each normalized weight map $\bar{\mathcal{W}}$, a Gaussian pyramid is built. The blending between the Laplacian inputs and Gaussian normalized weights is performed at

each level independently. Mathematically, each pyramid scale level l of the fused result is expressed as:

$$\mathcal{F}^l(x,y) = \sum_k G^l\left\{\bar{\mathcal{W}}^k(x,y)\right\} L^l\left\{\mathcal{I}^k(x,y)\right\} \tag{50}$$

where l represents the number of the pyramid levels (the number of pyramid levels is image dependent being determined as logarithm of the image size) and $L^l\{\mathcal{I}\}$ is the Laplacian version of the input $\mathcal{I}$ at scale l. $G^l\left\{\bar{\mathcal{W}}\right\}$ represents the Gaussian version of the normalized weight map of the $\bar{\mathcal{W}}$ at the same scale level l. The final decolorised image is obtained by summing the fussed contribution of all levels $\mathcal{F}^l$ of the resulted pyramid.

This multi-scale fusion strategy is able to avoid effectively seams artifacts by blending the features of the inputs selected by the weight maps. By employing independently a fusion process at every scale level the potential artifacts due to the sharp transitions of the weight maps are significantly reduced.

II.3.2 Results and Discussions

Our novel decolorization operator has been tested extensively for a large dataset that includes images of various natural scenes but also challenging synthetic images. In Fig. 57 are shown several comparative results. As ca be observed, compared with the local mapping state-of-the-art techniques [68, 128, 129, 135, 137] our fusion-based decolorization is less prone to artifacts being able to yield comparable results with the global mapping techniques, as well. Moreover, compared with Rache et al. [128] and Lu et al. [137] we are able to preserve the white color in the decolorized output which is an essential requirement of a grayscale operator (see the *butterfly* image in Fig. 57).

Additionally, our operator has demonstrated to produce consistent results also for videos. Decolorization of videos is not a trivial task being demonstrated for the first time in the work of Smith et al. [129]. In contrast, one of the most perceptually accurate technique (classified in the study of Cadik's [138]), Grundland and Dodgson [130] is not able to preserve the temporal coherence. This is proved in Fig.56 and also in our supplementary video.

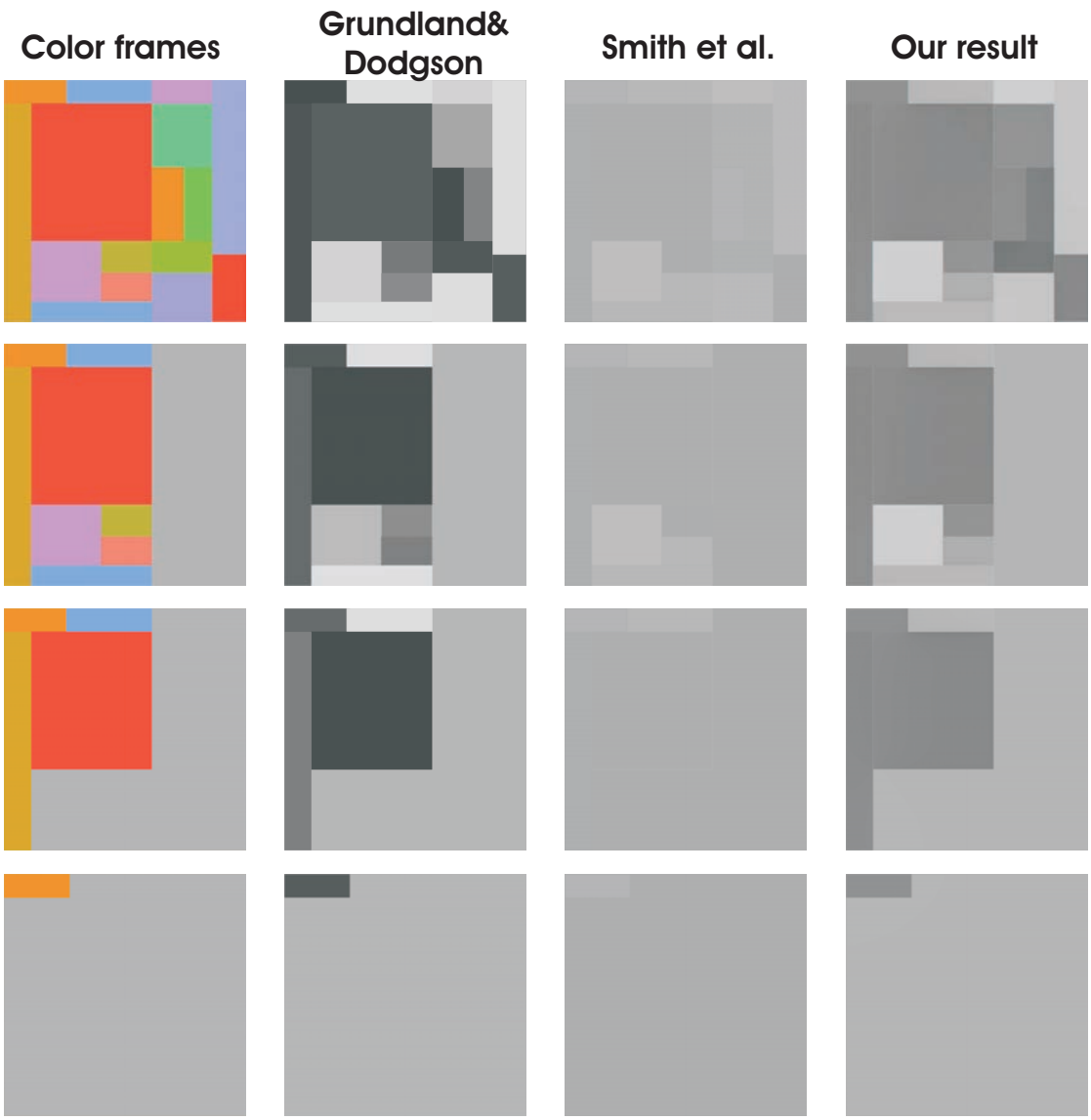

FIGURA 56: ***Temporal coherence.*** *Compared with Smith et al. [129] and our strategy, the global method of Grundland and Dodgson [130] is not able to preserve temporal coherence.* ***Please check the video of the supplementary material.***

In addition, our operator is simple and computationally inexpensive. We are able to process 10 fps (800x600 resolution) using a common portable PC with a 2.5 Ghz Intel Core i7 CPU and 8 GB memory (very close to the computation times reported by the optimized implementation of Lu et al [137]). We believe that this is an important advantage of our operator. In comparison, the approach of Ancuti et al. [68] processes the same image in approx 2 seconds, the approach of Smith et al. [129] takes more than 10 sec, the technique of Grundland and Dodgson [130] takes approx. 3.5 seconds while the method of Kim et al. [131] takes approx. 1 sec.

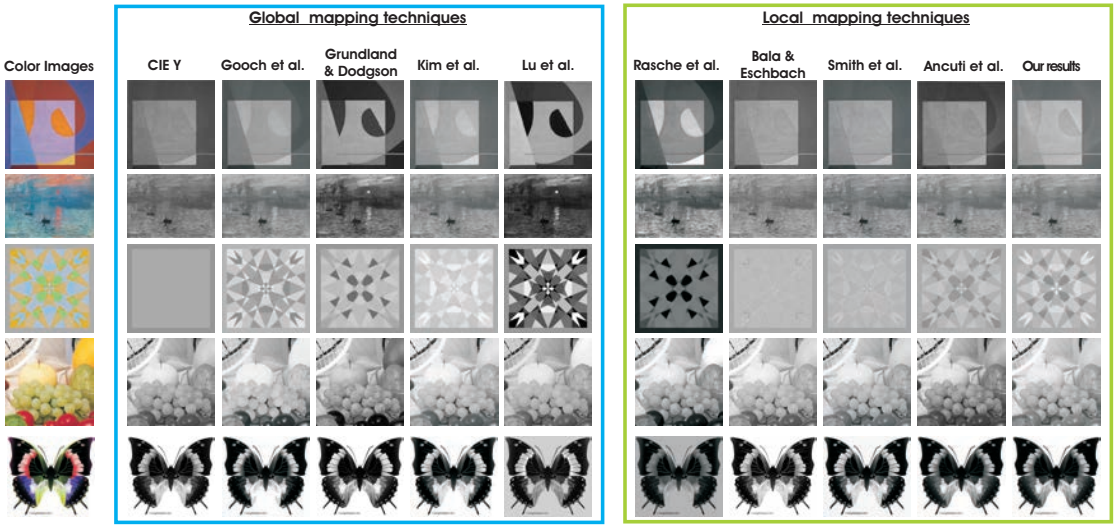

FIGURA 57: ***Comparison with the local and global mapping techniques.***

II.3.3 Validation

Besides the visual evaluation we propose a quantitative validation of the recent decolorization techniques. We define a visual descriptor to evaluate the level of local artifacts introduced by different methods. Basically, we are interested to measure how many pixels became black or white in the decolorized result compared with the original image. Inspired from the blind measure of Hautiere et al [24] we define the descriptor χ:

$$\chi = \frac{n_s}{M_I \times N_I} 100 \tag{51}$$

where n_s represents the number of pixels that become black or white while M_I and N_I denote respectively the width and the height of the image I. The descriptor χ measures the level of saturated pixels that as shown by [24] is a good estimator of the level of artifacts for local contrast manipulation techniques. In the work of Hautiere et al [24] a similar indicator was used to measures the local artifacts for two related applications: tone mapping and contrast enhancement. Basically, small values of χ are associated with a reduced level of artifacts.

The quantitative validation has been performed on the set of 24 images used in the perceptual evaluation of Cadik [138]. We evaluate the techniques of Gooch et al [127], Grundland and Dodgson [130], Kim et al. [131],Bala and Eschbach [135],Rasche et al. [128], Smith et al. [129], Ancuti et al. [68] and our strategy. The results shown in Fig.58 have been interpreted statistically using analysis of

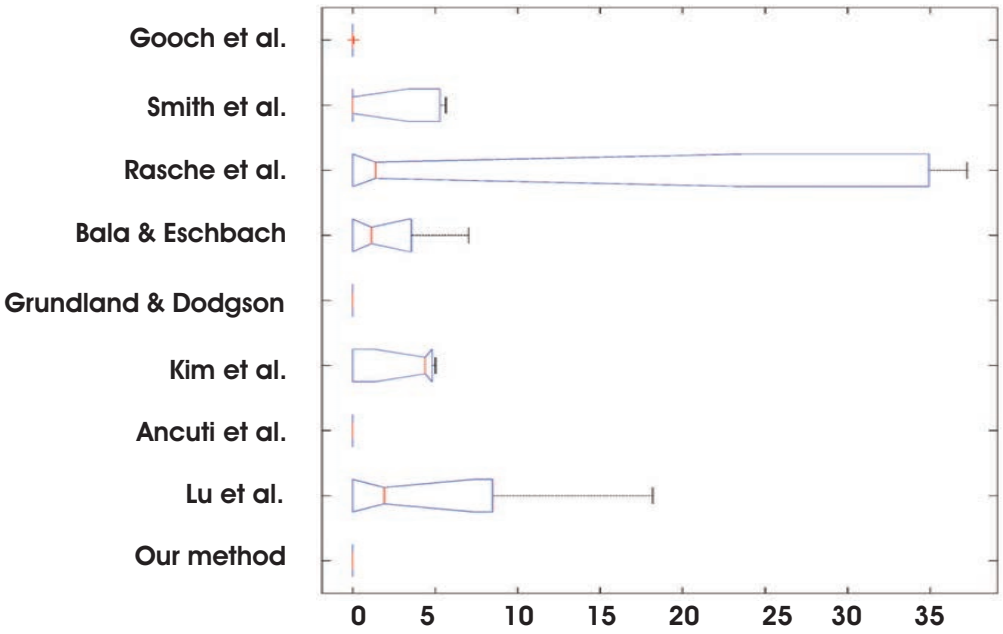

FIGURA 58: *Statistical (ANOVA) representation of the indicator χ.*

variance (ANOVA) [139]. In this representation the red line of the boxes correspond to the mean values of each method. A small mean value is desirable since it is associated with a reduced level of local distortions.

Interpreting the statistical results of descriptor χ shown in Fig.58, in general the class of local methods from which our techniques belongs to (Smith et al. [129], Rasche et al. [128] and Bala and Eschbach [135]) tend to introduce unpleasing degradations close to edges (please observe the level of the red line of each box). Indeed, these artifacts are visible by a close inspection. A similar conclusion corresponds also to the global method of Kim et al. [131]. Moreover, compared with the fusion-based approach of Ancuti et al. [68] the presented technique is a less complex and therefore it is able to perform faster. Moreover as demonstrated by our validation and the results shown in Fig.57 it yields results with less local distortions.

Our fusion strategy shown to perform similarly to the global mapping methods of Grundland and Dodgson [130] and Gooch et al. [127]. However, compared with these two global techniques, our technique has the advantage to be suitable for videos being also extremely fast compared with the optimization technique of Gooch et al. [127]approach (needs 25.7 seconds for a 200x200 image - GPU implementation).

II.4 Single-Scale Fusion: An Effective Approach to Merging Images

The advent of advanced image sensors has empowered effective and affordable applications such as digital photography, industrial vision, surveillance, medical applications, automotive, remote sensing, etc. However, in many cases the optical sensor is not able to accurately capture the scene content richness in a single shot. For example, the dynamic range of a real world scene is usually much higher than can be recorded with common digital imaging sensors, since the luminances of bright or highlighted regions can be 10,000 times greater than dark or shadowed regions. Therefore, such high dynamic range scenes captured by digital images are often degraded by under or over-exposed regions where details are completely lost. One solution to obtain a complete dynamic range depiction of scene content is to capture a sequence of LDR (low dynamic range) images captured with different exposure settings. The bracketed exposure sequence is then fused by preserving only well-exposed features from the different exposures. Similarly, night-time images are difficult to be processed due to poor illumination, making it difficult to capture a successful image even using the HDR (high dynamic range) method. However, by also capturing with a co-located infrared (IR) image sensor, it is possible to enrich the visual appearance of night-time by fusing complementary features from the optical and IR images.

Challenging problems like these require effective fusion strategies to blend information obtained from multiple-input imaging sources into visually agreeable images. Image fusion is a well-known concept that seeks to optimize information drawn from multiple images taken of the same sensor or different sensors. The aim of the fusion process is that the fused result yields a better depiction of the original scene, than any of the original source images.

Image fusion methods have been applied to a wide range of tasks including extended depth-of-field [36], texture synthesis [140], image editing [37], image compression [56], multi-sensor photography [141], context enhancement and surrealist video processing [142], image compositing [39], enhancing underexposed videos [90], multi-spectral remote sensing [143], medical imaging [144].

Many different strategies to fuse a set of images have been introduced in the literature [145]. The simplest methods, including averaging and principal component analysis (PCA) [146], straightforwardly fuse the input images' intensity values.

Multi-resolution analysis has also been extensively considered to match processing the human visual system. The discrete wavelet transform (DWT) was deployed by Li et al. [147] to accomplish multi-sensor image fusion. The DWT fusion method computes a composite multi-scale edge representation by selecting the most salient wavelet coefficients from among the inputs. To overcome the shift dependency of the DWT fusion approach, Rockinger [148] proposed using a shift invariant wavelet decomposition. Tessens et al. [149] used the directional curvelet transform (CVT) to separate high and low frequency image components while capturing image structures as a sparse set of coefficients. Another alternative is the contourlet transform [150], which combines the Laplacian pyramid with a directional filter bank. Zhang and Guo [151] deployed an undecimated, shift-invariant contourlet transform for image fusion. In another category of methods, Tang [152] introduced a discrete cosine transform (DCT)-based algorithm to enhance the contrast of the input images to be fused. Image fusion based on MRF models for remote sensing applications was described by Xu et al. [153]. Neural networks were employed by Fay et al. [154] to fuse night-vision images from multiple infra-red bands. A gradient-based method was introduced by Petrovic and Xydeas [155]. In their method, input images were represented at each resolution level using gradient map signals rather than absolute grey-level values. Liang et al. [156] formulated a tensor decomposition technique and used SVD to fuse multiple images. More recently, in the context of multi-exposure fusion, Shen et al. [157] introduced generalized random walks to achieve an optimal balance between two quality measures, i.e., local contrast and color consistency, while capturing scene details from different exposures. The problem of balancing color consistency and local contrast has been approached by estimating the probabilities of each output pixel belonging to one of the input images. Li et al. [158] proposed an effective framework built on guided filters [159] to improve the spatial consistency of fusion between the base and detail layers.

One of the most successful image fusion strategies is based on the Laplacian pyramid decomposition (see Fig. 59). Introduced by Burt and Adelson [56] in the context of extended depth of field, the Laplacian pyramid has been employed for applications ranging from image compression to image denoising. In the context of multi-scale fusion, the Laplacian pyramid decomposition has recently been demonstrated to be effective for several interesting tasks such as HDR imaging [41], image filtering [161, 162], single image dehazing [1, 69], image and video decolorization [68] and underwater image enhancement [4].

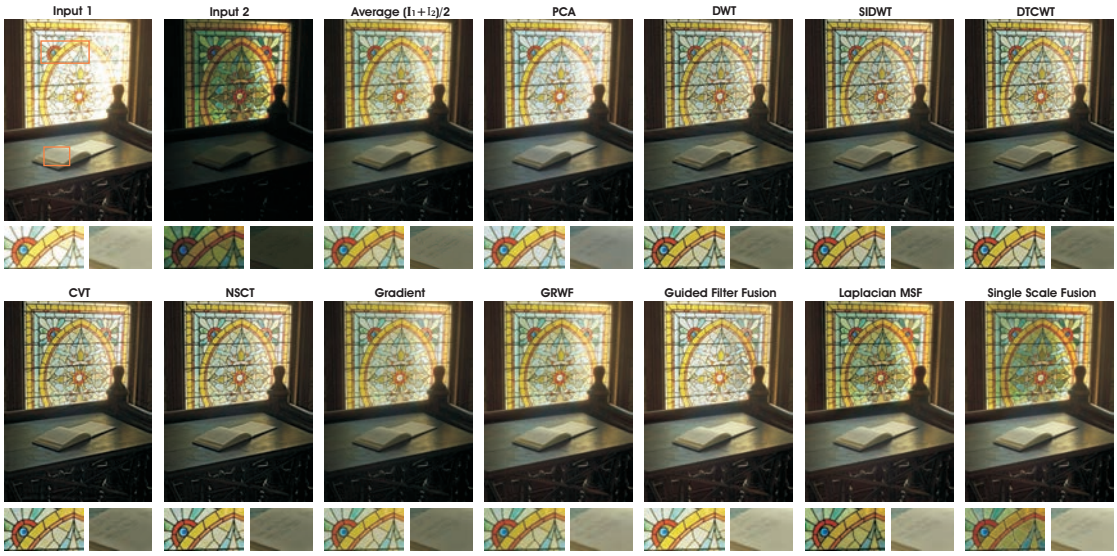

FIGURA 59: *Comparative results of different fusion techniques for merging multi-exposure images. Top row (from left to right), are shown the original two inputs, the result of averaging the inputs and the fusion results of [146–148, 160] . Bottom row (from left to right), are shown the results of the fusion approaches of [56, 149, 151, 155, 157, 158], and our single-scale fusion result . As can be seen, most of the fusion techniques yield results very similar to the simple average value of the two inputs. While most of the traditional fusion approaches yield results similar to the result obtained by simply averaging the inputs, the Laplacian multi-scale fusion [41, 56] is more robust and delivers results comparable to those of more recent fusion techniques such as GRWF fusion method of Shen et al. [157].*

Multi-scale fusion (MSF) based on the Laplacian pyramid became rapidly popular due to its effectiveness, but also to its intuitive method of deployment. The MSF process is guided by a set of measures (weights maps) that indicate the contribution of each pixel (of each input) to the final result. The weight maps capture the degree to which each input fits some desirable qualities (e.g. contrast, saliency) that are to be preserved in the fused result. Due to their inherent capacity to handle information at multiple scales, MSF based methods have been demonstrated to avoid the introduction of visual artifacts in image blending process.

However, despite of its popularity, MSF methods are generally computationally expensive and difficult to implement , especially in terms of data storage and transfer management [163, 164]. These limitations are particularly observable when processing large images since the number of levels of the multi-scale decomposition increases with the input image resolution. Decreasing the number of levels is not a solution, since it generally introduces unpleasing artifacts in the fused result (e.g. Fig. 67).

With these problems in mind, we have developed an easy-to-implement and computationally efficient alternative to the MSF strategy that fuses the multiple inputs in their native resolution, using weight maps defined on a single scale. Our fusion technique has been published recently in [7]. We first show how the MSF decomposition can be approximated using a single-scale decomposition in a way that eliminates redundant computations. Interestingly, the single-scale expression obtained from the MSF approximation also provides insightful cues regarding how the MSF process manipulates weights and image features to compute a visually pleasant outcome. It also helps explaining why MSF works, as compared to a simple weighted average of the inputs using low-pass weight maps.

We then demonstrate the generality and effectiveness of our proposed single scale fusion (SSF) in a variety of well-known fusion applications such as HDR imaging, image compositing, extended depth of field, medical imaging and blending IR with visible images. We also supply a quantitative evaluation that demonstrates that our single-scale fusion (SSF) strategy is able to yield results that are competitive with traditional multi-scale fusion (MSF) methods.

In summary, our work provides, as original contributions:
- the first single scale strategy for fusing multiple images that yields results that are highly competitive with classical multi-scale approaches;
- a mathematical derivation that identifies those components of the conventional MSF that are most critical to the blended image quality, which helps explain why MSF works;
- an extensive demonstration of the effectiveness of the single-scale fusion concept over a wide palette of applications;

II.4.1 Image Fusion: Background and Notations

Generally, image fusion can be defined as a process of effectively blending several input images (e.g. [41, 56]) or versions of the same original image (e.g. [1, 4]) into a single output image that retains the most naturalistic, high-quality elements from among all the source inputs. It is desirable that the fused results be free of any unpleasant artifacts not present in the scene. In order to generate a desired output, the fusion process is guided by several quality measures or weight maps. These

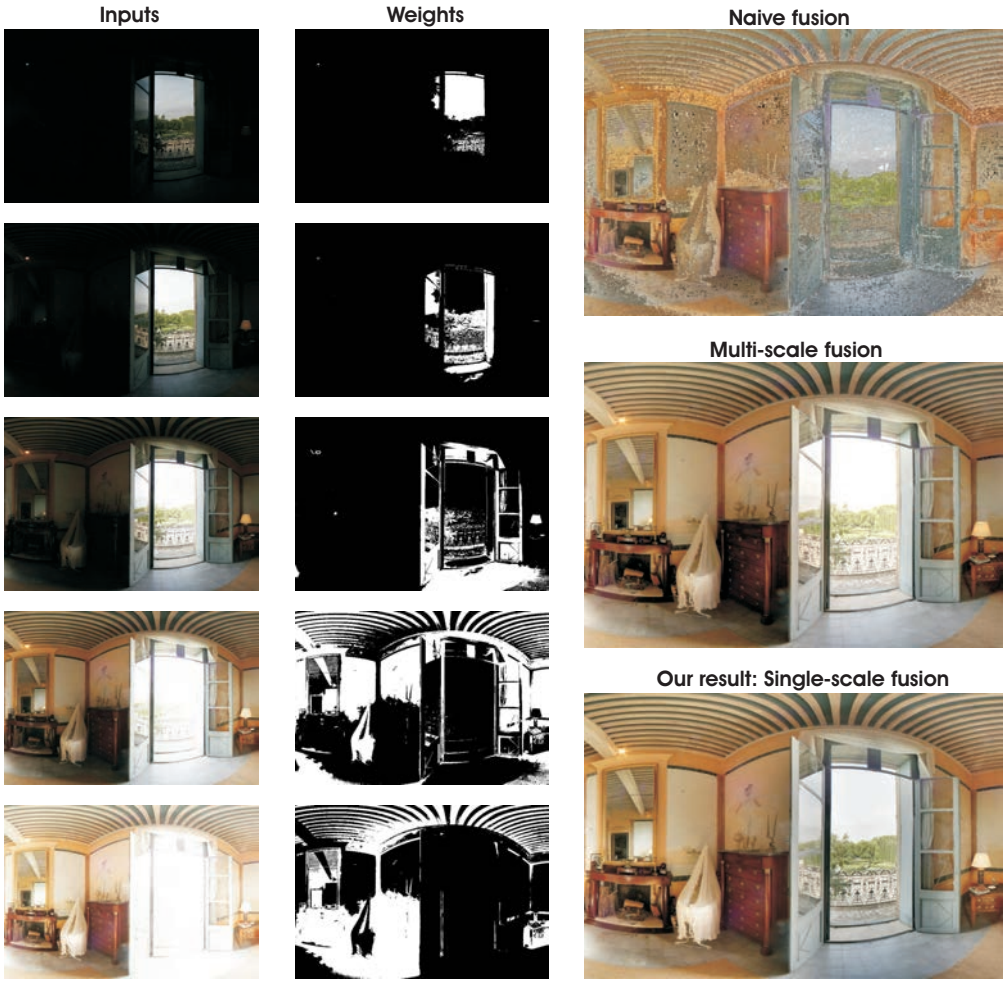

FIGURA 60: *Naive, multi-scale fusion and our single-scale result. Both involve, a similar degree of complexity, while our single-scale fusion method is able to deliver results competitive with the multi-scale approach.*

quality measures are generally defined dependent on the application, and aim to retain only those input features that transfer seamlessly to a visually satisfactory output result.

Before developing our single-scale fusion solution, we review the basic steps that define the classical image fusion process. We begin by briefly discussing the naive fusion solution, then we elaborate the multi-scale image fusion (MSF) approach based on the Laplacian decomposition.

Image fusion typically relies on set of weight maps that are used to transfer the most relevant features to the output. In its simplest form, the inputs $\mathcal{I}_k$ are directly weighted by some specific measures (weight maps) $\bar{\mathcal{W}}_k$, that indicate the amount that each image's pixels contribute to the final result. This approach, called naive image fusion (NF), is quite straightforward and computationally efficient. The

naive fusion result $\mathcal{R}_{NF}$ can be expressed as:

$$\mathcal{R}_{NF}(x) = \sum_{k}^{K} \bar{\mathcal{W}}_k(x)\mathcal{I}_k(x) \tag{52}$$

where K is the number of inputs. The weights $\bar{\mathcal{W}}_k$ are normalized to ensure that the intensity range of the result is similar to the dynamic range of the inputs: $\sum_k \bar{\mathcal{W}}_k(x) = 1$, for each coordinate x.

The naive fusion implementation involves a minimum number of operations, and has the additional advantage of preserving most of the available high frequencies in the final result. Unfortunately, the output of the naive fusion strategy contains distracting halos artifacts (see Fig. 60), especially in the regions containing strong transitions in the weight maps that have no correspondence with the input content. As pointed out in [41] and further discussed in next section (and depicted in Fig. 62), simple low-pass filtering of the weight maps is insufficient to remove those artifacts.

To overcome the limitations of the naive approach, the blending process can be performed in a multi-scale fashion. In order to explain our simplification we begin with the multi-scale image decomposition based on Laplacian pyramid originally described in Burt and Adelson [56]. The pyramid representation decomposes an image into a sum of bandpass images. In practice, each level of the pyramid does filter the input image using a low-pass Gaussian kernel G, and decimates the filtered image by a factor of 2 in both directions. It then subtracts from the input an up-sampled version of the low-pass image (thereby approximating a Laplacian), and uses the decimated low-pass image as the input for the subsequent level of the pyramid. Formally, using G_l to denote a sequence of l low-pass filtering and decimation, followed by l up-sampling operations, we define the N levels of the pyramid as follows:

$$\begin{aligned}
\mathcal{I}(x) &= \mathcal{I}(x) - G_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\} \triangleq L_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\} \\
&= L_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\} - G_2\{\mathcal{I}(x)\} + G_2\{\mathcal{I}(x)\} \\
&= L_1\{\mathcal{I}(x)\} + L_2\{\mathcal{I}(x)\} + G_2\{\mathcal{I}(x)\} \\
&= ... \\
&= \sum_{l=1}^{N} L_l\{\mathcal{I}(x)\} + G_N\{\mathcal{I}(x)\}
\end{aligned} \tag{53}$$

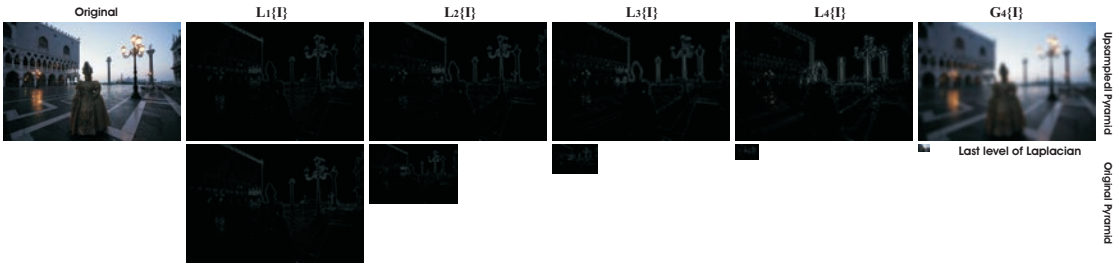

FIGURA 61: *The final level of a Laplacian pyramid includes a Gaussian blurred version of the original image. The top row depicts the upsampled versions of the five downsampled Laplacian and Gaussian signals shown in the bottom row. Note also that for better visualization the absolute value of each Laplacian image pixel is presented. This is to render the small Laplacian intensities in black and their large values in white, whatever their sign.*

As a result, the last component of the decomposition in (53), is a Gaussian blurred version of the input image with a large kernel (see Fig. 63). This is quite different from the other levels, which contain middle-to high frequencies. L_l and G_l represent the l^{th} level of the Laplacian and Gaussian pyramid, respectively. In the rest of this part all those images have been up-sampled to the original image dimension.

In the traditional multi-scale fusion (MSF) strategy [41], each source input $\mathcal{I}_k$, is decomposed into a Laplacian pyramid [56] 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 l :

$$\mathcal{R}_l(x) = \sum_{k=1}^{K} G_{l-1}\left\{\bar{\mathcal{W}}_k(x)\right\} L_l\left\{\mathcal{I}_k(x)\right\} \tag{54}$$

where $0 < l \leq N$ denotes the pyramid levels and k refers to the number of input images. The last component in (53) induces a last contribution $\mathcal{R}_{N+1} = \sum_k G_N\left\{\mathcal{W}_k\right\} G_N\left\{\mathcal{I}_k\right\}$. For a single level decomposition $N = 0$, $G_0\left\{\bar{\mathcal{W}}_k\right\}$ equals to $\bar{\mathcal{W}}_k$ and MSF reduces to naive fusion defined by equation (52). In practice, the number of levels N depends on the image size, and has a direct impact on the visual quality of the blended image (see Fig. 62).

This blending step is performed successively at each pyramid layer, in a bottom-up manner. The final multi-scale fused result $\mathcal{R}_{MSF}$ is obtained by simply summing

up the contribution from each level:

$$\mathcal{R}_{MSF}(x) = \sum_{l}^{N+1} \mathcal{R}_l(x) \tag{55}$$

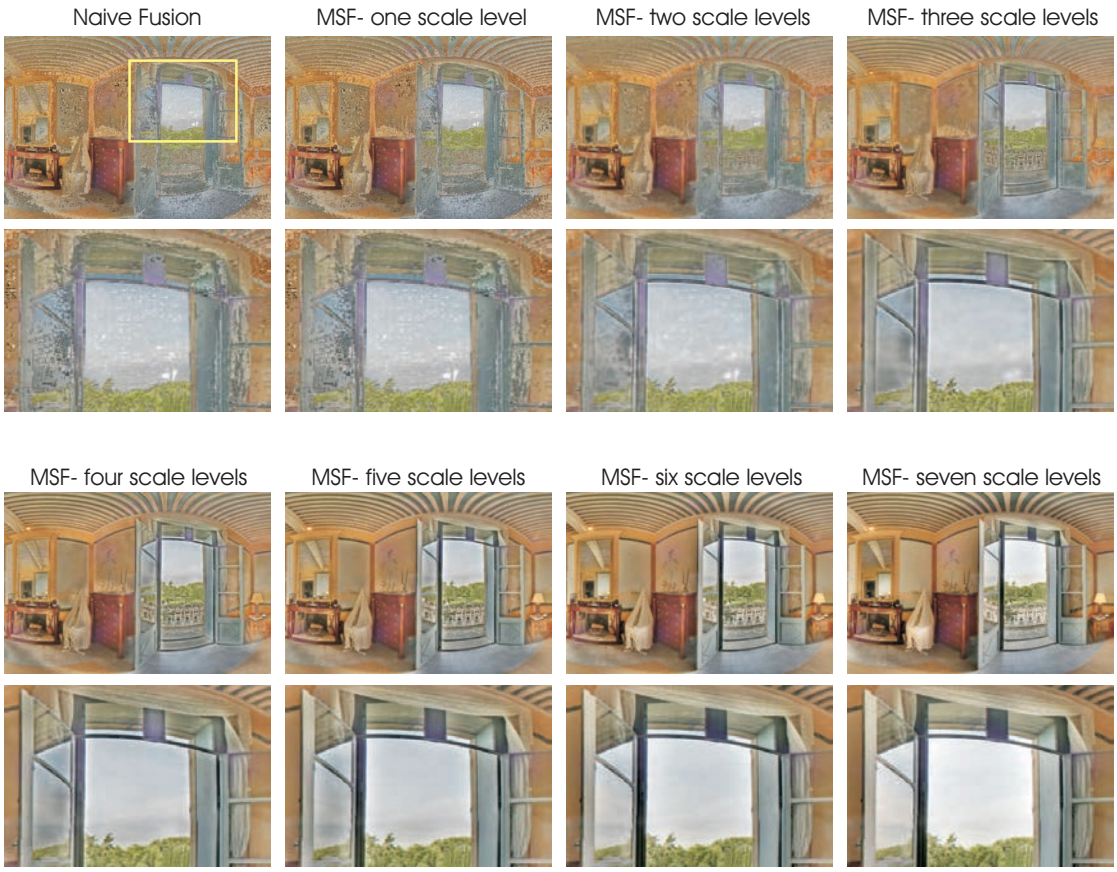

FIGURA 62: *Illustration of the influence of the number of levels in the multi-scale fusion approach. As can be observed, the number of levels affects the degree to which the higher image frequencies are revealed. Reducing the number of levels causes high frequency artifacts similarly to the naive fusion approach.*

II.4.2 Single Scale Fusion

This section derives our proposed single scale fusion strategy as an approximation of the conventional multiscale fusion approach. All along the section, to illustrate and justify our approximations, we present image samples corresponding to HDR imaging, using the well-known exposure fusion technique [41]. However, as will be illustrated in the next section, our approach is general, being suited to other scenarios and applications that are built on the multi-scale fusion process.

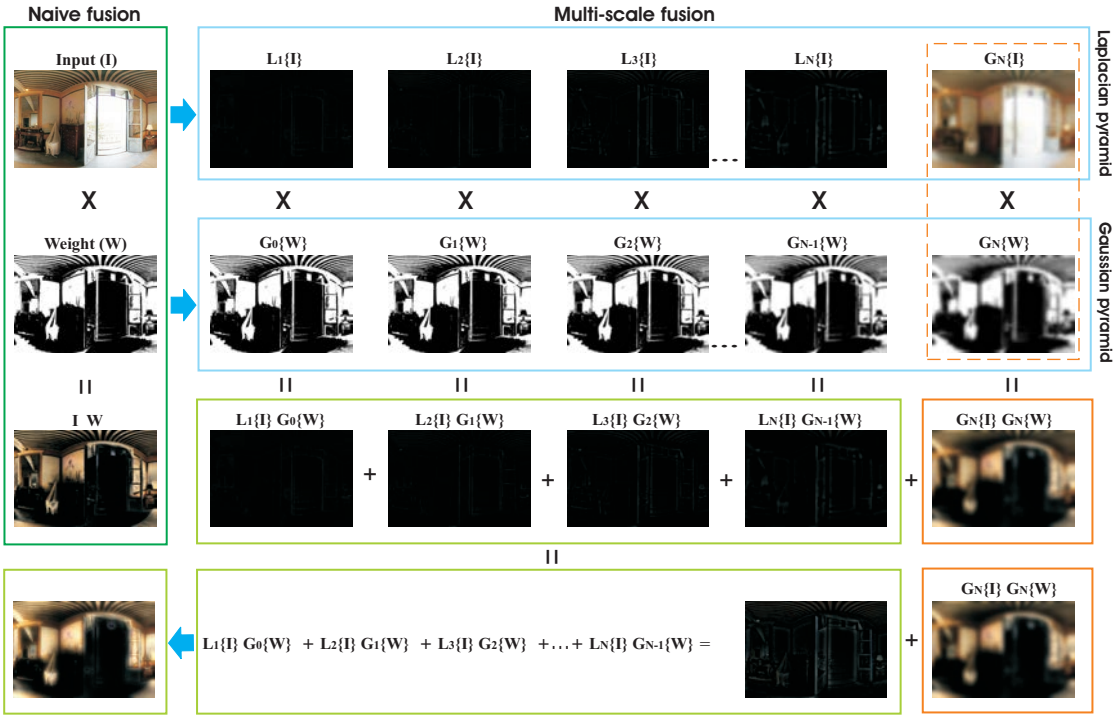

FIGURA 63: *Fusion pipeline. The Laplacian and Gaussian versions have been upsampled to the original size of the image.*

As discussed in previous section, the MSF builds on the Laplacian pyramid, and the contribution associated with the k^{th} input image $\mathcal{I}_k$ may be expressed (for simplicity, omitting index k and coordinate x):

$$\mathcal{R} = \sum_{l=1}^{N} G_{l-1}\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\} + G_N\left\{\bar{\mathcal{W}}\right\} G_N\left\{\mathcal{I}\right\} \tag{56}$$

To derive a single scale approximation of (56), we first observe that the empirical distribution of Laplacian of an image is heavily concentrated near zero, except near edges (black pixels are associated values near zero in Fig. 63). Hence, the lower levels of the pyramid only impact those regions that are characterized by significant gradient values. As a consequence, sharp transitions in the weight maps have little impact on the fusion process, unless they are aligned with similar events in the input. Based on this observation we could consider replacing $G_{l-1}\left\{\mathcal{W}\right\}$ by $G_N\left\{\mathcal{W}\right\}$ in (56). Then (56) becomes:

$$\mathcal{R} = \sum_{l=1}^{N} G_N\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\} + G_N\left\{\bar{\mathcal{W}}\right\} G_N\left\{\mathcal{I}\right\} \tag{57}$$

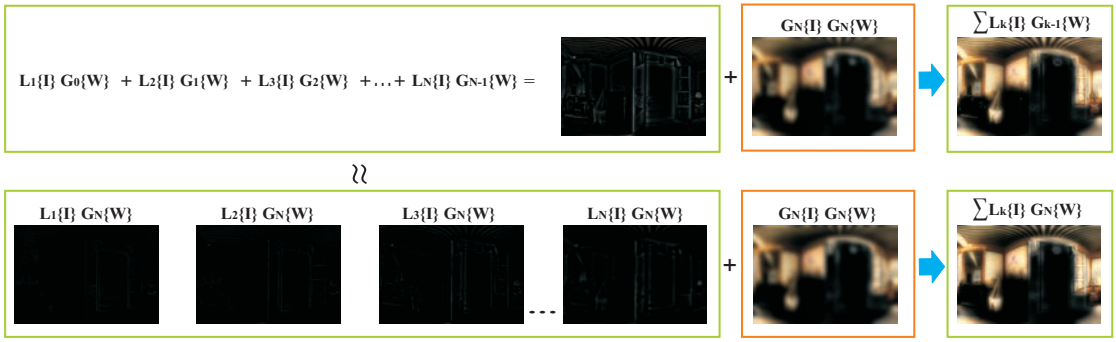

FIGURA 64: *As a first order MSF approximation, we had envisioned approximating the expression* $L_1\{\mathcal{I}\}G_0\{\mathcal{W}\}+L_2\{\mathcal{I}\}G_1\{\mathcal{W}\}+...+L_N\{\mathcal{I}\}G_{N-1}\{\mathcal{W}\}$ *by* $L_1\{\mathcal{I}\}G_N\{\mathcal{W}\}+L_2\{\mathcal{I}\}G_N\{\mathcal{W}\}+...+L_N\{\mathcal{I}\}G_N\{\mathcal{W}\}$*, thereby turning the MSF into a SSF. However, as illustrated in Fig.65, this approximation is not satisfying, thereby motivating our refined approximation derived from (58) to (66).*

In this equation, $G_N\{\bar{\mathcal{W}}\}$ can be put in evidence and the sum of $G_N\{\bar{\mathcal{I}}\}$ with the N Laplacians equals the image $\mathcal{I}$. Hence, this approximation reduces the fusion process to a single scale process, that is equivalent to the naive fusion strategy, but with Gaussian-filtered weights. Figures 65 and 67 reveal that, even if some image details are lost, the resulting outcome is free of any dramatic and visually disturbing artifacts.

This is an interesting finding, since until now it was commonly believed that smoothing the weight maps was inducing severe artifacts in the fused output (see, for example, the explanation and Figure 4 in *Mertens et al.* [41]). As may be seen in Fig. 65 , this observation from [41] is only partly valid since using a Gaussian filter with sufficiently large kernel size results in relatively artifact-free outcomes.

The reasonably good visual quality resulting from the simplification adopted in (57) also indirectly explains why MSF performs so well compared to the naive fusion strategy: abrupt transitions in the weight maps, which often introduce unpleasing artifacts into results produced by the naive method, tend to be canceled in the multi-scale fusion output, since discontinuities in the weight map tend to co-locate with abrupt changes in the inputs. Thus the MSF method also benefits by the contrast masking phenomenon [165], inherent in visual perception, which reduces the visibility of the artifacts in high contrast regions, especially when the artifact has similar orientation and location as the masking signal.

The importance of reducing or removing high-frequencies in the weight maps in regions that correspond to smooth inputs signals is confirmed when considering the number of levels involved in the MSF. In Fig. 62, we indeed observe that

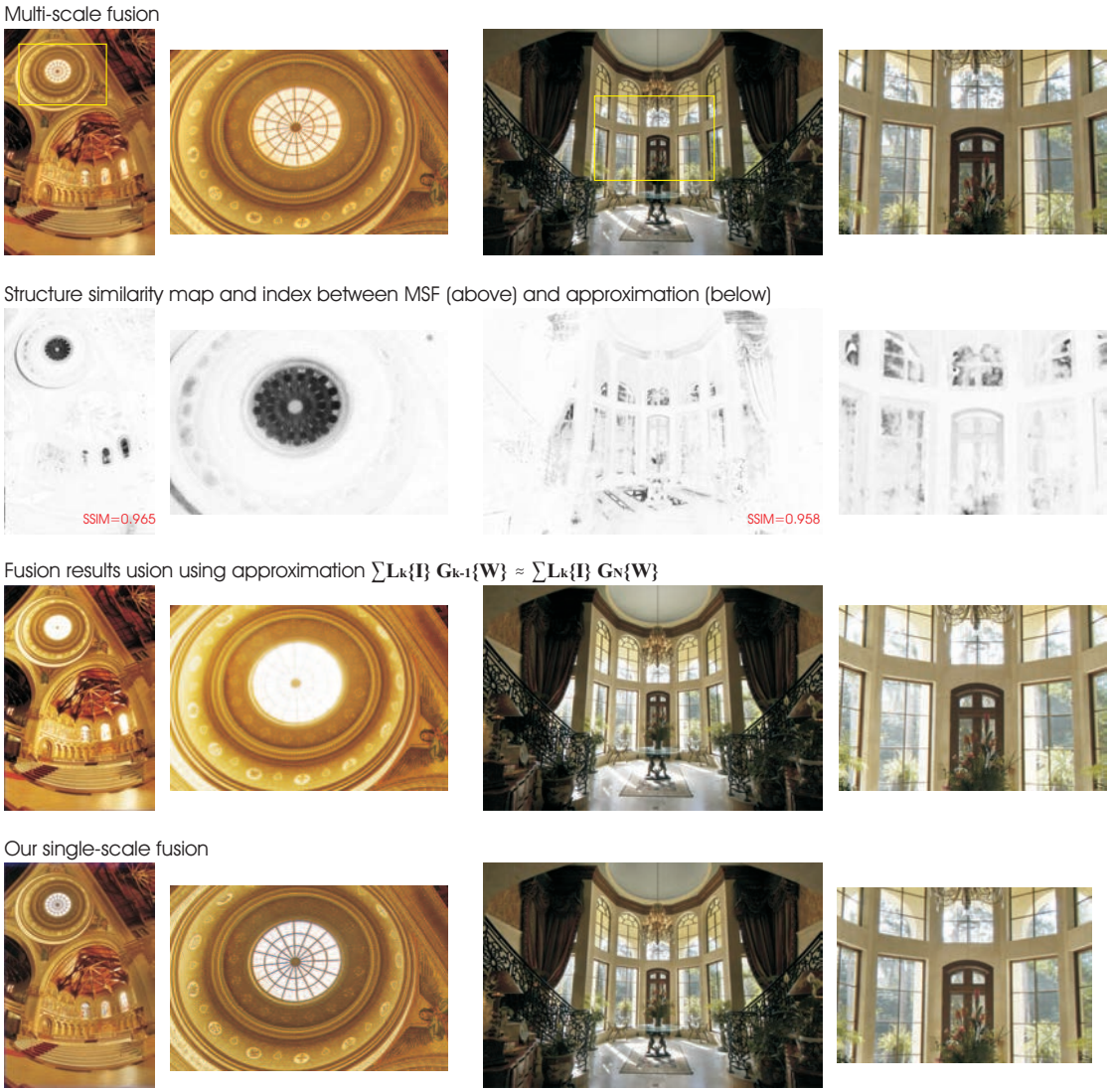

FIGURA 65: *Limits of straightforward single scale approximation derived from* $L_1\{\mathcal{I}\}G_0\{\mathcal{W}\} + L_2\{\mathcal{I}\}G_1\{\mathcal{W}\} + ... + L_N\{\mathcal{I}\}G_{N-1}\{\mathcal{W}\} \approx L_1\{\mathcal{I}\}G_N\{\mathcal{W}\} + L_2\{\mathcal{I}\}G_N\{\mathcal{W}\} + ... + L_{N-1}\{\mathcal{I}\}G_N\{\mathcal{W}\}$. *The top row shows the multi-scale fusion (MSF) results while the results obtained by the above approximation are shown in the third row. The second row shows structure similarity (SSIM) maps and index values computed between the MSF and the fusion results yielded by the mentioned approximation. The approximated MSF results are artifacts-free but important details are missing. In contrast, our SSF results (bottom row) preserve the details, as the MSF technique.*

decreasing the number of levels makes high frequencies in the weight maps much more disturbing, with unpleasant artifacts, similar to the naive strategy. This reveals that, to obtain a visually pleasant result, the multi-scale fusion strategy requires a sufficient number of pyramid levels which is computationally expensive and memory demanding on large images.

From the above discussion and observations, it should be clear that reducing high-frequencies in the weight maps is an important step towards obtaining visually

pleasing blended output images. However, even in the absence of any disturbing artifacts, the images resulting from (57) lack of details as compared to the MSF results (see Fig. 67 and Fig. 65). Therefore, we propose a second order approximation of (56), which aims to preserve details in the inputs, while remaining single scale. Formally, given that by definition of the Laplacian pyramid $G_{l-1}\left\{\bar{\mathcal{W}}\right\} = G_N\left\{\bar{\mathcal{W}}\right\} + \sum_{p=l}^{N} L_p\left\{\bar{\mathcal{W}}\right\}$, then (56) becomes:

$$\mathcal{R} = \sum_{l=1}^{N}\Big[\sum_{p=l}^{N} L_p\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\}\Big] + G_N\left\{\bar{\mathcal{W}}\right\}\sum_{l=1}^{N} L_l\left\{\mathcal{I}\right\} + G_N\left\{\bar{\mathcal{W}}\right\} G_N\left\{\mathcal{I}\right\} \quad (58)$$

By grouping the two last terms, the previous expression becomes:

$$\mathcal{R} = \sum_{l=1}^{N}\Big[\sum_{p=l}^{N} L_p\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\}\Big] + G_N\left\{\bar{\mathcal{W}}\right\}\mathcal{I} \quad (59)$$

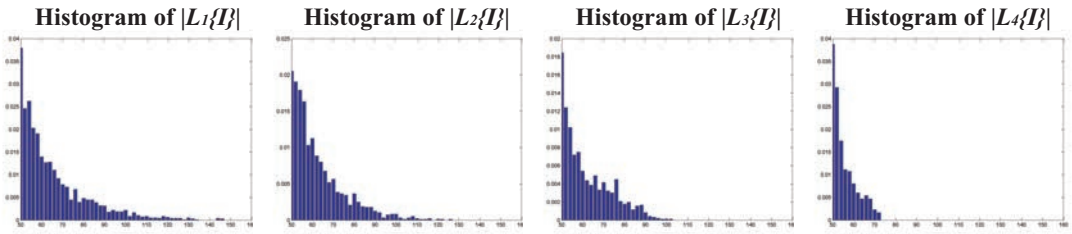

FIGURA 66: Histograms of $L_1\left\{\mathcal{I}\right\}$, $L_2\left\{\mathcal{I}\right\}$, $L_3\left\{\mathcal{I}\right\}$ and $L_4\left\{\mathcal{I}\right\}$. As expected, the histograms reveal that the largest values of $L_1\left\{\mathcal{I}\right\}$ are bigger than the largest values of $L_k\left\{\mathcal{I}\right\}$, with $k > 1$. This supports the approximation made to go from (59) to (60).

We can assume that the first term of the sum (l=1) dominates the others since the largest values, in $L_1\left\{\mathcal{I}\right\}$ tend to be much larger than the largest values in $L_k\left\{\mathcal{I}\right\}$, with $k > 1$ (see their histograms in Fig. 66). Here, we focus only on the largest values of $L_k\left\{\mathcal{I}\right\}$ because they are the only ones that matter when the products $L_k\left\{\mathcal{I}\right\} L_p\left\{\bar{\mathcal{W}}\right\}$ are added to $G_N\left\{\bar{\mathcal{W}}\right\}\mathcal{I}$.

We adopt a similar approximation for the Laplacian of the weight maps, (e.g $\sum_{p=1}^{N-1} L_p\left\{\bar{\mathcal{W}}\right\} \approx L_1\left\{\bar{\mathcal{W}}\right\}$). These approximations lead to the following expression:

$$\mathcal{R} \approx G_N\left\{\bar{\mathcal{W}}\right\}\mathcal{I} + L_1\left\{\bar{\mathcal{W}}\right\} L_1\left\{\mathcal{I}\right\} \quad (60)$$

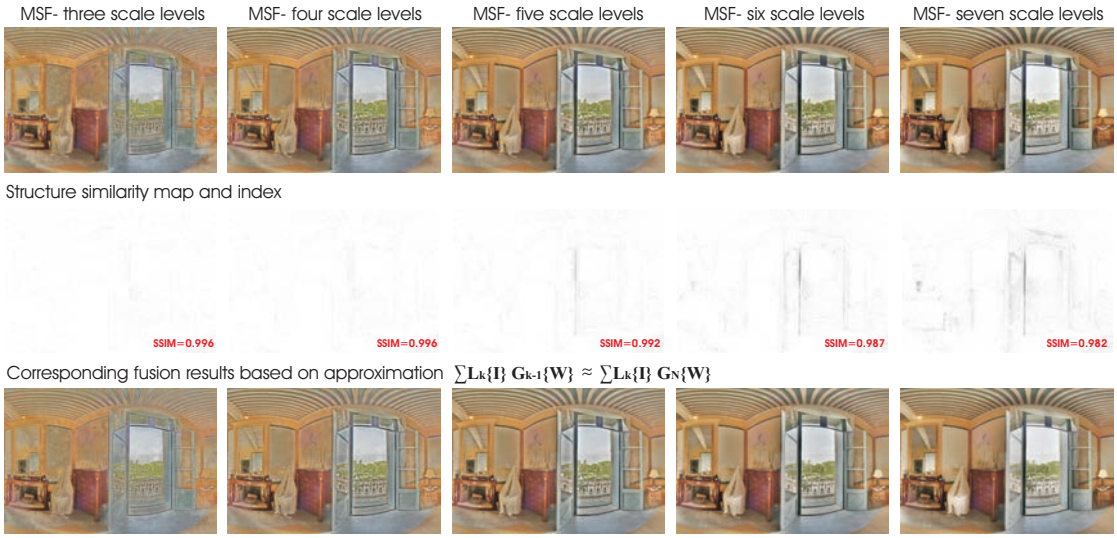

FIGURA 67: *The impact of the number of scales when approximating in (56)* $L_1\{\mathcal{I}\}G_0\{\mathcal{W}\}+L_2\{\mathcal{I}\}G_1\{\mathcal{W}\}+...+L_N\{\mathcal{I}\}G_{N-1}\{\mathcal{W}\}$ *by* $L_1\{\mathcal{I}\}G_N\{\mathcal{W}\}+L_2\{\mathcal{I}\}G_N\{\mathcal{W}\}+...+L_N\{\mathcal{I}\}G_N\{\mathcal{W}\}$. *We observe in this figure that (i) it is important to consider a sufficient number of scales in MSF to achieve artifact-free reconstruction, and (ii) the difference between MSF and the straightforward single scale approximation increases with the number of scale. We conclude that we should derive a more accurate SSF than the one simply replacing all Gaussian weights by* $G_N\{\mathcal{W}\}$.

By observing that $L_1\left\{\bar{\mathcal{W}}\right\}$ has a reasonably similar shape with $L_1\{\mathcal{I}\}$, in locations where $L_1\{\mathcal{I}\}$ is large (positive or negative), we obtain a preliminary version of our SSF simplification (see also Fig.68):

$$\mathcal{R}\approx G_N\left\{\bar{\mathcal{W}}\right\}\mathcal{I}+\beta L_1\{\mathcal{I}\}L_1\{\mathcal{I}\} \tag{61}$$

In this expression, $L_1\{\mathcal{I}\}$ is only significant at pixels that are close to an edge. Hence, we investigate how to approximate $L_1\{\mathcal{I}\}$ at a location x that lies close to an edge inflexion point x_0. For this purpose, we assume that the edge profile approximates a logistic function along the gradient orientation, in a small neighborhood around its inflexion point. The relevance of approximation is confirmed through extensive simulations in previous section, thereby experimentally validating our approach. Using a first order approximation, and the fact that at the inflexion point $\mathcal{I}(x_0)\approx G_1\{\mathcal{I}(x_0)\}$, we have:

$$L_1\{\mathcal{I}(x)\}=\mathcal{I}(x)-G_1\{\mathcal{I}(x)\}\approx(x-x_0)\left[\nabla\mathcal{I}(x_0)-\nabla G_1\{\mathcal{I}(x_0)\}\right] \tag{62}$$

and

$$\mathcal{I}(x)\approx\mathcal{I}(x_0)+(x-x_0)\nabla\mathcal{I}(x_0) \tag{63}$$

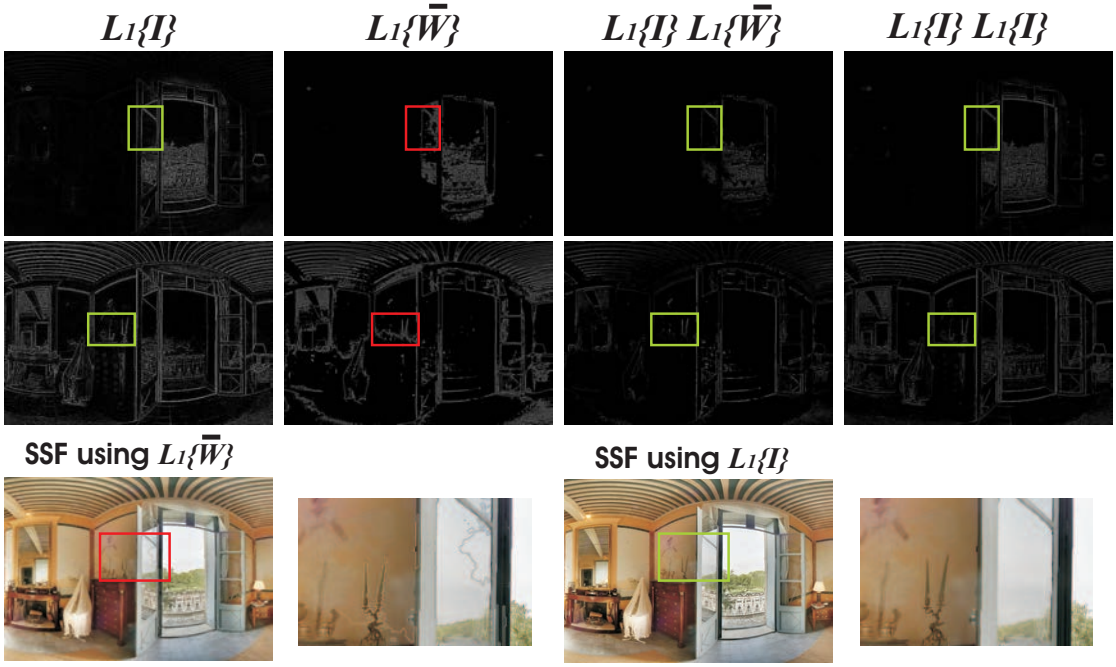

FIGURA 68: Top two rows show $L_1\{\mathcal{I}\}$, $L_1\{\bar{\mathcal{W}}\}$, $L_1\{\mathcal{I}\}.L_1\{\bar{\mathcal{W}}\}$ and $L_1\{\mathcal{I}\}.L_1\{\mathcal{I}\}$ of two inputs shown in Fig.60. *$L_1\{\bar{\mathcal{W}}\}$ has a shape that is similar to that of $L_1\{\mathcal{I}\}$, in locations where $L_1\{\mathcal{I}\}$ is large, and therefore we can simplify (60) to (61). Spurious edges could be transferred to the output if $L_1\{\mathcal{I}\}$ was approximated with $L_1\{\bar{\mathcal{W}}\}$ in (60) (see red boxes). In the bottom row are shown the results based on our SSF expression if using $L_1\{\bar{\mathcal{W}}\}$ and $L_1\{\mathcal{I}\}$, respectively.*

By merging (62) and (63), we have:

$$L_1\{\mathcal{I}(x)\} \approx [\mathcal{I}(x) - \mathcal{I}(x_0)] \frac{[\nabla\mathcal{I}(x_0) - \nabla G_1\{\mathcal{I}(x_0)\}]}{\nabla\mathcal{I}(x_0)} \tag{64}$$

where the factor $[\nabla\mathcal{I}(x_0) - \nabla G_1\{\mathcal{I}(x_0)\}]/\nabla\mathcal{I}(x_0)$ is smaller than one, and tends to zero when the width of the edge increases , *i.e.*, when the steepness of the logistic curve decreases. To simplify notation, we denote this factor $\gamma(x_0)$, and write the second term in (61) as:

$$\begin{aligned} \beta L_1\{\mathcal{I}(x)\} L_1\{\mathcal{I}(x)\} &\approx \beta\gamma(x_0) L_1\{\mathcal{I}(x)\} \left[1 - \frac{\mathcal{I}(x_0)}{\mathcal{I}(x)}\right] \mathcal{I}(x) \\ &\approx \alpha\, |L_1\{\mathcal{I}(x)\}|\, \mathcal{I}(x) \end{aligned} \tag{65}$$

The first approximation is obtained by replacing $L_1\{\mathcal{I}\}$ in (61) using the approximation derived in (64). The second approximation results from the fact that $(1 - \frac{\mathcal{I}(x_0)}{\mathcal{I}(x)})$ is a small value having the same sign as $L_1\{\mathcal{I}(x)\}$. Parameter α is introduced to reflect a reasonable average value for $\beta\gamma(x_0)\left[1 - \frac{\mathcal{I}(x_0)}{\mathcal{I}(x)}\right]$ around the inflexion point of different kinds of edges. In practice, it is set empirically, as discussed in previous section.

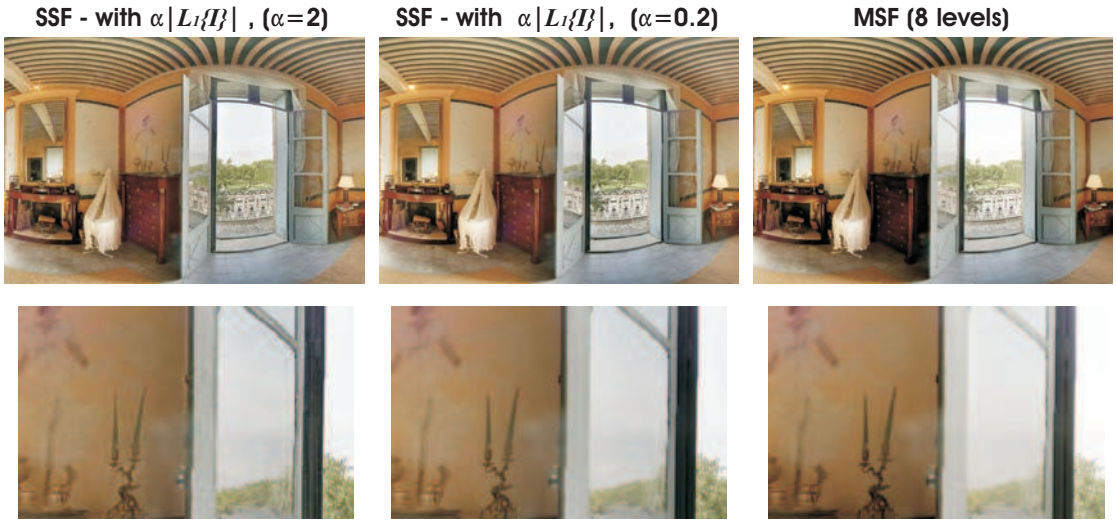

FIGURA 69: *The influence of the parameter α used in our single-scale fusion method (66). If we increase the impact (α=2) of $L_1\{\mathcal{I}\}$, some small artifacts appear around edges. Experimentally, we found that a default value of α=0.2 is well suited to all investigated scenarios.*

FIGURA 70: *Single-scale fusion results generated using (60) and (66). As may be observed, visually and also based on the SSIM evaluation [86] our approach yield very similar results as the traditional MSF approach. Since both single-scale fusion (60) and (66) produce almost identical results, all the results have been generated using the last derivation described in (66) using the default parameter $\alpha = 0.2$.*

Replacing the second term in (61) with the approximation (65), and coming back to the detailed notation (i.e., replacing $\mathcal{I}$ by $\mathcal{I}_k(x)$ and $\bar{\mathcal{W}}$ by $\bar{\mathcal{W}}_k(x)$), the contribution of the k^{th} input to our final simplified SSF formulation becomes:

$$\mathcal{R}_{SSF,k}(x) = \left[G_N\left\{\bar{\mathcal{W}}(x)\right\} + \alpha\left|L_1\left\{\mathcal{I}(x)\right\}\right|\right] \quad \mathcal{I}(x) \tag{66}$$

Since the convolution of two Gaussian kernels is a wider Gaussian kernel, $G_N\left\{\bar{\mathcal{W}}\right\}$ can be directly computed with a kernel whose variance is N times the variance of the initial Gaussian kernel. Hence, no need to apply N times the Gaussian filter to derive $G_N\left\{\bar{\mathcal{W}}\right\}$. By aggregating the contributions of all inputs, our SSF expression becomes $\mathcal{R}_{SSF}(x) = \sum_k \mathcal{R}_{SSF,k}(x)$.

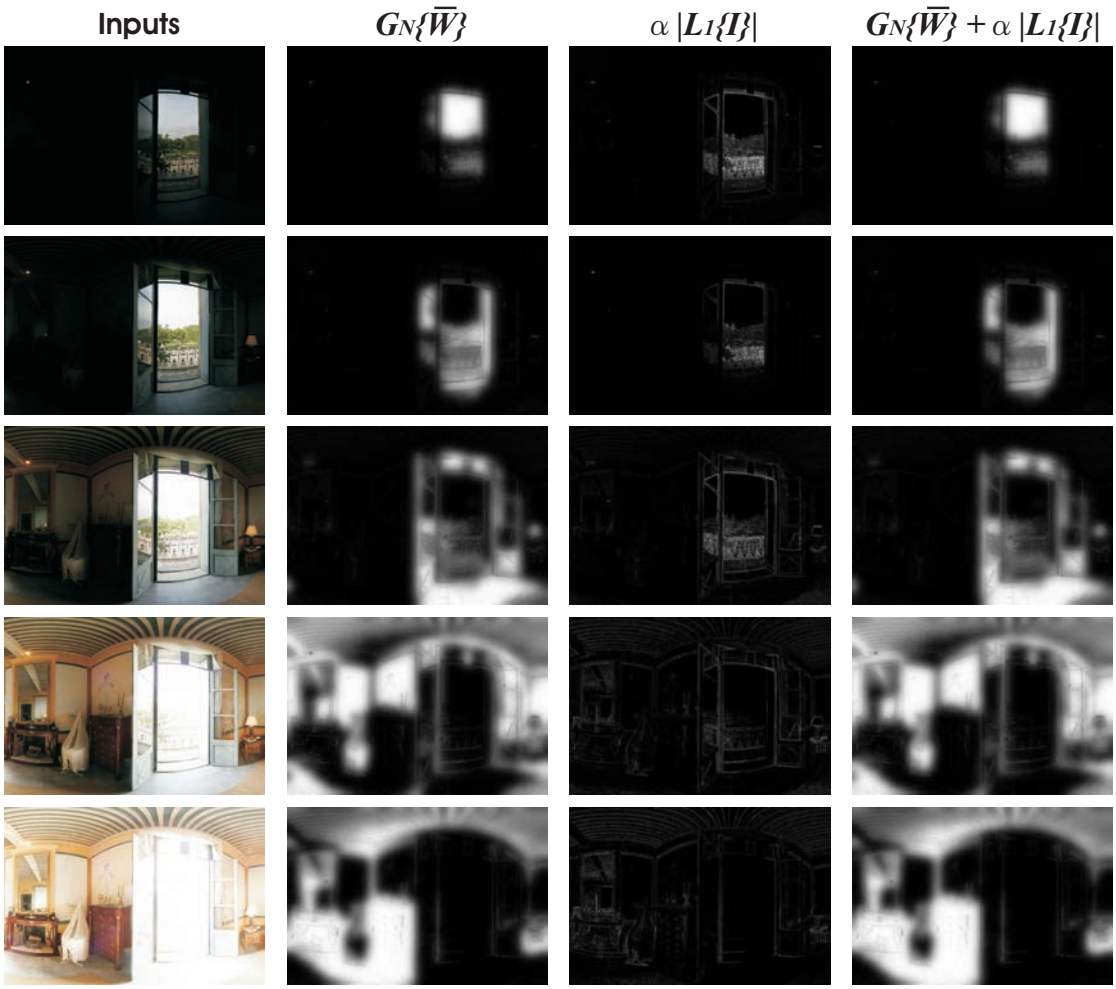

FIGURA 71: *The weights of our final SSF expression (66). From left to right: the inputs, the corresponding two weights ($G_N\{\bar{\mathcal{W}}\}$ and $\alpha\,|L_1\{\mathcal{I}\}|$) of (66) and their sum.*

Fig. 70 depicts differences between results obtained using our single-scale fusion (60) and (66) and the MSF expression. As can be observed visually and also based on the SSIM evaluation [86], both (60) and (66) yields very similar results compared as traditional MSF. While both SSF equations (60) and (66) produce almost identical outputs, all the results presented in this part have been generated using the last derivation described in (66) using the default parameter $\alpha = 0.2$. Interestingly, (66) increases the magnitude of the weights in the image regions with large Laplacian values, thereby reinforcing edges in the blended outcome. This is similar sharpening an image by subtracting from it a fraction of its Laplacian. This observation offers a novel perspective with respect to understanding the success of MSF: MSF promotes the regions of the images with high Laplacian magnitude, thereby reinforcing the contrast of the blended image.

II.4.3 Results and Discussion

Since the primary contribution of this part lies in the simplification of multi-scale fusion algorithm, our validation primarily aims at demonstrating that our proposed single scale simplification is valid in a large variety of use cases. We begin by introducing a set of weights commonly used in multiscale fusion, then validate our single scale fusion strategy on a variety of application problems.

As previously mentioned, due to its robustness and simplicity, multi-scale fusion (MSF) based on the Laplacian pyramid decomposition is employed in a wide variety of image processing tasks. However, although the concept remains the same, the solution may vary based on the inputs that are processed and on the criteria (quality measures) that are used to derive their associated weight maps. Here we employ four of the most general quality measures used in previous fusion-based approaches [1, 4, 41, 69]: local contrast, saturation, exposedness and saliency.

Local contrast weight map measures the amount of local variation of each input and is computed by applying a Laplacian filter to the luminance of each processed image. As shown in [1, 41] this assigns high values to sharp transitions in images such as edges and texture by computing the absolute value of the Laplacian response.

Saturation weight map enables algorithms to adapt to chromatic information by boosting the luminance of highly saturated regions. This measure is usually computed [4, 41] as the standard deviation within color channels around each pixel location. This is motivated by the observation that saturated color pixels take large values on at least one or two color channels. This weight map is simply computed (for each input $\mathcal{I}_k$) as the deviation (for every pixel location) between the R,G and B color channels and the luminance L of the input $\mathcal{I}_k$:

$$\mathcal{W}_{k,C} = \sqrt{1/3\left[(R_k - L_k)^2 + (G_k - L_k)^2 + (B_k - L_k)^2\right]} \tag{67}$$

Exposedness weight map estimates the degree to which a pixel is exposed. This weight promotes a constant appearance of local contrast, neither exaggerated nor understated. Pixel values are generally better exposed when they have normalized values, closer to the average value, as in [41, 69]. This measure for input $\mathcal{I}_k$ is

FIGURA 72: *HDR imaging. Comparison of single-scale fusion (SSF) with multi-scale fusion (MSF) for HDR imaging. Also shown are comparative results with several well known tone mapping techniques [166–172].*

expressed as a Gaussian-modeled distance to the average normalized value (0.5):

$$\mathcal{W}_{k,E} = exp\left(-\frac{(\mathcal{I}_k - 0.5)^2}{2\sigma^2}\right) \tag{68}$$

where the standard deviation is set to $\sigma = 0.25$.

Saliency weight map identifies the degree of local visual conspicuity, by highlighting visually attractive regions of an image. As in the recent fusion techniques of [1, 69] we employ the well-known saliency technique of Achanta et al. [55]. Its computation is inspired by the biological concept of center-surround contrast being computed as a difference between a Gaussian smoothed version of the input and its mean value. The saliency weight is defined as:

$$\mathcal{W}_{k,S} = \|\mathcal{I}_{k,\omega_{hc}} - \mathcal{I}_{k,\mu}\| \tag{69}$$

where $\mathcal{I}_{k,\mu}$ is the arithmetic mean of the input $\mathcal{I}_k$ while $\mathcal{I}_{k,\omega_{hc}}$ is a Gaussian filtered version of the same input.

To derive the $\bar{\mathcal{W}}_k$ maps, those four weight maps are first summed up for each input image k. The K resulting maps are then normalized on a pixel-per-pixel basis, by dividing the weight of each pixel in each map by the sum of the weights of the same pixel over all maps.

In the following sections we will briefly discuss several well-known fusion-based applications and show that our single-scale fusion (SSF) method produces highly competitive results compared to traditional multi-scale fusion (MSF). We will then discus the advantages of SSF in terms of computational complexity and ease of implementation.

High Dynamic Range Imaging.

Various tone mapping techniques [166–172] aim to create a LDR depiction from an HDR image by compressing the wide dynamic range to a narrower range. Conversely, a well-known HDR imaging approach, exposure

fusion [41] skips the step of computing a HDR image, and immediately fuses the multiple exposures into a high-quality, low dynamic range image that is ready for display.

We compare our single-scale fusion approach with the well-known exposure fusion technique of Mertens et al. [41], which extends the original MSF approach of Burt and Adelson [56]. For a fair evaluation we use the same weight maps in the process of fusing the multiple exposure images.

FIGURA 73: *HDR imaging. Comparison of the MSF with SSF for several set of images used in the TMQI evaluation shown in Table 3. From left to right and top to bottom (**arno, belgium house, cave, chairs, chinese garden, kluki, mask, ostrow, memorial, laurentian library**).*

Figure 72 shows comparative results between SSF and MSF (the exposure fusion approach of [41]) and also the results generated by several tone mapping techniques [166–172] that have been generated by using the publicly available software *Luminance HDR*[3].

[3]http://qtpfsgui.sourceforge.net/

	S score		**N** score		**Q** score		
Image Name	MSF	SSF	MSF	SSF	MSF	SSF	*SSIM*
arno	0.8305	0.7745	0.7535	0.8311	0.9197	0.9156	0.9640
belgium house	0.8422	0.8274	0.9817	0.9853	0.9566	0.9530	0.9649
cave	0.7453	0.7265	0.7545	0.7730	0.8954	0.8925	0.9571
chairs	0.7313	0.7540	0.8353	0.8412	0.9034	0.9110	0.9854
chateau	0.7971	0.8110	0.8134	0.8518	0.9195	0.9291	0.9793
chinese garden	0.8554	0.8480	0.8137	0.8342	0.9358	0.9368	0.9869
foyer	0.8345	0.8262	0.8861	0.9021	0.9407	0.9407	0.9862
grandcanal	0.8347	0.8273	0.7850	0.7868	0.9257	0.9240	0.9483
kluki	0.8658	0.8484	0.7805	0.8885	0.9336	0.9449	0.9770
laurentian library	0.8431	0.8419	0.8184	0.8275	0.9331	0.9341	0.9684
mask	0.8351	0.8087	0.9535	0.9461	0.9506	0.9421	0.9853
memorial	0.8746	0.8756	0.7882	0.7813	0.9371	0.9363	0.9636
ostrow	0.8696	0.8302	0.7310	0.7683	0.9270	0.9220	0.9131
Average	0.8276	0.8154	0.8227	0.8475	0.9291	0.9294	0.9677

TABELA 3: Qualitative comparison between MSF and our SSF approach based on the tone mapping metric quality index TMQI [173]. All three constituent sub-indices (S- structural fidelity, N- statistical naturalness, Q- TMQI score) of TMQI are shown, while the last column shows the values of the structure similarity index (SSIM) between the MSF and SSF results.

Qualitative visual evaluation of Fig. 72 and Fig. 73 reveals minor differences between our strategy and the multi-scale fusion approach. We have also performed a detailed quantitative evaluation, by employing a recent specialized model of the quality of images produced by tone mapping operators TMQI [173].

TMQI uses the well known structural similarity (SSIM) index [86, 174, 175] along with a natural scene statistics (NSS) model [176]. TMQI evaluates the quality of the resulted LDR images using the HDR image as a reference. It combines the multi-scale SSIM [174] with a statistical naturalness measure to generate a general TMQI index.

For the quantitative evaluation we tested 13 sets of images (results on ten of them are shown in Fig. 73 while the other three are included in Fig. 72). Table 3 contains the values of all three TMQI indexes (S- structural fidelity, N- statistical naturalness, Q- TMQI overall score) that comprise the TMQI quality assessment model (the values of the TMQI indexes are in the range [0,1]). Besides the TMQI indexes, the last column of the Table shows the SSIM values between MSF and SSF results.

As a general remark, it may be observed that our single-scale fusion strategy delivers similar TMQI results as MSF. However, some structure information may be lost (the index S is slightly lower on SSF as compared with MSF) the naturalness

appearance of the SSF results are slightly improved compared with the MSF results (index N). Indeed, a close inspection of the level of similarity between the SSF and MSF results reveals very little difference. These observations are also supported by the SSIM index values shown in the last column of the Table 3.

General evaluation.

To evaluate the applications described in the following series of sections, we compare the results of multi-scale fusion (MSF) with the output of our single-scale fusion (SSF)approach. Since PSNR has been proven to be an ineffective way of predicting human visual responses to image quality [77], we also compute the well-known structural similarity (SSIM) index [86, 174, 175] on the results. Analyzing the resulting PSNR and SSIM values reveals that both indicate that SSF delivers a good approximation (e.g. the SSIM values were greater than 0.95 for all examples in our experiments) to the MSF technique.

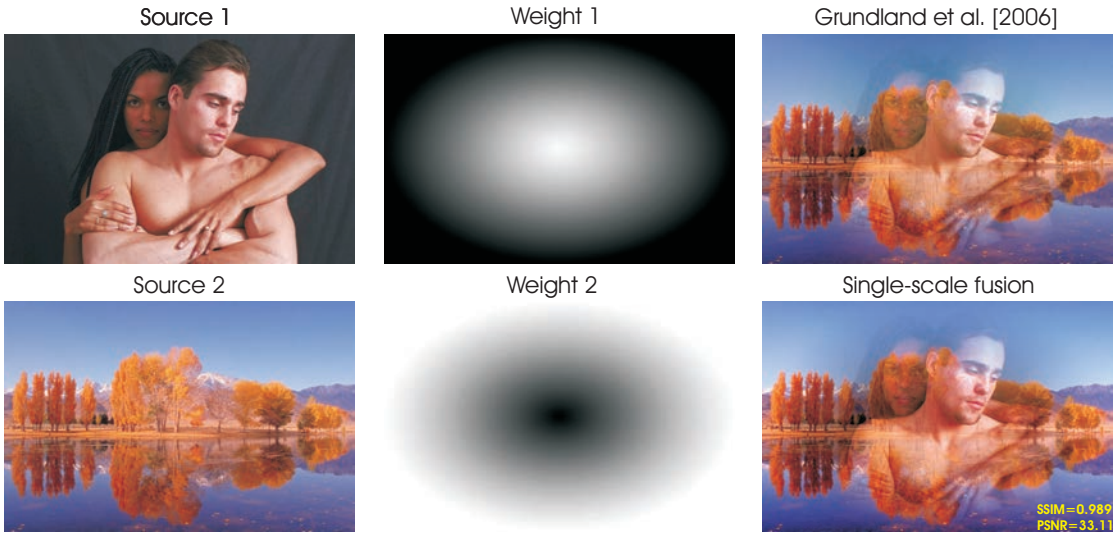

FIGURA 74: *Image Compositing.*

Image Compositing.

Image compositing is an important image/video editing task that deals with the problem of combining component images in order to generate an integrated composite image. Known also as photomontage, this artistic technique has been considered since the advent of photography [177]. Overlayed or superimposed images are combined with the aim of transmitting artistic thoughts or expressions to the viewer. Image compositing challenges consist of preserving the contrast, the sharpness and creating seamless transitions in the composed output.

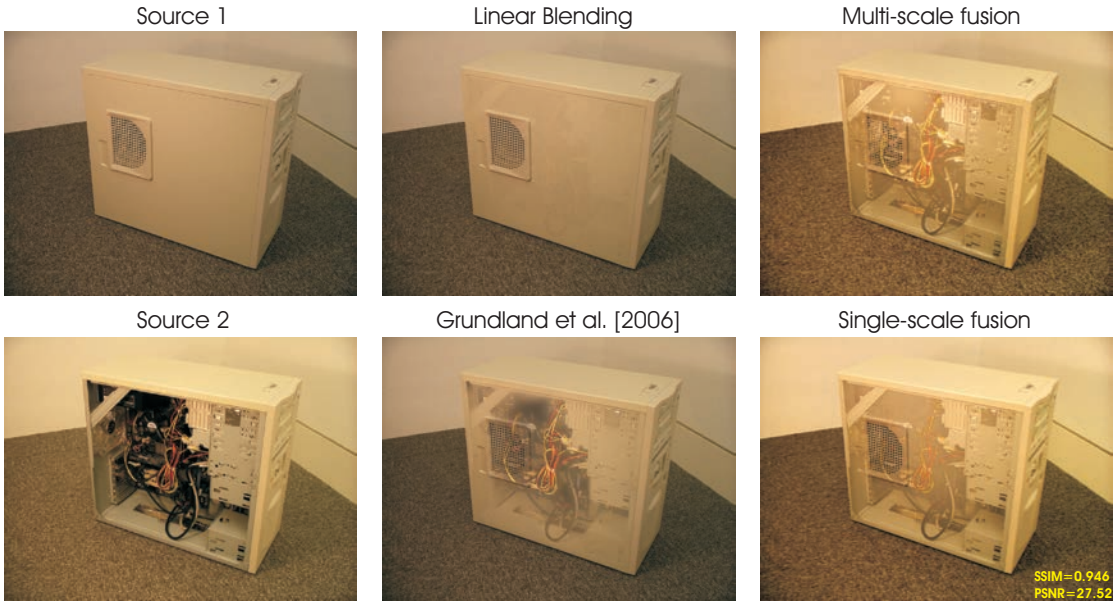

FIGURA 75: *Image Compositing.*

Multi-scale fusion has been successfully applied for this task [39]. As shown in the examples in Fig. 74, our technique performs on a par with the specialized multi-scale fusion approach of [39] and also with the classical MSF approach using the same weight maps as the ones described in the beginning of this section. Moreover, it may be observed that our algorithm preserves the degree of apparent local contrast as well as salient regions, while seamlessly blending multiple inputs (see Fig. 75 for another example).

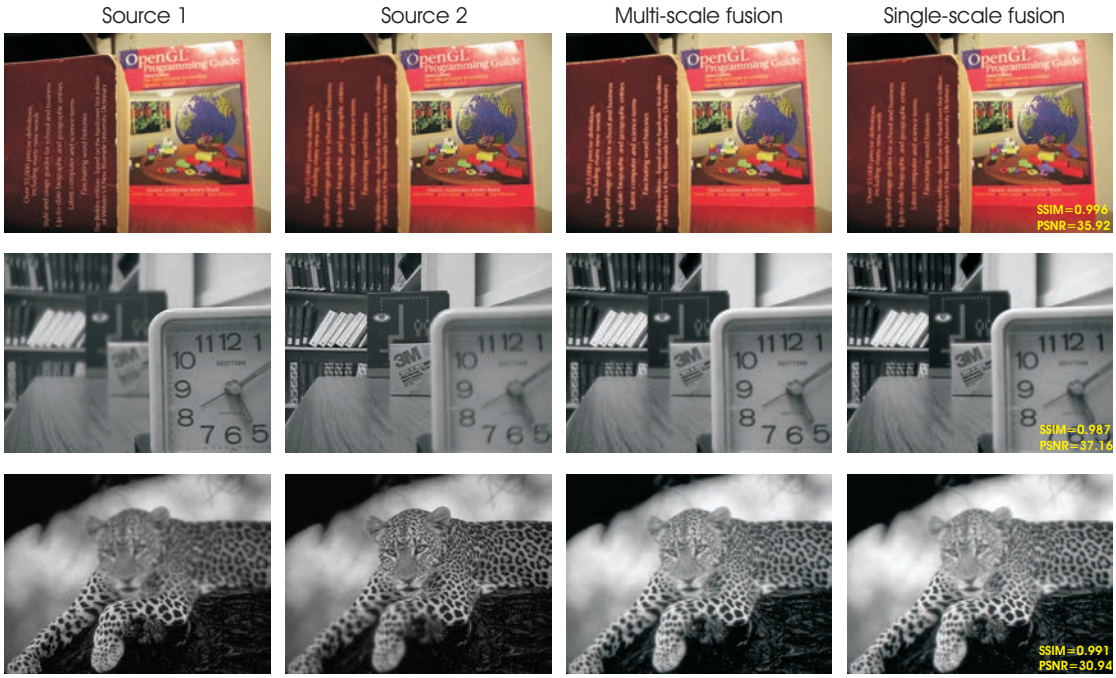

FIGURA 76: *Extended Depth of Field.*

Extended Depth-of-Field.

This task seeks to blend several images that were obtained by focusing at different depths to create an output image having an extended focal range. Extended depth-of-field methods have utility in fields such macro photography and microscopy [36], where the depth of field may be extremely limited. This task was performed automatically over the entire image for the first time in[56] using a multi-scale fusion strategy based on Laplacian pyramids. Figure 76 demonstrates that SSF is able to produce comparable results as traditional MSF techniques (using the same weight maps).

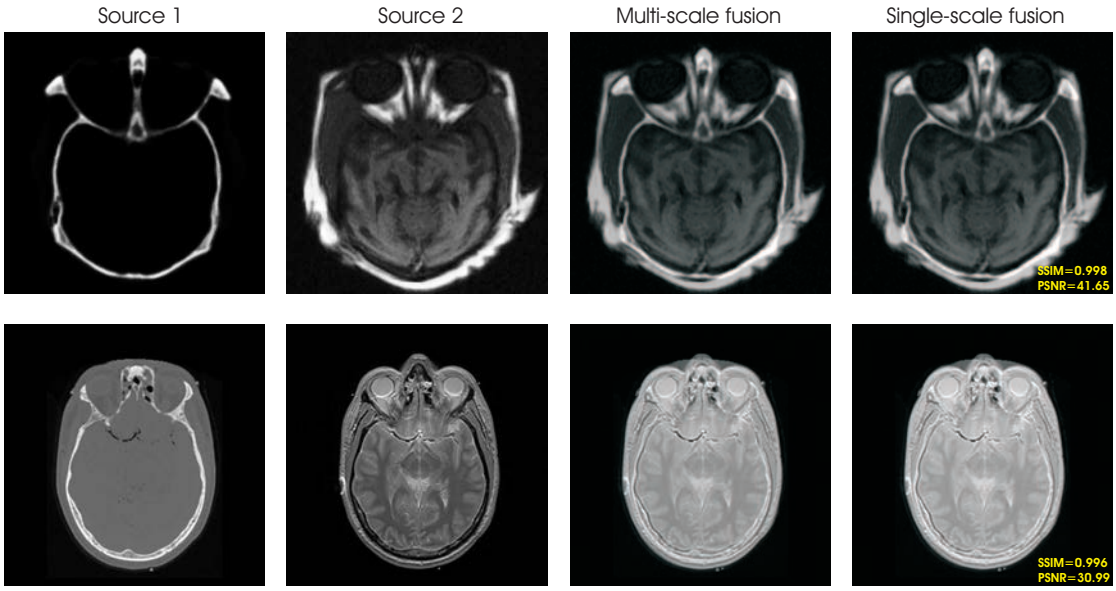

FIGURA 77: *Medical imaging: fusing MRI/CT images.*

Medical Imaging.

In the medical field, image fusion is important for integrating multi-modal images into a single output result that may contain more details and a more complete depiction. For instance, combining MRI with CT images [144, 178] is a common strategy that yields a more accurate description of the scanned body, since information provided by these different scanning techniques may provide complementary information. Such MRI/CT fused output images have been shown to provide both anatomical and functional information that can be important for planning surgical procedures. Using two well-known MRI/CT image pairs [178], Fig. 77 demonstrates that our simplified approach is able to yield results that preserve the contrast and finest details in a manner similar to the classical MSF approach (observe also the overlaid SSIM index values).

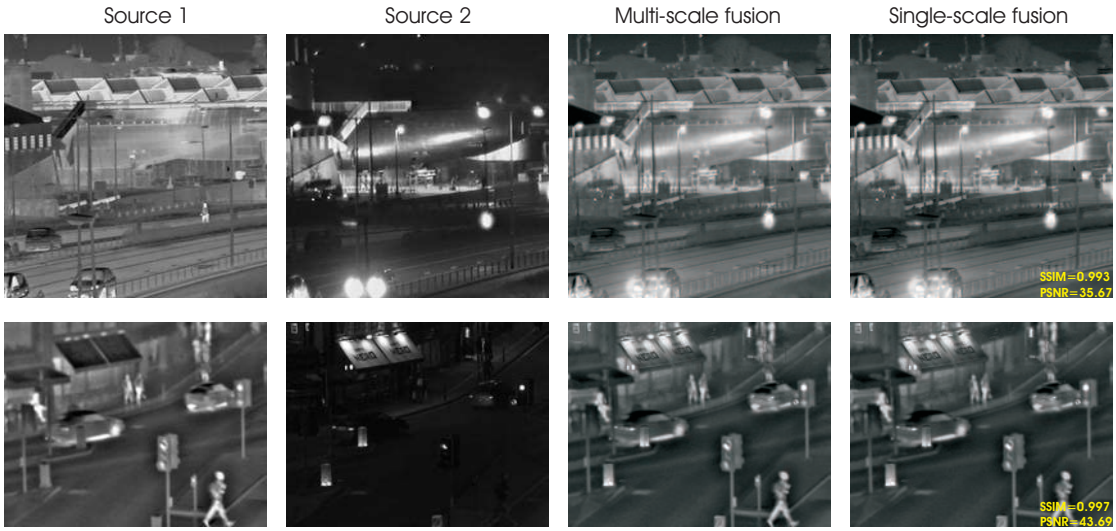

FIGURA 78: *Fusing visible and infra-red information.*

Multi-band Image Fusion.

Multi-band image fusion considers the composition of images from different light frequency bands, such as visible light and IR images. For instance, fusing radar data and IR images can considerably enhance accuracy when estimating the positions of different objects [141, 145]. Additionally, in the context of nighttime surveillance, existing techniques combine the IR image information with visible image data in order to better detect and localize persons in an analyzed scene. Figure 78 presents two examples that fuse the visible with IR information. Both close visual inspection and structure similarity (SSIM index) validation, strongly indicates that the SSF technique produces very similar results as the MSF approach.

	256 × 256		512 × 512		1024 × 1024		2048 × 2048	
	Comp. times	No. of levels	Comp. times	No. of levels	Comp. times	No. of levels	Comp. times	No. of levels
Naive fusion (NF)	0.11	1	0.21	1	0.59	1	2.11	1
Multi-scale fusion (MSF)	0.41	6	0.79	7	2.57	8	8.61	9
Single-scale fusion (SSF)	0.13	1	0.27	1	0.85	1	3.19	1

FIGURA 79: Comparative computation times (expressed in seconds of MATLAB code) of NF, MSF and SSF strategies for different input sizes. As shown , our approach has similar complexity (the same number of levels) as the naive fusion approach, but able to deliver comparable or better results than the multi-scale fusion approach.

While the computation complexity of our SSF technique is similar to that of the naive fusion implementation (please refer to the Table 79), our single scale fusion technique it is able to produce high quality results. Unlike naive fusion (NF), the multi-scale fusion (MSF) approach has the advantage that it can preserve

both relevant high and low frequencies into the final result, while strongly mitigating most visual artifacts (see Fig. 60). In contrast, the main disadvantage of MSF is a higher computation complexity and more complex memory management procedures. As described in the literature dealing with the resource efficient implementation of multiresolution processes [163, 164], data transfer and cache memory management is non-trivial in systems that process signals at distinct resolution levels, simply because manipulating multiple resolutions penalizes memory access bandwidth (more data are manipulated) and/or memory access locality (when subband coefficients associated to the same image location are stored in distinct parts of the memory). The problem is especially critical for embedded systems with highly constrained resources. On general purpose platforms, the memory access platform is less prominent. In this case however, the computational complexity is still larger for MSF than for SSF.To compare the SSF and MSF computational complexity, Table 79 presents the running time of different fusion algorithms for different image sizes. Codes have been written in Matlab, and run on [CPU i7 , 8GB RAM]. As expected, our single-scale fusion (SSF) approach has the same running time as naive fusion (NF) strategy , which is significantly faster than MSF. This reflects the advantage of implementing the fusion as a single scale procedure.

II.4.4 Conclusions

We have introduced a simplified single-scale approximation to the well-known multi-scale fusion based on the Laplacian decomposition. Before introducing our single scale strategy for fusing multiple images, we first identify the most critical components of the traditional MSF that helps to explain why MSF performs so well. Our SSF method has a complexity comparable to the naive fusion solution. However, our extensive qualitative and quantitative evaluations demonstrate that our simplified fusion approach has the advantage to produce similar high quality results as the multi-scale fusion approach.

In the future work, we plan to explore the use of perceptually relevant natural scene statistics [179] to perceptually optimize the fusion process [180].

Partea III

Development and Future Directions of the Academic Career

III.1 Proposal for career development at the didactic level

Currently I am lecturer in Department of Measurements and Optical Electronics (MEO) of the Faculty of Electronics, Telecommunications and Information Technologies (ETcTI), Timisoara (UPT). I started my academic activity within the UPT, as professor assistant in 2012. On an academic level, I had a considerable experience, considering that since 2014, I was meeting the minimum criteria for getting the Habillitation in Romanian academic system.

Within this department, the didactic activity has included the subjects for the bachelor's and master's degree program: *Electronic and Optoelectronic Devices (Dispozitive Electronice si Optoelectronice)*, *Computer Programming and Utilization (Programarea si Utilizarea Calculatoarelor)*, *Interface of Measurement and Testing Systems Techniques (Interfatarea sistemelor de masurare si testare)* and *Computer Vision (Vederea Artificiala)*. Also, at the University of Bremen, Germany, at Master degree I taught the course on *Real-Time Image Processing* in the years 2015 and 2016 through the Erasmus Plus program.

During 2018, I proposed the introduction of a number of 7 new courses (some of them required by the Timisoara industrial segment) that include: *Computer vision*, *Pattern Recognition and Image Analysis*, *Advanced Image Processing*, *Deep Learning*, *LabView machine vision*, *Techniques of Measurement in Bioinformatics*, *ADAS Imaging and Computer Vision Techniques*, *Medical Image Processing*. Among these courses, only the *Computer Vision* course has been accepted for exchange with *Interface of Measurement and Testing Systems Interface* course.

I consider that an important aspect related to didactic activity is the elaboration of didactic materials that aim to facilitate a better understanding of the taught subjects. Together with my colleagues within the MEO department, I was involved in developing specialized materials for the courses *Electronic and Optoelectronic Devices (Dispozitive Electronice si Optoelectronice)*, *Computer Programming and Utilization (Programarea si Utilizarea Calculatoarelor)* and *Computer Vision (Vederea Artificiala)*. Additionally, I have prepared the *Real-Time Image Processing course* (at University of Bremen, Germany). Through these materials, I'm convinced that I've succeed to awaken the interest of my students who have begun to ask each year for more bachelor and dissertation themes related to image processing and computer vision.

In my didactic activity I intend to use interactive teaching methods (such as Brainstorming) to stimulate active participation and involvement of students in coursework, or practice work centered on discovery, team learning and group learning. I propose to diversify the teaching methods towards collaborative creativity and educational partnership with universities and companies from Romania and abroad. In this context, I want to improve both the teaching activity and the didactic material required for the students. The main purpose of this didactic development plan is to familiarize the students with the development of the projects and to understand the requirements of the research activity.

In addition, in the near future I would like to propose several new courses tailored to the industry/market requirements and meet the required quality standards. I want to get involved in introducing and implementing courses that use the knowledge of image processing for mobile platforms, programming concepts, advanced computer vision/pattern recognition and machine learning / deep neural networks, with up-to-date materials showing the latest achievements in the research field.

At the moment I have some important achievements in the field of Computer Vision / Image Processing research, which allow me to yield high qualitative courses in line with international standards. Note that the journal *Color Balance and Fusion for Underwater Image Enhancement* published at the prestigious IEEE Transactions on Image Processing (Impact Factor 5.071) (link access: https://ieeexplore.ieee.org/document/8058463) **is ranked on the first place as the most popular article published all time in IEEE Transactions on Image Processing** (achieving more than 30,000 full-text views in 15 months since publishing).

During my academic career I have been involved in coordinating more than 12 students for both the bachelor's and the dissertation projects, with topics related to image processing and computer vision. Among the elaborated topics I would like to mention: Techniques for stereo matching and stereo occlusion filling, Fusion of medical images such as MRI-PET-CT, Multi-sensor fusion, Saliency detection with Raspberry Pi, Face detection with RaspberryPi, Advanced fusion techniques, Image processing developed with RaspberryPi, Management system building (BMS) developed on the Arduino platform. I would like to mention that the number of supervised students (within the UPT) has been influenced by the fact that in 2013 I was continuing my activity at UPT during my 2 years maternity leaving (which I did not take it, since I started my activity as professor

assistant at MEO in that period). Also during 2015-2017 I suspended my teaching activity at UPT, in order to carry out the research activities at the University of Girona (Spain) within the project financed by Techniospring TECSPR14-2 -0023, (128,920.00 euro) *Enhancing Techniques for Navigation System of Auv.* During this time I was involved in managing activities regarding the PhD student Klemen Istenik of University of Girona, whose research theme was closely related to the processing of underwater images.

Moreover, in the next period I would like to be involved in training of specialists in measurement-related areas such as image acquisition sensors and image processing for the monitoring and testing process. For this purpose, I have proposed the new course *Measurement and evaluation techniques in imaging* that I hope in the not too distant future to be accepted in one of the master's programs of the faculty.

I also intend to participate in the development of laboratories by purchasing high-end equipment and making the materials necessary for the teaching process by attracting public funds (national and international projects) and private (local and not only) companies. Over the past years, I have sent several project proposals (*Tinere Echipe*, *ARUT*, *Techniospring*, *Marie Curie*) where I usually have obtained high scores (of over 90%), but unfortunately these efforts did not allow me to obtain finance for all of them.

In addition, I would like to mention that I am part of the doctoral commission of PhD student Kiss Arpad (who is coordinated by S.l. Dr. Ing. Habil. Cosmin Ancuti), whose doctoral thesis *Improving the visibility for performance enhancement in ADAS systems* is closely related to my current research directions. Recently, my project *Baze de imagini afectate de ceață pentru testarea și evalurea metodelor de clasificare în traficul rutier* was awarded with a 10,000 euro grant in the ARUT 2018 competition. Ing. Kiss Arpad (PhD candidate) is currently included in the team of this research project. This grant includes funds for professional photo equipment, one perfromant computer equipped with NVIDIA and also to fund my participation and the PhD student Kiss Arpad at least to one international conference.

Regarding future PhD candidates management, they will have the possibility to collaborate and learn from the research teams that already exists in ETcTI. Currently, I am invited at monthly meetings of the PhD and master students of Prof.

Caleanu Catalin. I am aware of the current resources and I am sure that this team will welcome new dedicated students.

I also intend to motivate the outstanding undergraduate students to integrate into my research team, with the prospect of enrolling in doctoral studies under my leadership and beyond (after acquiring the skills). An important aspect that I will consider is the possibility of involving my future PhD students in collaboration with local industry partners, being convinced that in this way the students will acquire the competence to participate in the more complex research projects within the university as well as in collaborative projects with the private partners.

Last but not least, I propose to promote and update the ETcTI faculty specializations in order to attract future undergraduate students, master students and PhD students. For this purpose, together with Prof. Dr. ing. Habil. Catalin Caleanu and S.l. Dr. Ing. Habil. Cosmin Ancuti we started the process of introducing a new master-research oriented program in Computer Vision and Artificial Intelligence that we believe will have an important impact in the near future both in the research activities of our faculty but also in collaborations with local industry partners.

III.2 Proposal for career development at the research level

My extensive experience in academic research has been motivated me to continue my work and build an academic carrier. The habilitation will enable me to undertake a novel and an exciting role and also the possibility to engage future PhD students to achieve professional maturity and independence.

During my early academic career, I was focused on several fundamental problems of computational photography. Computational photography is a fairly novel research topic that merges computer graphics and computer vision techniques, referring broadly to sensing strategies and algorithms that extend the capabilities of digital photography. Among them image decolorization [68, 132] was one of the most important direction. Since image decolorization is quite subjective, we believe that one single solution will hardly satisfy all the expectations. The main reason is the fact the color is a perception information and is the way our brain respond to the wavelengths. However, the main target of image processing is to

preserve the perceptual aspects of the image. One of the major drawbacks is influenced by the saliency algorithm, that identifies the most salient regions in the image. Visual salience is considered to be the strategy of our brain that speed-up selection and processing of items and locations. In many situations the image content plays a key role and therefore saliency detection may fail. In future work, we intend to extend the saliency algorithm by extending its processed information with additional data such as scene classification, or objects identification. During tests of the fusion algorithm, we have observed that since only the global contrast differences are accounted, the employed saliency can fail to produce a consistent regional contrast. We plan to extend this technique in order to account simultaneously for global contrast and the spatial coherence. We believe that the algorithm can demonstrates its robustness and to be included by other applications such as edge preserving filtering, single image depth estimation and segmentation. Additionally, we intend to investigate the potential of our algorithm for several other pattern recognition applications that are still relying to the standard decolorizing technique. Last but not least, we would like to introduce an objective color-to-grayscale measure capable to deal with asses certain image qualities such as the potential of preserving the original discriminability and also the finest details.

Another direction is image dehazing [1, 181, 182]. We have addressed the more complex case when the images are characterized by non-homogenous haze for single images [2, 5] for both night-time and underwater images. We would like to investigate a more comprehensive optical model that would allow to restore more generally the images. We also aim to introduce new statistical based observations, that will allow better understanding of the haze physical phenomena. We believe that a promising direction is to combine the dehazing with superresolution and image denoising techniques, since in many cases the recovered regions require further enhancement mainly due to the noise.

I have also extended my experience in several fundamental computer vision problems that are built on local feature points [61, 183–186]. Feature points have extensively demonstrated their utility in various computer vision-related applications, considering their ability to depict distinctively local but also global content of images. In the future work we would like to extend the classic approach by considering the global information, that we believe also needs to be more deeply analyzed. Aside from segmentation, information of the texture represents an important global characteristic that is a challenge that should be explored more

efficiently, in order to filter feature points that have similar properties. In addition, since we involved only SIFT in our robustness demonstrations (regarding photometric transformations), we plan to built up a comprehensive evaluation of the proposed approach and to test it against other detectors descriptors which are concerned.

Regarding the image deblurring in our previous work [187, 188] the strategy is to take additional information by using sharp images of the same scene. Moreover, to increase the robustness of our outputs, we plan to consider several changes into the reference image such as additional noise and variation of the lighting. Moreover, I plan to improve the robustness of our algorithm for larger geometric distortion between images and also to consider more challenging blurs (e.g. shift-variant, out-of-focus). A more challenging research direction is the case when both photographs have been damaged by unknown motion blur. This research area presents a great potential for complex applications, and it is known that deconvolution technique can deal with other restoration problems that require edge preservation, such as image denoising or surface reconstruction.

To validate and asses the performance of the s.o.t.a. dehazing methods we have introduced D-HAZY (http://ancuti.meo.etc.upt.ro/D_Hazzy_ICIP2016/) the first database for dehazing, with artificially generated images that is containing a number of +1400 images, of various scenes. Recently, in the summer of 2018 we have introduced the new O-HAZE(http://ancuti.meo.etc.upt.ro/O-HAZE/) and I-HAZE(http://ancuti.meo.etc.upt.ro/I-HAZE/) bases, which are the first databases containing both foggy images and the ground-truth reference images for outdoor and indoor scenes.

Considering our expertise using fusion fusion strategies [1, 68, 186], we plan to enlarge the explored area to various topics such as satellites remote sensing image fusion, image saliency, image detail manipulation and segmentation.

I would also like to extend my experience with image enhancement using machine learning. We plan to investigate the possibility to generalize the optical model for all hazy cases (day-time, night-time and underwater).

Last but not least, in collaboration with my future PhD students, I will continue to carry out and manage different research directions in computer vision and computational photography. This will be reflected in my publication activity and also by

the activity as reviewer for different prestigious publications such as IEEE Transactions on Image Processing, International Journal of Computer Vision (IJCV), Computer Graphics Forum Journal, Pattern Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Pattern Recognition Letters, Visual Computer but also to the main computer vision conferences: IEEE Computer Vision and Pattern Recognition (CVPR), IEEE International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV).

As a professor, my intentions are to cover my area of expertise such as classes in computer science but also subjects related to computational photography, although there are many adjacent subjects that also has arose my interest. I'm convinced that it represents a great opportunity to teach and review fundamental techniques in related areas such as mathematics and computer science. I strongly believe that my devotion for research will be reflected also in comprehensive and well-illustrated courses.

As a researcher, my eagerness to learn new things reflects a high degree of motivation. I believe that a professor helps constructing the flow of the lab, both scientifically and personally. As well, I aim to foster and stimulate scientific discussions, write papers and grants, teach others experimental design and analysis, and present the findings of my research. My interests in academic education range widely, although teaching experience shaped and is been shaped by my research interests. My plans revolve around the development of usable, high-fidelity computational photography and computer vision techniques. This habilitation will reinforce my role and will enable me to undertake new collaborations by exchanging experience with researchers and students from other universities. It is a great opportunity to motivate my PhD students to develop novel and effective algorithms and practical tools in subjects such as: HDR imaging, super resolution, deblurring, white balancing, image matching, dehazing, underwater image and video restoration, machine learning, etc. In addition, considering my personal experience, I will actively support them to get involved in different national and international grants.

Bibliografie

[1] 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.

[2] C. De Vleeschouwer C. Ancuti, C. O. Ancuti and A. C. Bovik, "Night-time dehazing by fusion," *IEEE ICIP*, 2016.

[3] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-hazy: A dataset to evaluate quantitatively dehazing algorithms," *IEEE International Conference on Image Processing (IEEE ICIP)*, 2016.

[4] Cosmin Ancuti, Codruta O. Ancuti, Tom Haber, and Philippe Bekaert, "Enhancing underwater images and videos by fusion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[5] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Multi-scale underwater descattering," *International Conference on Pattern Recognition (ICPR 2016)*, 2016.

[6] C. Ancuti and C. O. Ancuti, "Laplacian-guided image decolorization," *IEEE ICIP*, 2016.

[7] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and A. C. Bovik, "Single-scale fusion: An effective approach to merging images," *IEEE Transactions on Image Processing*, 2016.

[8] P.S. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sensing of Environment*, 1988.

[9] G. Dal Moro and L. Halounova, "Haze removal and data calibration for high-resolution satellite data," *Int. Journal of Remote Sensing*, 2006.

[10] S.G. Narasimhan and S.K. Nayar, "Contrast restoration of weather degraded images," *IEEE Trans. on Pattern Analysis and Machine Intell.*, 2003.

[11] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo- Model-based photograph enhancement and viewing," in *ACM Transactions on Graphics*, 2008.

[12] T. Treibitz and Y. Y. Schechner, "Polarization: Beneficial for visibility enhancement?," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[13] Raanan Fattal, "Single image dehazing," *SIGGRAPH, ACM Transactions on Graphics*, 2008.

[14] Robby T. Tan, "Visibility in bad weather from a single image," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[15] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[16] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *IEEE International Conference on Computer Vision*, 2009.

[17] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," *In IEEE International Conference on Computer Vision*, 2009.

[18] H. Koschmieder, "Theorie der horizontalen sichtweite," in *Beitrage zur Physik der freien Atmosphare*, 1924.

[19] Giuseppe Mancino, Angelo NoleCorresponding, Francesco Ripullone, and Agostino Ferrara, "Landsat tm imagery and ndvi differencing to detect vegetation change: assessing natural forest expansion in basilicata, southern italy," *Journal of Biogeosciences and Forestry*, 2013.

[20] G. dal Moro and L. Halounovra, "Haze removal for high-resolution satellite data: a case study," *International Journal of Remote Sensing*, 2007.

[21] B. Guindon Y. Zhang and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images," *Remote Sensing of Environment*, , no. 82, 2002.

[22] Y. Zhang and B. Guindon, "Quantitative assessment of a haze suppression methodology for satellite imagery: effect on land cover classification performance," *IEEE Transactions on Geoscience and Remote Sensing*, , no. 41, 2003.

[23] H. B. Mitchell, "Image fusion: Theories, techniques and applications," *Springer*, 2010.

[24] N. Hautiere, J.-P. Tarel, D. Aubert, and E. Dumont, "Blind contrast enhancement assessment by gradient ratioing at visible edges," *Img. Anal. and Stereology*, 2008.

[25] Ko Nishino, Louis Kratz, and Stephen Lombardi, "Bayesian defogging," *International Journal on Computer Vision*, 2012.

[26] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images," *Remote Sensing of Environment*, 2002.

[27] S.G. Narasimhan and S.K. Nayar, "Vision in bad wheather," in *In IEEE ICCV*, 1999.

[28] S.G. Narasimhan and S.K. Nayar, "Chromatic framework for vision in bad weather," in *In IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[29] S.G. Narasimhan and S.K. Nayar, "Interactive de-wheathering of an image using physical models," *ICCV Workshop CPMVC*, 2003.

[30] N. Hautiere, J.-P. Tarel, and D. Aubert, "Towards fog-free in-vehicle vision systems through contrast restoration," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[31] Y. Y. Schechner, S.G. Narasimhan, and S.K. Nayar, "Polarization-based vision through haze," *App. Opt.*, , no. 42, pp. 511–525, 2003.

[32] S. Shwartz, E. Namer, and Y.Y. Schechner, "Blind haze separation," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[33] Y.Y. Schechner and Y. Averbuch, "Regularized image recovery in scattering media.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007.

[34] E. Namer, S. Shwartz, and Y.Y. Schechner, "Skyless polarimetric calibration and visibility enhancement," *Optic Express*, , no. 17, pp. 472–493, 2009.

[35] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," *In IEEE TPAMI*, 2008.

[36] A. Agarwala, M. Dontcheva, M. Agrawala, S. M. Drucker, A. Colburn, B. Curless, D. Salesin, and M. F. Cohen, "Interactive digital photomontage," *ACM Trans. Graph (SIGGRAPH)*, 2004.

[37] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph (SIGGRAPH)*, 2003.

[38] R. Brinkmann R., "The art and science of digital compositing," *Morgan Kaufmann*, 1999.

[39] 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.

[40] P. J. Burt, K. Hanna, and R. J. Kolczynski, "Enhanced image capture through fusion," *In IEEE Int. Conf. on Computer Vision*, 1993.

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

[42] L. Schaul, C. Fredembach, and S. Susstrunk, "Color image dehazing using the near-infrared," *In IEEE Int. Conf. on Image Processing*, 2009.

[43] E. J. McCartney, "Optics of the atmosphere:scattering by molecules and particles," *JohnWiley and Sons*, 1976.

[44] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, 1980.

[45] E. Land and J. McCann, "Lightness and retinex theory," *The Journal of the Optical Society of America A*, vol. 61, 1971.

[46] V. Cardei, B. Funt, , and K. Barnard, "Estimating the scene illumination chromaticity using a neural network," *Journal of the Optical Society of America A*, 2002.

[47] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," *12th Color Imaging Conference*, 2004.

[48] J. van de Weijer, Th. Gevers, and A. Gijsenij, "Edge-based color constancy," *In IEEE Tran. on Image Processing*, 2007.

[49] Marc Ebner, *Color Constancy, Wiley 1st edition*, 2007.

[50] Arjan Gijsenij and Theo Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 687–698, 2011.

[51] Joost van de Weijer and Theo Gevers, "Color constancy based on the grey-edge hypothesis," in *IEEE International Conference on Image Processing (ICIP)*, 2005.

[52] J. K. Tsotsos, S. M. Culhane, W. Y. K.Wai, Y. Lai, N. Davis, and F. Nuflo., "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1–2, pp. 507–545, 1995.

[53] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *ACM Multimedia*, pp. 374–381, 2003.

[54] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[55] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

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

[57] Z. Farbman et al., "Edge-preserving decompositions for multi-scale tone and detail manipulation," *SIGGRAPH*, 2008.

[58] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Effective single-image dehazing by fusion," in *In IEEE ICIP*, 2010.

[59] Codruta Orniana Ancuti, Cosmin Ancuti, Chris Hermans, and Philippe Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," *ACCV*, 2010.

[60] R. Fattal, "Dehazing using color-lines," *ACM Trans. on Graph.*, 2014.

[61] Cosmin Ancuti and Codruta O. Ancuti, "Effective contrast-based dehazing for robust image matching," *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 2014.

[62] S. Emberton, L. Chittka, and A. Cavallaro, "Hierarchical rank-based veiling light estimation for underwater dehazing," *Proc. of British Machine Vision Conference (BMVC)*, 2015.

[63] K. Tang, J. Yang, and J. Wang, "Investigating haze relevant features in a learning framework for image dehazing," *IEEE CVPR*, 2009.

[64] S. C. Pei and T.Y. Lee, "Nighttime haze removal using color transfer pre-processing and dark channel prior," *In IEEE Int. Conf. Image Processing*, 2012.

[65] J. Zhang, Y. Cao, and Z. Wang, "Nighttime haze removal based on a new imaging model," *In IEEE Int. Conf. Image Processing*, 2014.

[66] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," *In IEEE Int. Conf. on Computer Vision*, 2015.

[67] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. on Pattern Analysis and Machine Intell.*, 2011.

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

[69] 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.

[70] C. Ancuti and C. O. Ancuti, "Effective contrast-based dehazing for robust image matching," *IEEE Geoscience and Remote Sensing Letters*, 2014.

[71] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," *In IEEE Int. Conf. on Computer Vision*, 2013.

[72] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[73] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *In IEEE Trans. on Image Processing*, 2012.

[74] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *In IEEE Signal Processing Letters*, 2013.

[75] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *In IEEE Trans. on Image Processing*, 2012.

[76] J.-P. Tarel, N. Hautière, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, 2012.

[77] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Morgan and Claypool Publishers*, 2006.

[78] G. Sharma, W. Wu, and E. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research and Applications*, 2005.

[79] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth.," *In German Conference on Pattern Recognition*, 2014.

[80] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light.," *In IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[81] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," *International Conference on Computer Vision*, 2013.

[82] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[83] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *In SIGGRAPH*, 2004.

[84] K. He, J. Sun, and X. Tang, "Guided image filtering," *In IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2013.

[85] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics Gems IV*, 1994.

[86] H. Wang, R. Raskar, and N. Ahuja, "Seamless video editing," *In International Conference on Pattern Recognition*, 2004.

[87] S. Westland, C. Ripamonti, and V. Cheung, "Computational colour science using matlab, 2nd edition," *Wiley*, 2005.

[88] D.-M. He and G. G. L. Seet, "Divergent-beam lidar imaging in turbid water," *Optics and Lasers in Engineering*, 2004.

[89] Y.Y. Schechner and Y. Averbuch, "Regularized image recovery in scattering media," *IEEE Trans Pattern Anal Mach Intell.*, 2007.

[90] Eric P. Bennett, John L. Mason, and Leonard McMillan, "Multispectral bilateral video fusion," *IEEE Trans. on Image Processing*, 2007.

[91] Codruta O. Ancuti, Cosmin Ancuti, and Philippe Bekaert, "Fusion-based restoration of the underwater images," in *In IEEE ICIP*, 2011.

[92] Raimondo Schettini and Silvia Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *J. on Adv. in Sig. Proc.*, 2010.

[93] Donna M. Kocak, Fraser R. Dalgleish, Frank M. Caimi, and Yoav Y. Schechner, "A focus on recent developments and trends in underwater imaging," *Marine Technology Society Journal*, 2008.

[94] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic aperture confocal imaging," *SIGGRAPH, ACM Trans. Graph.*, 2004.

[95] S. Bazeille, L. Jaulin I. Quidu, and J. P. Malkasse, "Automatic underwater image pre-processing," *In Proc. of CMM*, 2006.

[96] A. Arnold-Bos, J.-P. Malkasset, and G. Kervern, "Towards a model-free denoising of underwater optical images," *IEEE Europe Oceans Conf.*, 2005.

[97] F. Petit, A.-S. Capelle-Laize, and P. Carre , "Underwater image enhancement by attenuation inversion with quaternions," *IEEE ICASSP*, 2009.

[98] S.K. Nayar and S.G. Narasimhan, "Vision in bad weather," *IEEE Int. Conf. on Comp. Vision*, 1999.

[99] D.B. Chenault and J.L. Pezzaniti, "Polarization imaging through scattering media," *In Proc. SPIE*, 2000.

[100] Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE Journal of Oceanic Engineering*, 2005.

[101] B. L. McGlamery, "A computer model for underwater camera systems," *Ocean Optics*, 1979.

[102] Marc Ebner, *Color Constancy, Wiley 1st edition*, 2007.

[103] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," *IS&T/SID Twelfth Color Imaging Conference: Color Science,Systems and Applications, Society for Imaging Science and Technology*, pp. 37–41, 2004.

[104] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge based color constancy," *IEEE Transactions on Image Processing*, 2007.

[105] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *In IEEE International Conference on Computer Vision*, 1998.

[106] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *Int. Journal of Computer Vision*, 2009.

[107] J. Zimmerman, "An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement," *IEEE Trans. on Med. imaging*, 1988.

[108] 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 Transactions on Image Processing*, 1997.

[109] T. Pu and G. Ni, "Contrast-based image fusion using the discrete wavelet transform," *Opt. Eng.*, 2000.

[110] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-S. Seidel, "Dynamic range independent image quality assessment," in *SIGGRAPH, ACM Trans. Graph.*, 2008.

[111] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Comp. Vision*, vol. 20, pp. 91–110, 2004.

[112] G. Papandreou and P. Maragos, "Multigrid geometric active contour models," *In IEEE Trans. on Image Processing*, 2007.

[113] B. A. Levedahl and L. Silverberg, "Control of underwater vehicles in full unsteady flow," *IEEE Journal of Oceanic Engineering*, 2009.

[114] C. H. Mazel, "In situ measurement of reflectance and fluorescence spectra to support hyperspectral remote sensing and marine biology research," *In IEEE Oceans*, 2006.

[115] G. L. Forest, "Visual inspection of sea bottom structures by an autonomous underwater vehicle," *IEEE Trans. Syst. Man and Cyber*, vol. 31, pp. 691–705, 2001.

[116] Y. Kahanov and J. Royal, "Analysis of hull remains of the dor d vessel, tantura lagoon, israel," *International Journal of Nautical Archeology*, 2001.

[117] T. Treibitz and Y.Y. Schechner, "Active polarization descattering," *IEEE Trans Pattern Anal Mach Intell.*, 2009.

[118] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," *In Proc. of IEEE OCEANS*, 2010.

[119] J. Y. Chiang and Y. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *In IEEE Trans. on Image Processing*, 2012.

[120] P. Drews-Jr, E. Nascimento, F. Moraes, S. Botelho, M. Campos, and R. Grande-Brazil, "Transmission estimation in underwater single images," *IEEE Workshop ICCV*, 2013.

[121] A. Galdran, D. Pardo, A. Picon, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, 2015.

[122] J. N. Lythgoe and C. C. Hemmings, "Polarized light and underwater vision," *Nature*, vol. 213, no. 79, 1967.

[123] H. Kim, H. Jin, S. Hadap, and I. Kweon, "Specular reflection separation using dark channel prior," *In IEEE CVPR*, 2013.

[124] D. Zhengguo, "A new visibility metric for haze images," *http://uk.mathworks.com/matlabcentral/fileexchange/33529-a-new-visibility-metric-for-haze-images*.

[125] L.S. Wang, K. Ma, and H. Yeganeh, "A patch-structure representation method for quality assessment of contrast changed images," *IEEE Signal Processing Letters*, 2015.

[126] G. Wyszecki and W.S. Stiles, "Color science: Concepts and methods, quantitative data and formulae, 2nd ed.," *Wiley-Interscience*, 2000.

[127] 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.

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

[129] 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.

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

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

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

[133] C. Lu, L. Xu, and J. Jia, "Contrast preserving decolorization," *In IEEE International Conference on Computational Photography*, 2012.

[134] 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.

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

[136] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Decolorizing images for robust matching," in *In IEEE ICIP*, 2010.

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

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

[139] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, Allyn & Bacon, 5th edition, 2005.

[140] D. J. Heeger and J. R. Bergeny, "Pyramid-based texture analysis/synthesis," *ACM Transactions on Graphics (SIGGRAPH)*, 1995.

[141] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Transactions on Graphics (SIGGRAPH)*, 2004.

[142] Ramesh Raskar, Adrian Ilie, and Jingyi Yu, "Image fusion for context enhancement and video surrealism," *In NPAR*, 2004.

[143] S. Zheng, W.-Z. Shi, J. Liu, and J. Tian, "Remote sensing image fusion using multiscale mapped ls-svm," *IEEE Transactions on Geoscience and Remote Sensing*, 2008.

[144] F. Laliberte, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images-an evaluation study," *IEEE Trans. on Medical Imag.*, 2003.

[145] A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Information Fusion*, 2007.

[146] H. Yesou, Y. Besnus, and J. Rolet, "Extraction of spectral information from landsat tm data and merger with spot panchromatic imagery—a contribution to the study of geological structures," *ISPRS J. Photogramm. Remote Sens.*, 1993.

[147] H. Li, S. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, 1995.

[148] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," *IEEE Int. Conf. on Image Processing*, 1997.

[149] L. Tessens, A. Ledda, A. Pizurica, and W. Philips, "Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion," *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2007.

[150] M.N.Do and M.Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, 2005.

[151] Q. Zhang and B. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *IEEE Signal Process.*, 2009.

[152] J. Tang, "A contrast based image fusion technique in the dct domain," *Digital Signal Processing*, 2004.

[153] M. Xu, H. Chen, and P. K. Varshney, "An image fusion approach based on markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, 2011.

[154] D. Fay, A. Waxman, M. Aguilar, D. Ireland, J. Racamato, W. Ross, W. Streilein, and M. Braun, "Fusion of multi-sensor imagery for night vision," *Information Fusion*, 2000.

[155] V. Petrovic and C. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Transactions on Image Processing*, 2004.

[156] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Trans. on Image Proc.*, 2012.

[157] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Transactions on Image Processing*, 2011.

[158] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, 2013.

[159] K. He, J. Sun, and X. Tang, "Guided image filtering," *In Proc. European Conf. on Comp. Vision*, 2010.

[160] J. Lewis, R. OCallaghan, S. Nikolov, D. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, 2007.

[161] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: Edge-aware image processing with a laplacian pyramid," *ACM Transactions on Graphics (SIGGRAPH)*, 2011.

[162] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast local laplacian filters: Theory and applications," *ACM Transactions on Graphics (SIGGRAPH)*, 2014.

[163] B. Geelen, A. Ferentinos, F. Catthoor, G. Lafruit, D. Verkest, R. Lauwereins, and T. Stouraitis, "Modeling and exploiting spatial locality trade-offs in wavelet-based applications under varying resource requirements, acm transactions on embedded computing systems," *ACM Transactions on Embedded Computing Systems*, vol. 9, no. 3, 2010.

[164] Y. Andreopoulos, P. Schelkens G. Lafruit, K. Masselos, and J. Cornelis, "High-level cache modeling for 2-d discrete wavelet transform implementations," *International Journal of Computer Aided VLSI Design*, vol. 34, no. 3, 2003.

[165] J.F. Delaigle, C. De Vleeschouwer, B. Macq, and L. Langendijk, "Human visual system features enabling watermarking," *IEEE Int. Conf. on Multimedia and Expo*, 2002.

[166] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-rangeimages," *ACM Trans. Graph.*, 2002.

[167] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph.Forum*, 2003.

[168] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, 2002.

[169] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graph.*, 2005.

[170] Raanan Fattal, Dani Lischinski, and Michael Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, 2002.

[171] R. Mantiuk, K. Myszkowski, and H-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Perception*, 2006.

[172] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, 2008.

[173] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, 2013.

[174] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. Signals, Syst., Comput.*, 2003.

[175] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, 2011.

[176] Z. Wang and A. C. Bovik, "Reduced and no reference visual quality assessment- the natural scene statistic model approach," *IEEE Signal Process. Mag.*, 2011.

[177] E.H. Adelson, "Depth-of-focus imaging process method," *United States Patent 4,661,986*, 1985.

[178] A.P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information Fusion*, 2014.

[179] A.C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. of the IEEE*, 2013.

[180] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Transactions on Image Processing*, 2008.

[181] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," *In Asian Conference on Computer Vision (ACCV)*, 2010.

[182] Codruta O. Ancuti Cosmin Ancuti and Philippe Bekaert, "Single image restoration of outdoor scenes," *In Lecture Notes in Computer Science*, 2012.

[183] Cosmin Ancuti, Codruta O. Ancuti, and Philippe Bekaert, "An efficient two steps algorithm for wide baseline image matching," in *Journal of Visual Computer, (Computer Graphics International Conference)*, May 2009.

[184] Codruta O. Ancuti, Cosmin Ancuti, and Philippe Bekaert, "Decolorizing images for robust matching," in *Proceedings of 17th IEEE International Conference on Image Processing (IEEE ICIP)*, 2010.

[185] Codruta O. Ancuti, Cosmin Ancuti, and Philippe Bekaert, "Cc-sift: Exploiting chromatic contrast for wide-baseline matching," in *Proceedings of 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP)*, 2010.

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

[187] Cosmin Ancuti, Codruta O. Ancuti, and Philippe Bekaert, "Image deblurring by corresponding regions," in *Conference Abstracts and Applications ACM SIGGRAPH*, August 11-15 2008.

[188] Cosmin Ancuti, Codruta O. Ancuti, and Philippe Bekaert, "Deblurring by matching," in *Computer Graphics Forum, (**Eurographics**)*, 2009.