Single-Scale Fusion: An Effective Approach to Merging Images

Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*— Due to its robustness and effectiveness, multi-scale fusion (MSF) based on the Laplacian pyramid decomposition has emerged as a popular technique that has shown utility in many applications. Guided by several intuitive measures (weight maps) the MSF process is versatile and straightforward to be implemented. However, the number of pyramid levels increases with the image size, which implies sophisticated data management and memory accesses, as well as additional computations. Here, we introduce a simplified formulation that reduces MSF to only a single level process. Starting from the MSF decomposition, we explain both mathematically and intuitively (visually) a way to simplify the classical MSF approach with minimal loss of information. The resulting single-scale fusion (SSF) solution is a close approximation of the MSF process that eliminates important redundant computations. It also provides insights regarding why MSF is so effective. While our simplified expression is derived in the context of high dynamic range imaging, we show its generality on several well-known fusion-based applications, such as image compositing, extended depth of field, medical imaging, and blending thermal (infrared) images with visible light. Besides visual validation, quantitative evaluations demonstrate that our SSF strategy is able to yield results that are highly competitive with traditional MSF approaches.**

***Index Terms*— Multi-scale image fusion, Laplacian pyramid, image enhancement.**

I. Introduction

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 [1], texture synthesis [2], image editing [3], image compression [4], multi-sensor photography [5], context enhancement and surrealist video processing [6], image compositing [7], enhancing under-exposed videos [8], multi-spectral remote sensing [9], medical imaging [10].

Many different strategies to fuse a set of images have been introduced in the literature [11]. The simplest methods, including averaging and principal component analysis (PCA) [12], 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. [13] 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 [14] proposed using a shift

Manuscript received February 25, 2016; revised July 3, 2016 and September 13, 2016; accepted October 12, 2016. Date of publication October 26, 2016; date of current version November 15, 2016. This work was supported in part by Belgian NSF, in part by a MOVE-IN Louvain, Marie Curie Fellowship, in part by the Marie Curie Actions of EU-FP7 under REA Grant TECNIOspring Programme under Grant 600388, and in part by the Agency for Business Competitiveness of the Government of Catalonia under Grant ACCIO: TECSPR14-2-0023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Martin Kleinsteuber.

C. O. Ancuti is with the Computer Vision and Robotics Group, University of Girona, 17071 Girona, Spain, and also with MEO, Universitatea Politehnica Timișoara, 300723 Timișoara, Romania. (e-mail: codruta.ancuti@gmail.com).

C. Ancuti is with ICTEAM, Universite Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium, and also with MEO, Universitatea Politehnica Timișoara, 300006 Timișoara, Romania.

C. De Vleeschouwer is with ICTEAM, Universite Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium.

A. C. Bovik is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2621674

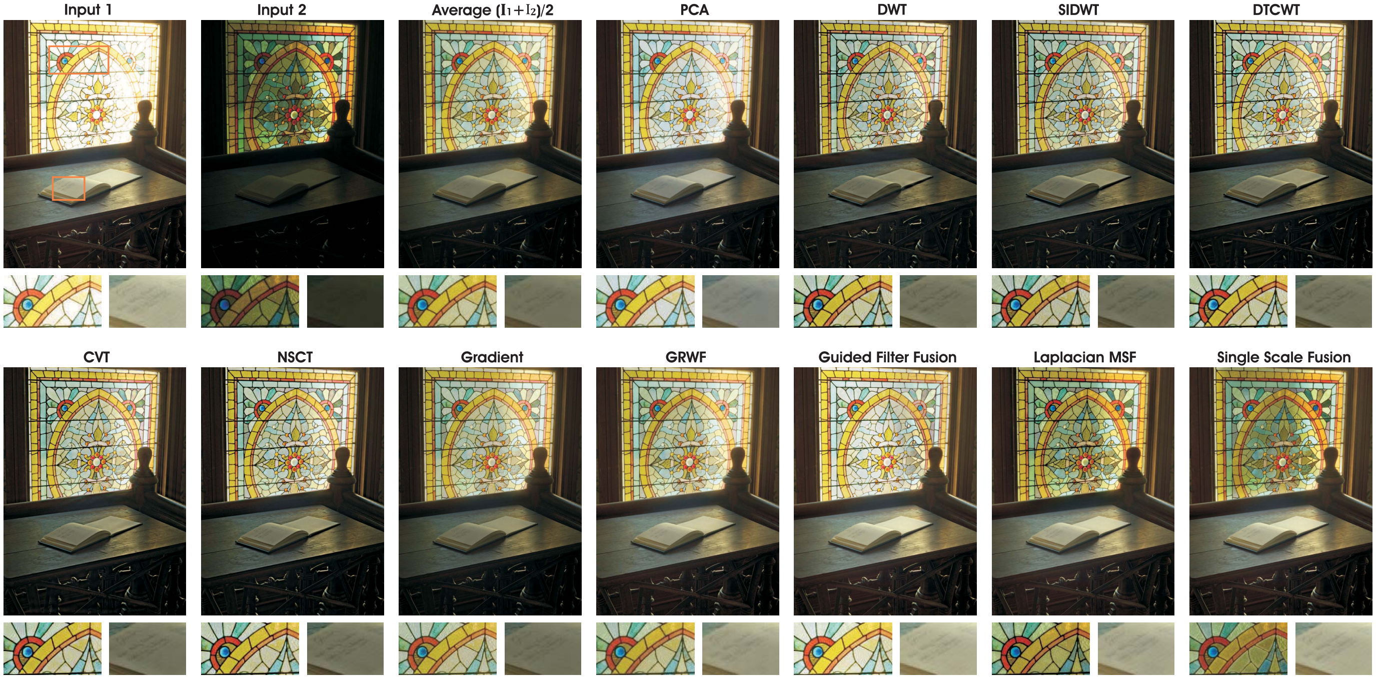


Fig. 1. Comparative results of different fusion techniques for merging multi-exposure images. Top row (from left to right), are shown the original two inputs, the result of averaging the inputs and the fusion results of [12]–[14] and [26]. Bottom row (from left to right), are shown the results of the fusion approaches of [4], [15], [17], [21], [23], and [24], 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 [4], [27] is more robust and delivers results comparable to those of more recent fusion techniques such as GRWF fusion method of Shen et al. [23].

invariant wavelet decomposition. Tessens et al. [15] 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 [16], which combines the Laplacian pyramid with a directional filter bank. Zhang and Guo [17] deployed an undecimated, shift-invariant contourlet transform for image fusion. In another category of methods, Tang [18] 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. [19]. Neural networks were employed by Fay et al. [20] to fuse night-vision images from multiple infra-red bands. A gradient-based method was introduced by Petrovic and Xydeas [21]. In their method, input images were represented at each resolution level using gradient map signals rather than absolute grey-level values. Liang et al. [22] formulated a tensor decomposition technique and used SVD to fuse multiple images. More recently, in the context of multi-exposure fusion, Shen et al. [23] 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. [24] proposed an effective framework built on guided filters [25] 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. 1). Introduced by Burt and Adelson [4] 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 [27], image filtering [28], [29], single image dehazing [30], [31], image and video decolorization [32] and underwater image enhancement [33].

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 [34], [35]. 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. 9).

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. 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 paper 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;

The rest of the paper is organized as follows. Section II briefly reviews the process of multi-scale fusion (MSF) based on the Laplacian pyramid decomposition and also discusses limitations of the naive fusion process. Section III introduces approximations that are performed to derive a single-scale fusion (SSF) algorithm from the MSF formulation. Section IV presents both the qualitative and quantitative performance of our SSF operator for a large range of applications, while Section V concludes the paper.

II. Image Fusion: Background and Notations

Generally, image fusion can be defined as a process of effectively blending several input images (e.g. [4], [27]) or versions of the same original image (e.g. [30], [33]) 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 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.

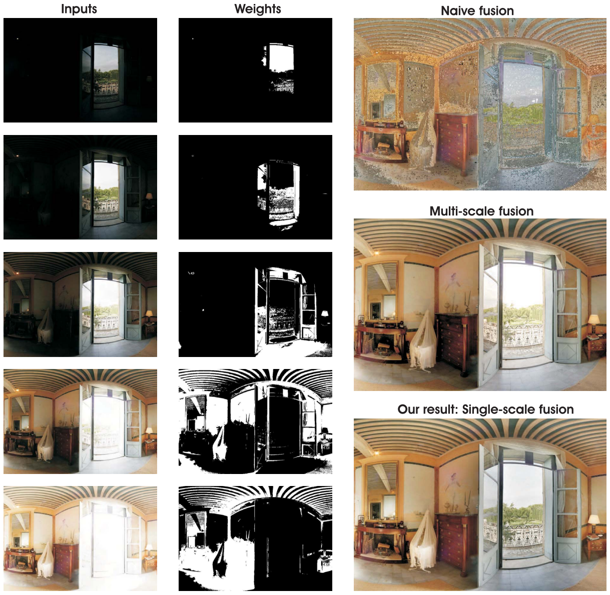


Fig. 2. Naive, multi-scale fusion and our single-scale result. Both involve, a similar degree of complexity, while our single-scale fusion method is able to deliver results competitive with the multi-scale approach.

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.

A. Naive Image Fusion

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.[1] This approach, called naive image fusion (NF), is quite straightforward and computationally efficient. The naive fusion result $\mathcal{R}_{NF}$ can be expressed as:

$$\mathcal{R}_{NF}(x) = \sum_k^K \bar{\mathcal{W}}_k(x)\mathcal{I}_k(x) \tag{1}$$

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. 2), especially in the regions containing strong transitions in the weight maps that have no correspondence with the input content. As pointed out in [27] and further discussed in Section III (and depicted in Fig. 4),

[1] *See the experimental section for a description of the weight map is computed in practice.*

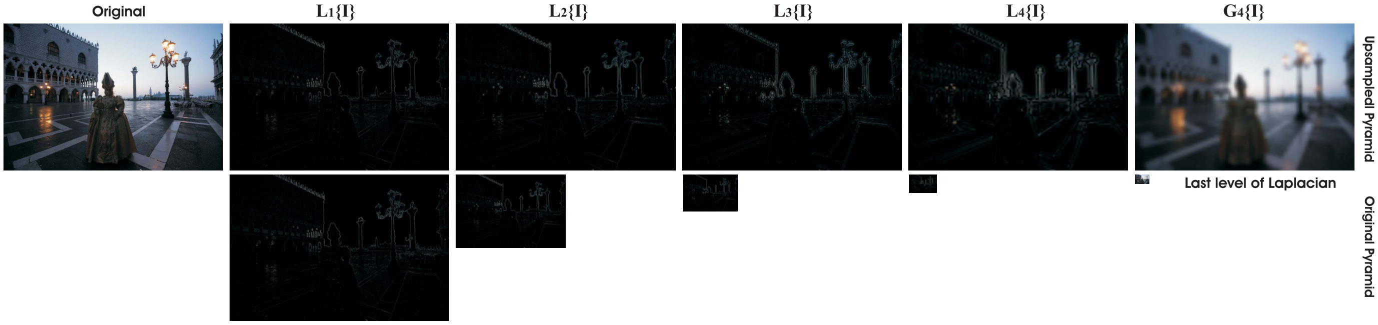


Fig. 3. The final level of a Laplacian pyramid includes a Gaussian blurred version of the original image. The top row depicts the upsampled versions of the five downsampled Laplacian and Gaussian signals shown in the bottom row. Note also that for better visualization the absolute value of each Laplacian image pixel is presented. This is to render the small Laplacian intensities in black and their large values in white, whatever their sign.

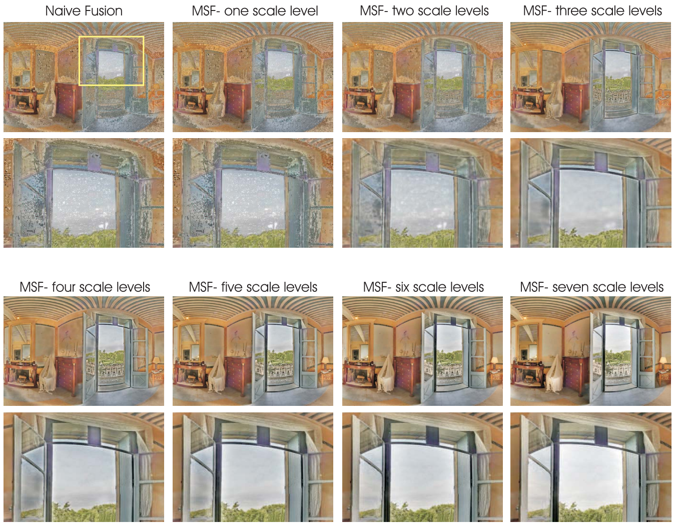


Fig. 4. Illustration of the influence of the number of levels in the multi-scale fusion approach. As can be observed, the number of levels affects the degree to which the higher image frequencies are revealed. Reducing the number of levels causes high frequency artifacts similarly to the naive fusion approach.

simple low-pass filtering of the weight maps is insufficient to remove those artifacts.

B. Multi-Scale Fusion

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 [4]. 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:

$$\begin{aligned}\mathcal{I}(x) &= \mathcal{I}(x) - G_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\}\\ &\triangleq L_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\}\\ &= L_1\{\mathcal{I}(x)\} + G_1\{\mathcal{I}(x)\} - G_2\{\mathcal{I}(x)\} + G_2\{\mathcal{I}(x)\}\\ &= L_1\{\mathcal{I}(x)\} + L_2\{\mathcal{I}(x)\} + G_2\{\mathcal{I}(x)\}\\ &= \ldots\\ &= \sum_{l=1}^{N} L_l\{\mathcal{I}(x)\} + G_N\{\mathcal{I}(x)\} \qquad (2)\end{aligned}$$

As a result, the last component of the decomposition in (2), is a Gaussian blurred version of the input image with a large kernel (see Fig. 5). 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 the paper all those images have been up-sampled to the original image dimension.

In the traditional multi-scale fusion (MSF) strategy [27], each source input $\mathcal{I}_k$, is decomposed into a Laplacian pyramid [4] 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[2]:

$$\mathcal{R}_l(x) = \sum_{k=1}^{K} G_{l-1}\left\{\bar{\mathcal{W}}_k(x)\right\} L_l\{\mathcal{I}_k(x)\} \qquad (3)$$

where $0 < l \leq N$ denotes the pyramid levels and k refers to the number of input images. The last component in (2) induces a last contribution $\mathcal{R}_{N+1} = \sum_k G_N\{\mathcal{W}_k\} G_N\{\mathcal{I}_k\}$. 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 (1). 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. 4).

This blending step is performed successively at each pyramid layer, in a bottom-up manner. The final multi-scale

[2] *An efficient multi-scale fusion (MSF) implementation manipulates downsampled signals at each level of resolution to save memory and computation, and upsamples the outcome of each level only before the aggregation procedure*

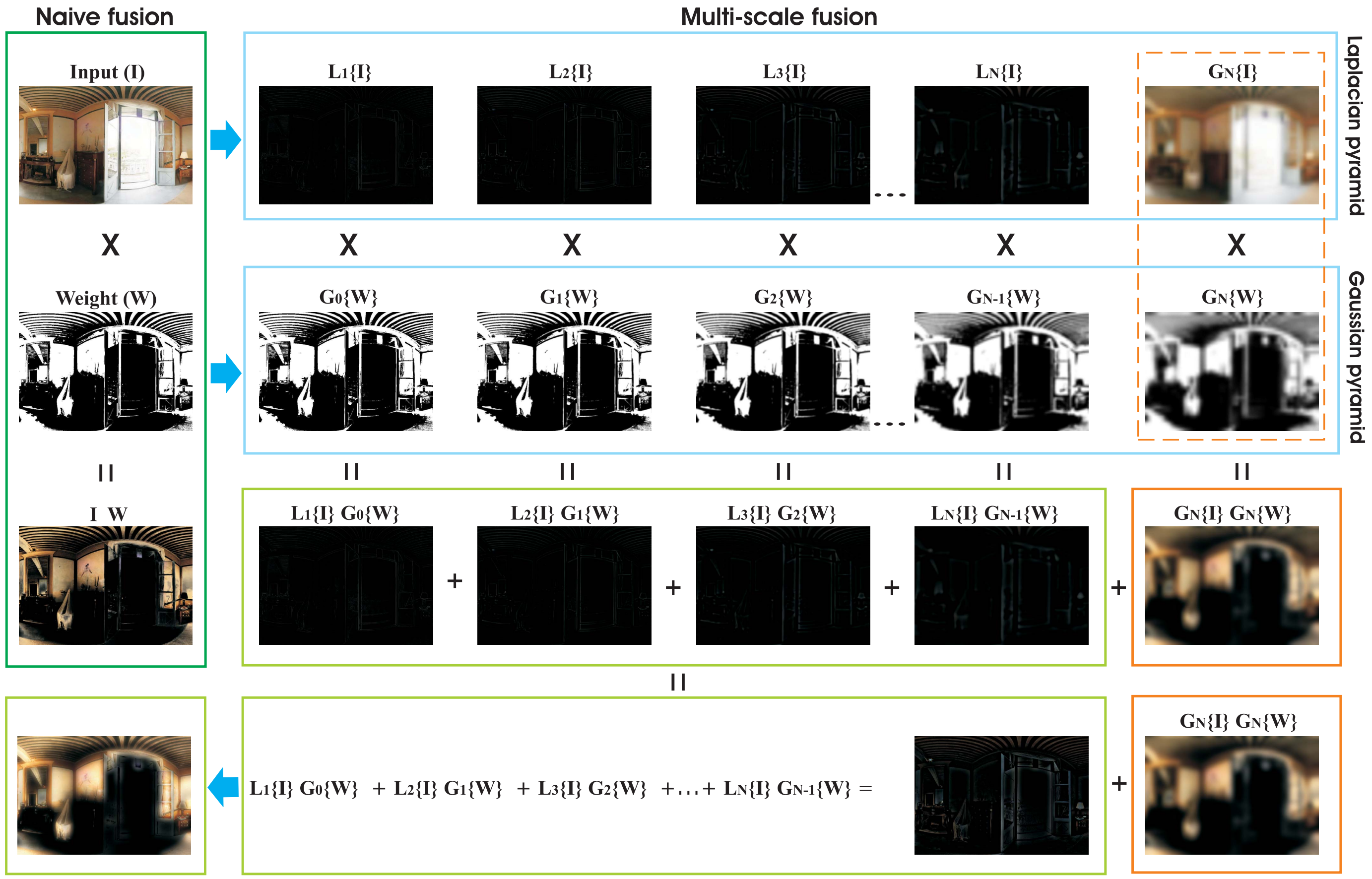


Fig. 5. Fusion pipeline. The Laplacian and Gaussian versions have been upsampled to the original size of the image.

fused result $\mathcal{R}_{MSF}$ is obtained by simply summing up the contribution from each level:

$$\mathcal{R}_{MSF}(x) = \sum_{l}^{N+1} \mathcal{R}_l(x) \tag{4}$$

III. 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 [27]. However, as will be illustrated in the Section IV, our approach is general, being suited to other scenarios and applications that are built on the multi-scale fusion process.

As discussed in Section II, 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):

$$\mathcal{R} = \sum_{l=1}^{N} G_{l-1}\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\} + G_N\left\{\bar{\mathcal{W}}\right\} G_N\left\{\mathcal{I}\right\} \tag{5}$$

To derive a single scale approximation of (5), 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. 5). 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 (5). Then (5) becomes:

$$\mathcal{R} = \sum_{l=1}^{N} G_N\left\{\bar{\mathcal{W}}\right\} L_l\left\{\mathcal{I}\right\} + G_N\left\{\bar{\mathcal{W}}\right\} G_N\left\{\mathcal{I}\right\} \tag{6}$$

In this equation, $G_N\left\{\bar{\mathcal{W}}\right\}$ can be put in evidence and the sum of $G_N\left\{\bar{\mathcal{I}}\right\}$ 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 7 and 9 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.* [27]). As may be seen in Fig. 7 , this observation from [27] 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 (6) also indirectly explains why

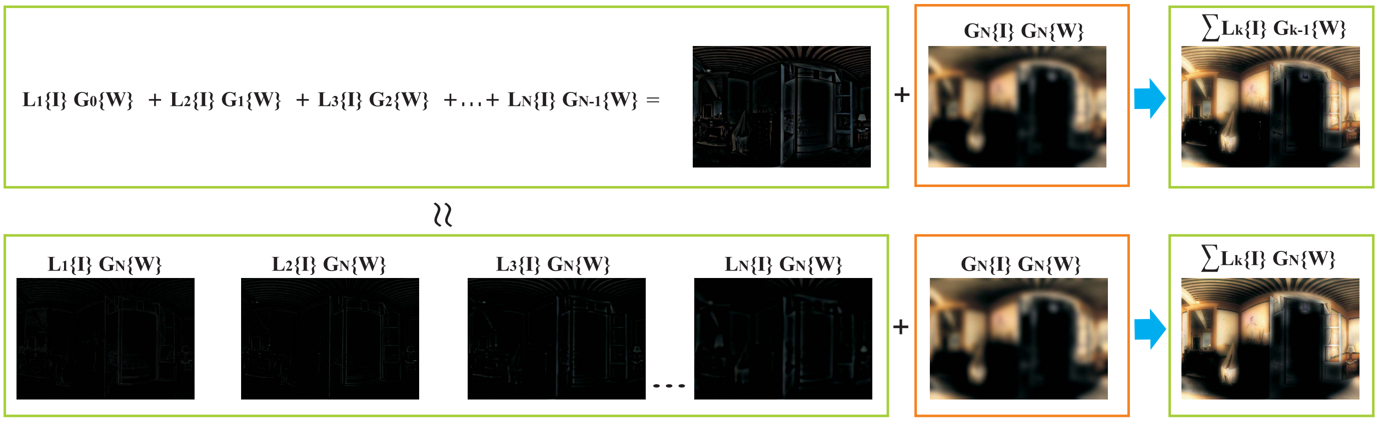


Fig. 6. As a first order MSF approximation, we had envisioned approximating the expression $L_1\{\mathcal{I}\}G_0\{\mathcal{W}\} + L_2\{\mathcal{I}\}G_1\{\mathcal{W}\} + \cdots + L_N\{\mathcal{I}\}G_{N-1}\{\mathcal{W}\}$ by $L_1\{\mathcal{I}\}G_N\{\mathcal{W}\} + L_2\{\mathcal{I}\}G_N\{\mathcal{W}\} + \cdots + L_N\{\mathcal{I}\}G_N\{\mathcal{W}\}$, thereby turning the MSF into a SSF. However, as illustrated in Fig.7, this approximation is not satisfying, thereby motivating our refined approximation derived from (7) to (15).

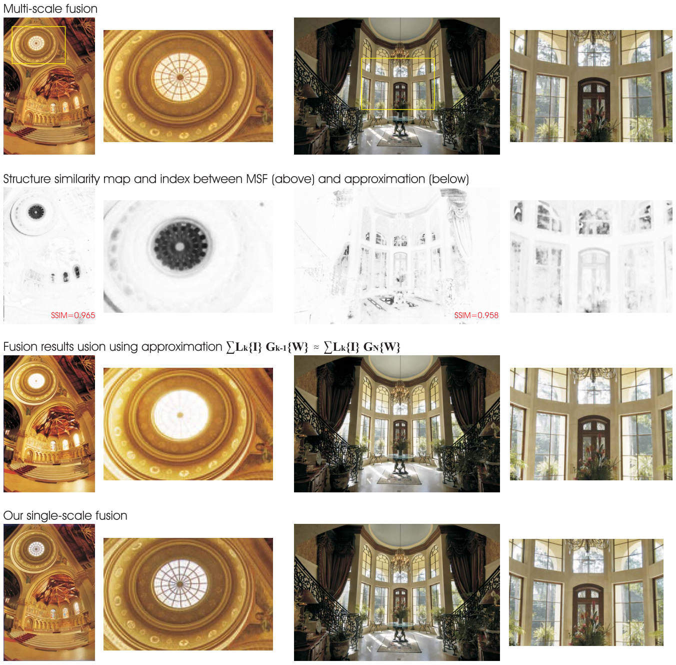

Multi-scale fusion
Structure similarity map and index between MSF (above) and approximation (below)
SSIM=0.965
SSIM=0.958
Fusion results usion using approximation ∑Lk{I} Gk-1{W} ≈ ∑Lk{I} GN{W}
Our single-scale fusion

Fig. 7. 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.

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 [36], 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.

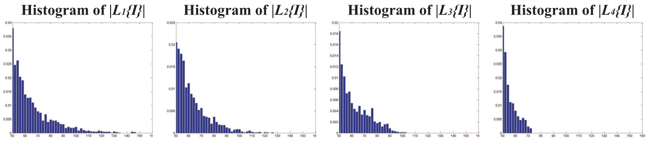


Fig. 8. Histograms of $L_1\{\mathcal{I}\}$, $L_2\{\mathcal{I}\}$, $L_3\{\mathcal{I}\}$ and $L_4\{\mathcal{I}\}$. As expected, the histograms reveal that the largest values of $L_1\{\mathcal{I}\}$ are bigger than the largest values of $L_k\{\mathcal{I}\}$, with $k > 1$. This supports the approximation made to go from (8) to (9).

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. 4, we indeed observe that 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 (6) lack of details as compared to the MSF results (see Fig. 9 and Fig. 7). Therefore, we propose a second order approximation of (5), which aims to preserve details in the inputs, while remaining single scale. Formally, given that by definition of the Laplacian pyramid (Section III) $G_{l-1}\{\bar{\mathcal{W}}\} = G_N\{\bar{\mathcal{W}}\} + \sum_{p=l}^{N} L_p\{\bar{\mathcal{W}}\}$, then (5) becomes:

$$\mathcal{R} = \sum_{l=1}^{N}\left[\sum_{p=l}^{N} L_p\{\bar{\mathcal{W}}\} L_l\{\mathcal{I}\}\right] + G_N\{\bar{\mathcal{W}}\}\sum_{l=1}^{N} L_l\{\mathcal{I}\} + G_N\{\bar{\mathcal{W}}\}G_N\{\mathcal{I}\} \qquad (7)$$

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

$$\mathcal{R} = \sum_{l=1}^{N}\left[\sum_{p=l}^{N} L_p\{\bar{\mathcal{W}}\} L_l\{\mathcal{I}\}\right] + G_N\{\bar{\mathcal{W}}\}\mathcal{I} \qquad (8)$$

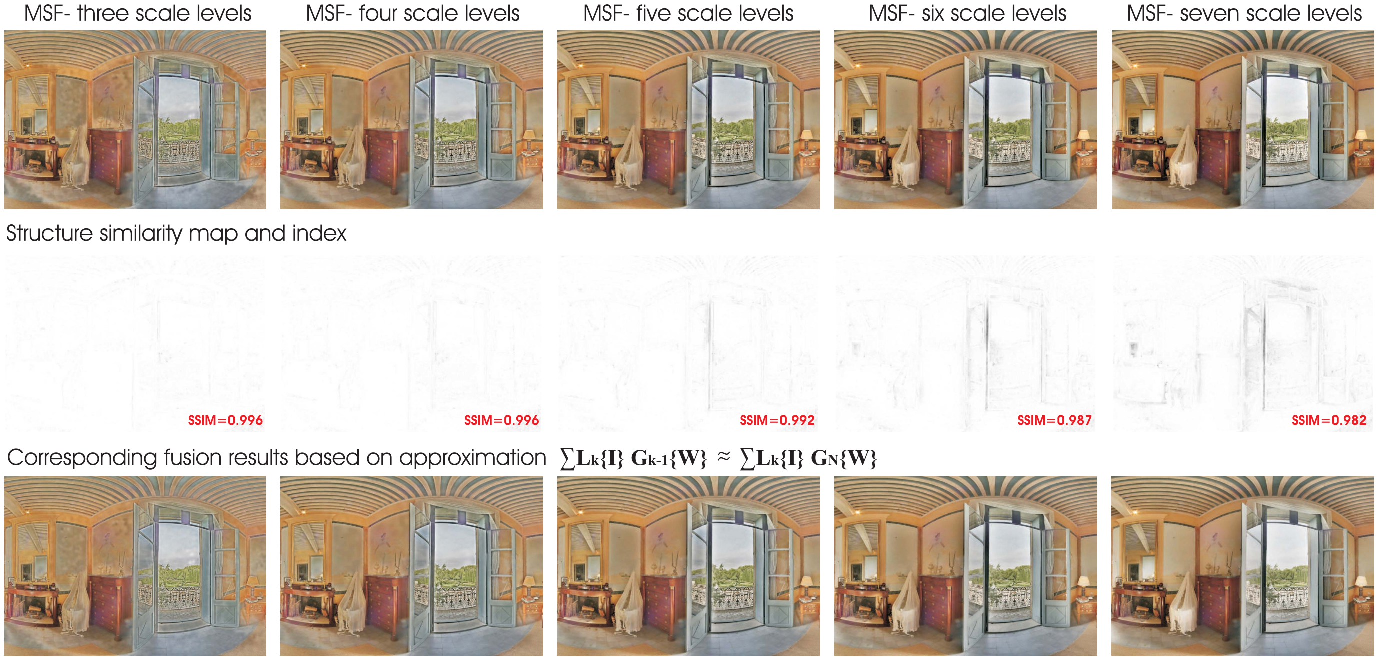


Fig. 9. The impact of the number of scales when approximating in (5) $L_1\{\mathcal{I}\}\, G_0\{\mathcal{W}\} + L_2\{\mathcal{I}\}\, G_1\{\mathcal{W}\} + \ldots + L_N\{\mathcal{I}\}\, G_{N-1}\{\mathcal{W}\}$ by $L_1\{\mathcal{I}\}\, G_N\{\mathcal{W}\} + L_2\{\mathcal{I}\}\, G_N\{\mathcal{W}\} + \ldots + L_N\{\mathcal{I}\}\, G_N\{\mathcal{W}\}$. 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 $G_N\{\mathcal{W}\}$.

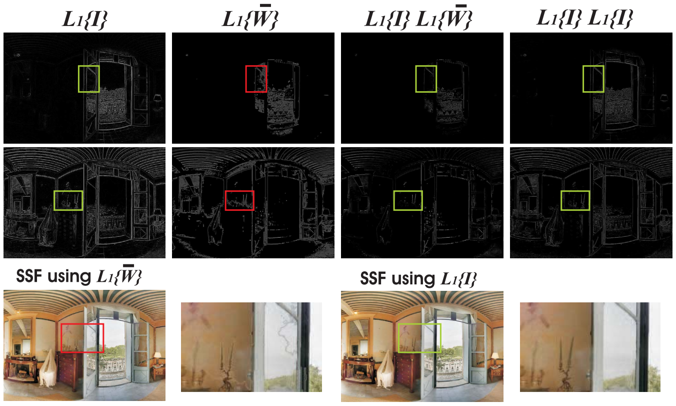


Fig. 10. Top two rows show $L_1\{\mathcal{I}\}$, $L_1\{\bar{\mathcal{W}}\}$, $L_1\{\mathcal{I}\}.L_1\{\bar{\mathcal{W}}\}$ and $L_1\{\mathcal{I}\}.L_1\{\mathcal{I}\}$ of two inputs shown in Fig.2. *$L_1\{\bar{\mathcal{W}}\}$ has a shape that is similar to that of $L_1\{\mathcal{I}\}$, in locations where $L_1\{\mathcal{I}\}$ is large, and therefore we can simplify (9) to (10). Spurious edges could be transferred to the output if $L_1\{\mathcal{I}\}$ was approximated with $L_1\{\bar{\mathcal{W}}\}$ in (9) (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.*

We can assume that the first term of the sum (l=1) dominates the others since the largest values, in $L_1\{\mathcal{I}\}$ tend to be much larger than the largest values in $L_k\{\mathcal{I}\}$, with $k > 1$ (see their histograms in Fig. 8). Here, we focus only on the largest values of $L_k\{\mathcal{I}\}$ because they are the only ones that matter when the products $L_k\{\mathcal{I}\}\, L_p\{\bar{\mathcal{W}}\}$ are added to $G_N\{\bar{\mathcal{W}}\}\mathcal{I}$.

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

$$\mathcal{R} \approx G_N\{\bar{\mathcal{W}}\}\mathcal{I} + L_1\{\bar{\mathcal{W}}\}\, L_1\{\mathcal{I}\} \tag{9}$$

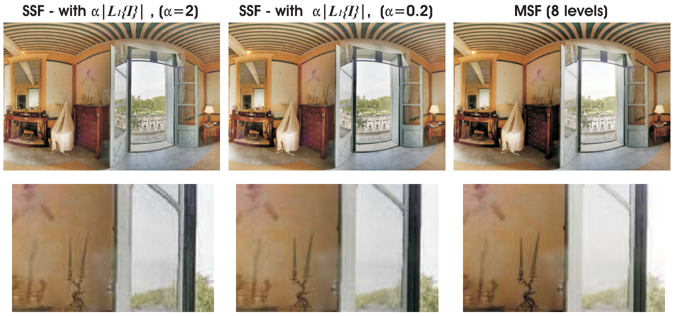

SSF - with $\alpha|L_1\{I\}|$, (α=2)
SSF - with $\alpha|L_1\{I\}|$, (α=0.2)
MSF (8 levels)

Fig. 11. The influence of the parameter α used in our single-scale fusion method (15). If we increase the impact ($\alpha = 2$) of $L_1\{\mathcal{I}\}$, some small artifacts appear around edges. Experimentally, we found that a default value of $\alpha = 0.2$ is well suited to all investigated scenarios.

By observing that $L_1\{\bar{\mathcal{W}}\}$ 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.10):

$$\mathcal{R} \approx G_N\{\bar{\mathcal{W}}\}\mathcal{I} + \beta L_1\{\mathcal{I}\}\, L_1\{\mathcal{I}\} \tag{10}$$

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.[3] The relevance of approximation is confirmed through extensive

[3] *This assumption is certainly not strictly valid. It does however, support the developments (9) to (15), and leads to the approximation proposed in (15)*

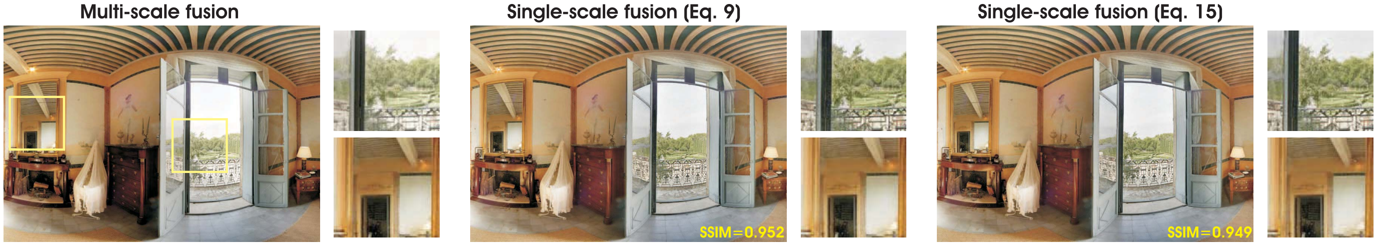


Fig. 12. Single-scale fusion results generated using (9) and (15). As may be observed, visually and also based on the SSIM evaluation [37] our approach yield very similar results as the traditional MSF approach. Since both single-scale fusion (9) and (15) produce almost identical results, all the results have been generated using the last derivation described in (15) using the default parameter $\alpha = 0.2$.

simulations in Section IV, 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:

$$\begin{aligned} L_1\{\mathcal{I}(x)\} &= \mathcal{I}(x) - G_1\{\mathcal{I}(x)\} \\ &\approx (x - x_0)\left[\nabla\mathcal{I}(x_0) - \nabla G_1\{\mathcal{I}(x_0)\}\right] \end{aligned} \tag{11}$$

and

$$\mathcal{I}(x) \approx \mathcal{I}(x_0) + (x - x_0)\nabla\mathcal{I}(x_0) \tag{12}$$

By merging (11) and (12), we have:

$$L_1\{\mathcal{I}(x)\} \approx [\mathcal{I}(x) - \mathcal{I}(x_0)]\,\frac{[\nabla\mathcal{I}(x_0) - \nabla G_1\{\mathcal{I}(x_0)\}]}{\nabla\mathcal{I}(x_0)} \tag{13}$$

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 (10) as:

$$\begin{aligned} \beta L_1\{\mathcal{I}(x)\}\, L_1\{\mathcal{I}(x)\} &\approx \beta\gamma(x_0) L_1\{\mathcal{I}(x)\}\left[1 - \frac{\mathcal{I}(x_0)}{\mathcal{I}(x)}\right]\mathcal{I}(x) \\ &\approx \alpha\,|L_1\{\mathcal{I}(x)\}|\,\mathcal{I}(x) \end{aligned} \tag{14}$$

The first approximation is obtained by replacing $L_1\{\mathcal{I}\}$ in (10) using the approximation derived in (13). 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 Section IV.

Replacing the second term in (10) with the approximation (14), 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:

$$\mathcal{R}_{SSF,k}(x) = \left[G_N\left\{\bar{\mathcal{W}}(x)\right\} + \alpha\,|L_1\{\mathcal{I}(x)\}|\right]\mathcal{I}(x) \tag{15}$$

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 $\mathcal{R}_{SSF}(x) = \sum_k \mathcal{R}_{SSF,k}(x)$.

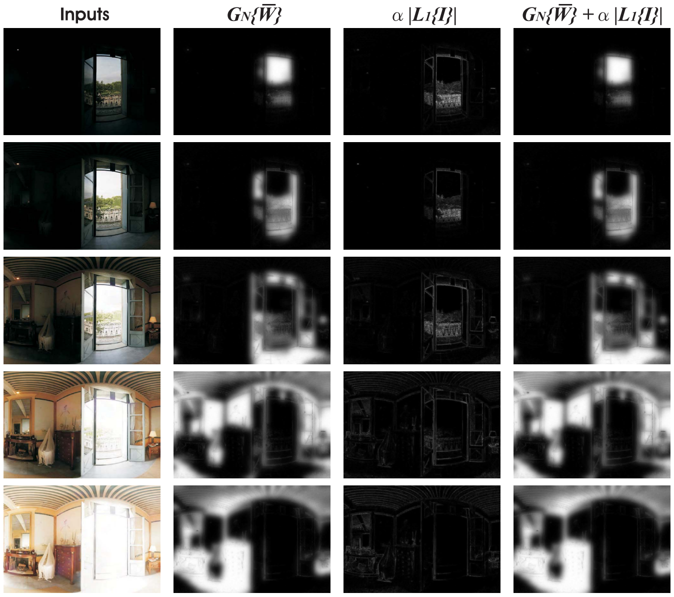


Fig. 13. The weights of our final SSF expression (15). From left to right: the inputs, the corresponding two weights ($G_N\left\{\bar{\mathcal{W}}\right\}$ and $\alpha\,|L_1\{\mathcal{I}\}|$) of (15) and their sum.

Fig. 12 depicts differences between results obtained using our single-scale fusion (9) and (15) and the MSF expression. As can be observed visually and also based on the SSIM evaluation [37], both (9) and (15) yields very similar results compared as traditional MSF. While both SSF equations (9) and (15) produce almost identical outputs, all the results presented in the remaining of the paper have been generated using the last derivation described in (15) using the default parameter $\alpha = 0.2$. Interestingly, (15) 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.

IV. Results and Discussion

Since the primary contribution of our paper 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 [27], [30], [31], [33]: 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 [27] and [30] 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 [27], [33] 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 $\mathcal{I}_k$:

$$\mathcal{W}_{k,C} = \sqrt{1/3\left[(R_k - L_k)^2 + (G_k - L_k)^2 + (B_k - L_k)^2\right]} \tag{16}$$

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 [27] and [31]. This measure for input $\mathcal{I}_k$ is expressed as a Gaussian-modeled distance to the average normalized value (0.5):

$$\mathcal{W}_{k,E} = exp\left(-\frac{(\mathcal{I}_k - 0.5)^2}{2\sigma^2}\right) \tag{17}$$

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 [30] and [31] we employ the well-known saliency technique of Achanta et al. [45]. 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:

$$\mathcal{W}_{k,S} = \left\|\mathcal{I}_{k,\omega_{hc}} - \mathcal{I}_{k,\mu}\right\| \tag{18}$$

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.

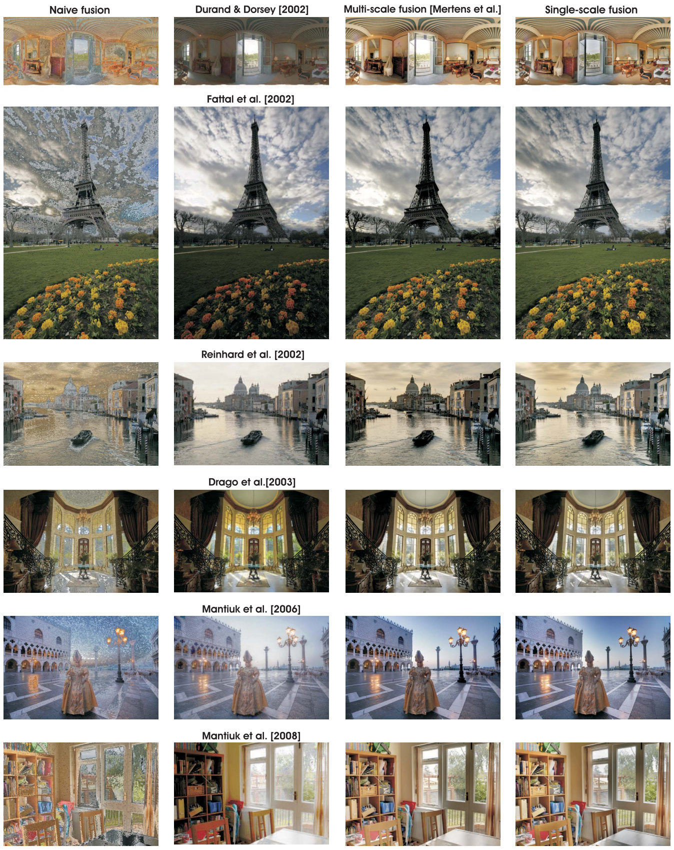


Fig. 14. HDR imaging. Comparison of single-scale fusion (SSF) with multi-scale fusion (MSF) for HDR imaging. Also shown are comparative results with several well known tone mapping techniques [38]–[44].

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.

A. High Dynamic Range Imaging

Various tone mapping techniques [38]–[44] 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 [27] 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. [27], which extends the original MSF approach of Burt and Adelson [4]. For a fair evaluation we use the same weight maps in the process of fusing the multiple exposure images.

Figure 14 shows comparative results between SSF and MSF (the exposure fusion approach of [27]) and also the results

TABLE I

QUALITATIVE COMPARISON BETWEEN MSF AND OUR SSF APPROACH BASED ON THE TONE MAPPING METRIC QUALITY INDEX TMQI [46]. 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

	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

generated by several tone mapping techniques [38]–[44] that have been generated by using the publicly available software *Luminance HDR*.[4]

Qualitative visual evaluation of Fig. 14 and Fig. 15 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 [46].

TMQI uses the well known structural similarity (SSIM) index [37], [47], [48] along with a natural scene statistics (NSS) model [49]. TMQI evaluates the quality of the resulted LDR images using the HDR image as a reference. It combines the multi-scale SSIM [47] 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. 15 while the other three are included in Fig. 14). Table I 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 I.

B. 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 [50], we also compute the well-known structural similarity (SSIM) index [37], [47], [48] 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.

C. 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 [51]. 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.

Multi-scale fusion has been successfully applied for this task [7]. As shown in the examples in Fig. 16, our technique performs on a par with the specialized multi-scale fusion approach of [7] 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. 17 for another example).

D. 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 [1], where the depth of field may be extremely limited. This task was performed automatically over the entire image for the first time in [4] using a multi-scale fusion strategy based on Laplacian pyramids. Figure 18 demonstrates that SSF is able to produce comparable results as traditional MSF techniques (using the same weight maps).

E. 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 [10], [52] is a common strategy that yields a more accurate description of the scanned body, since information provided by these

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

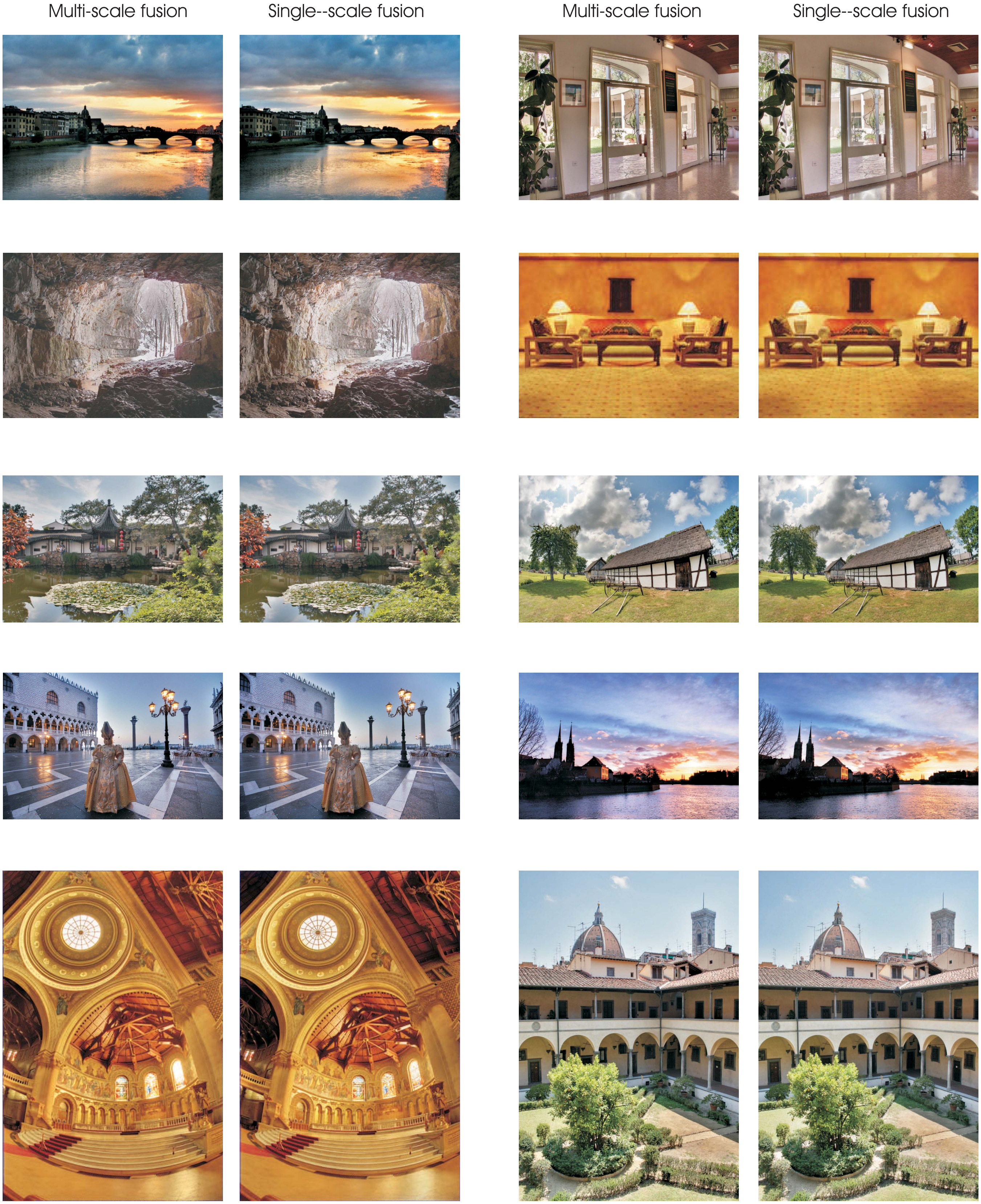


Fig. 15. HDR imaging. Comparison of the MSF with SSF for several set of images used in the TMQI evaluation shown in Table I. From left to right and top to bottom (**arno, belgium house, cave, chairs, chinese garden, kluki, mask, ostrow, memorial, laurentian library**).

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 [52], Fig. 19 demonstrates that our simplified approach is able to yield results that

TABLE II

COMPARATIVE COMPUTATION TIMES (EXPRESSED IN SECONDS OF MATLAB CODE) OF NF, MSF AND SSF STRATEGIES FOR DIFFERENT INPUT SIZES. AS SHOWN IN SECTION III, 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

	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

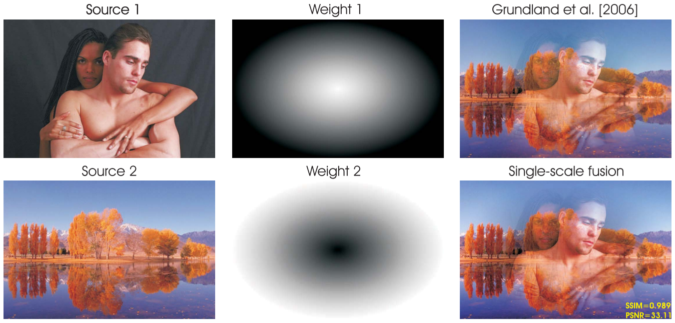


Fig. 16. Image Compositing.

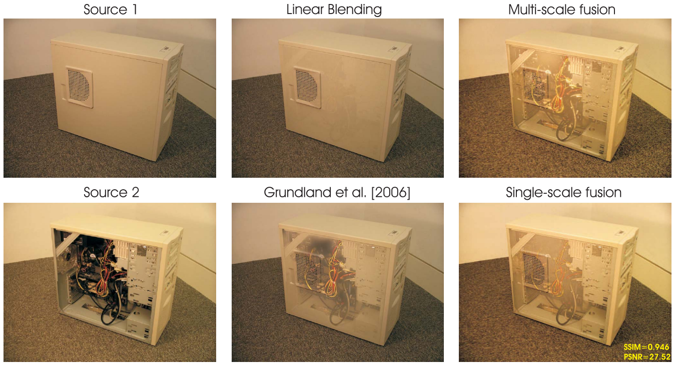


Fig. 17. Image Compositing.

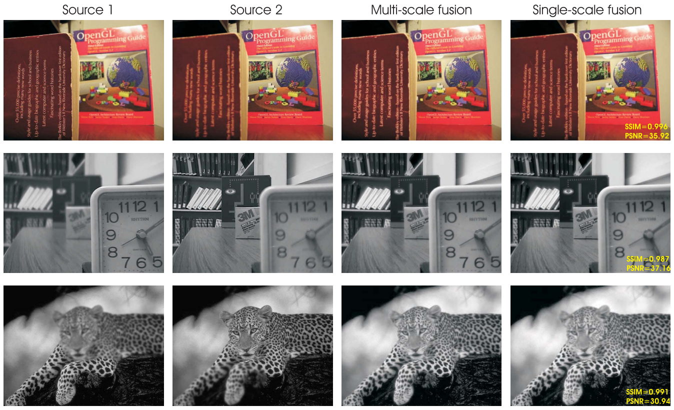


Fig. 18. Extended Depth of Field.

preserve the contrast and finest details in a manner similar to the classical MSF approach (observe also the overlaid SSIM index values).

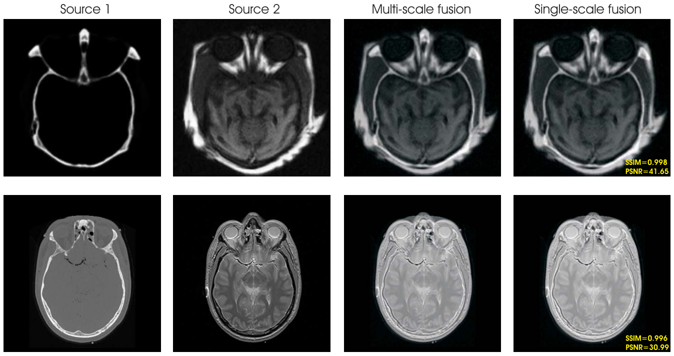


Fig. 19. Medical imaging: fusing MRI/CT images.

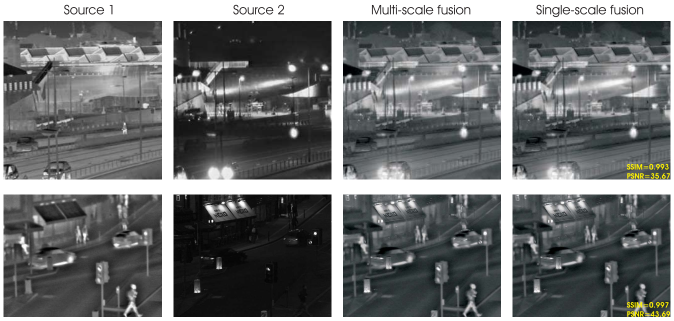


Fig. 20. Fusing visible and infra-red information.

F. 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 [5], [11]. 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 20 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.

While the computation complexity of our SSF technique is similar to that of the naive fusion implementation (please refer to the Table II), 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. 2). 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 [34], [35], 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 II 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.

V. Conclusions

In this paper 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 [53] to perceptually optimize the fusion process [54].

References

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

[2] D. J. Heeger and J. R. Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. SIGGRAPH*, 1995, pp. 229–238.

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

[4] P. Burt and T. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[5] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph. (SIGGRAPH)*, vol. 23, no. 3, pp. 664–672, 2004.

[6] R. Raskar, A. Ilie, and J. Yu, "Image fusion for context enhancement and video surrealism," in *Proc. NPAR*, 2004, pp. 85–152.

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

[8] E. P. Bennett, J. L. Mason, and L. McMillan, "Multispectral bilateral video fusion," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1185–1194, May 2007.

[9] S. Zheng, W.-Z. Shi, J. Liu, and J. Tian, "Remote sensing image fusion using multiscale mapped LS-SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1313–1322, May 2008.

[10] F. Laliberte, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images-an evaluation study," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 661–673, May 2003.

[11] A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion*, vol. 8, no. 2, pp. 114–118, 2007.

[12] H. Yésou, 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. Photogram. Remote Sens.*, vol. 48, no. 5, pp. 23–36, 1993.

[13] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.

[14] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Oct. 1997, pp. 288–291.

[15] L. Tessens, A. Ledda, A. Pizurica, and W. Philips, "Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. I-861–I-864.

[16] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[17] Q. Zhang and B.-L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, 2009.

[18] J. Tang, "A contrast based image fusion technique in the dct domain," *Digit. Signal Process.*, vol. 14, no. 3, pp. 218–226, 2004.

[19] M. Xu, H. Chen, and P. K. Varshney, "An image fusion approach based on Markov random fields," *IEEE Trans. Geosc. Remote Sens.*, vol. 49, no. 12, pp. 5116–5127, Dec. 2011.

[20] D. Fay *et al.*, "Fusion of multi-sensor imagery for night vision: Color visualization, target learning and search," in *Proc. Inf. Fusion*, vol. 1. Jul. 2000, pp. TUD3/3–TUD310.

[21] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.

[22] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2898–2909, May 2012.

[23] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec. 2011.

[24] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[25] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.

[26] J. J. Lewis, R. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Inf. Fusion*, vol. 8, no. 2, pp. 119–130, 2007.

[27] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[28] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: Edge-aware image processing with a laplacian pyramid," *ACM Trans. Graph. (SIGGRAPH)*, vol. 58, no. 3, pp. 81–91 2011.

[29] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast local laplacian filters: Theory and applications," *ACM Trans. Graph. (SIGGRAPH)*, vol. 33, no. 5, 2014, Art. no. 167.

[30] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.

[31] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[32] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "Image and video decolorization by fusion," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 79–92.

[33] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2012, pp. 81–88.

[34] B. Geelen *et al.*, "Modeling and exploiting spatial locality trade-offs in wavelet-based applications under varying resource requirements," *ACM Trans. Embedded Comput. Syst.*, vol. 9, no. 3, 2010, Art. no. 17.

[35] Y. Andreopoulos, P. S. G. Lafruit, K. Masselos, and J. Cornelis, "High-level cache modeling for 2-D discrete wavelet transform implementations," *J. VLSI Signal Process. Syst. Signal, Image Video Technol.*, vol. 34, no. 3, pp. 209–226, 2003.

[36] J. Delaigle, C. D. Vleeschouwer, B. Macq, and L. Langendijk, "Human visual system features enabling watermarking," in *Proc. IEEE Int. Conf. Multimedia Expo*, vol. 2. Aug. 2002, pp. 489–492.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[38] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, 2002.

[39] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 419–426, 2003.

[40] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, 2002.

[41] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 1, pp. 13–24, Jan./Feb. 2005.

[42] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 249–256, 2002.

[43] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Perception*, vol. 3, no. 3, pp. 286–308, 2006.

[44] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, 2008, Art. no. 68.

[45] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, Jun. 2009, pp. 1597–1604.

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

[47] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.

[48] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.

[49] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[50] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. San Mateo, CA, USA: Morgan Claypool, 2006.

[51] E. Adelson, "Depth-of-focus imaging process method," U.S. Patent 4 661 986, Apr. 28, 1987.

[52] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 19, pp. 4–19, Sep. 2014.

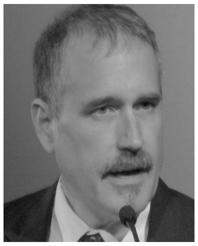
[53] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[54] S. S. Channappayya, A. C. Bovik, C. Caramanis, and R. W. Heath, "Design of linear equalizers optimized for the structural similarity index," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 857–872, Jun. 2008.

Codruta O. Ancuti received the M.Sc. degree from Universitatea Politehnica Timisoara in 2003 and the Ph.D. degree from Hasselt University in 2011. She was a Research Engineer with Siemens VDO from 2003 to 2005, where she was involved in developing different embedded system platforms for automotive industry. In 2005, she joined the Human–Computer Interaction Group and the Vision Computing Group, Hasselt University. She is currently a Senior Researcher/Lecturer with the MEO Group, Universitatea Politehnica Timisoara, and a Research Fellow with the Vi-COROB Group, University of Girona. Her main interest of research includes image understanding and visual perception. She is the first that introduced several single images-based enhancing techniques built on the multiscale fusion, such as color-to grayscale, image dehazing, underwater image, and video restoration.

Cosmin Ancuti received the M.Sc. degree from Universitatea Politehnica Timisoara (UPT), Romania, in 2003, and the Ph.D. degree from Hasselt University, Belgium, in 2009. From 2010 to 2012, he was a Post-Doctoral Fellow with IMINDS and Intel Exascience Lab, imec, Leuven, Belgium. He is currently a Senior Researcher/Lecturer with UPT and a Research Fellow with the Universite Catholique de Louvain, Belgium (European Marie-Curie mobility fellowship). He has authored over 40 papers published in international conference proceedings and journals. His area of interests includes image/video enhancement techniques, computational photography, and low level computer vision.

Christophe De Vleeschouwer was a Senior Research Engineer with imec from 1999 to 2000, a Post-Doctoral Research Fellow with the University of California at Berkeley from 2001 to 2002 and EPFL in 2004, and a Visiting Scholar with CMU from 2014 to 2015. He is currently a Senior Research Associate with Belgian NSF, and an Associate Professor with the ISP Group, Universite Catholique de Louvain. He has co-authored over 35 journal papers or book chapters. He holds two patents. His main interests lie in video and image processing for content management, transmission and interpretation. He is enthusiastic about nonlinear and sparse signal expansion techniques, ensemble of classifiers, multiview video processing, and graph based formalization of vision problems. He served as an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA.He has been a Co-Founder of Keemotion using video analysis for automatic sport coverage.



Alan C. Bovik (F'96) is currently the Director of the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, Institute for Neuroscience, The University of Texas at Austin. He has authored over 800 technical articles. He holds several U.S. patents. His research interests include image and video processing, digital television and digital cinema, computational vision, and visual perception. He is a member of the Television Academy, the National Academy of Television Arts and Sciences, and the Royal Society of Photography. He received the Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Television Academy in 2015, for his work on the development of video quality prediction models which have become standard tools in broadcast and postproduction houses throughout the television industry. He received a number of major awards from the IEEE Signal Processing Society, including the Society Award, the Technical Achievement Award, the Best Paper Award, the Signal Processing Magazine Best Paper Award, the Education Award, the Meritorious Service Award, and (co-author) the Young Author Best Paper Award. He also received the 2016 IEEE Circuits and Systems for Video Technology Best Paper Award. He was named Honorary Member Award of the Society for Imaging Science and Technology and received the Society of Photo-Optical and Instrumentation Engineers (SPIE) Technology Achievement Award. He was a IS&T/SPIE Imaging Scientist of the Year. He also received the Joe J. King Professional Engineering Achievement Award and the Hocott Award for Distinguished Engineering Research in 2008, from the Cockrell School of Engineering at The University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Urbana–Champaign in 2008. He is a fellow of the Optical Society of America and SPIE. He holds the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin. He also co-founded and was the longest-serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, and created and served as the first General Chair of the IEEE International Conference on Image Processing, Austin, TX, USA, in 1994. He was a registered Professional Engineer in the State of Texas. He is also a frequent consultant to legal, industrial, and academic institutions.