Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 360

Variational Image Fusion with Optimal Local Contrast

David Hafner and Joachim Weickert

Saarbrücken 2015

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 360
submitted: April 1, 2015

Variational Image Fusion with Optimal Local Contrast

David Hafner

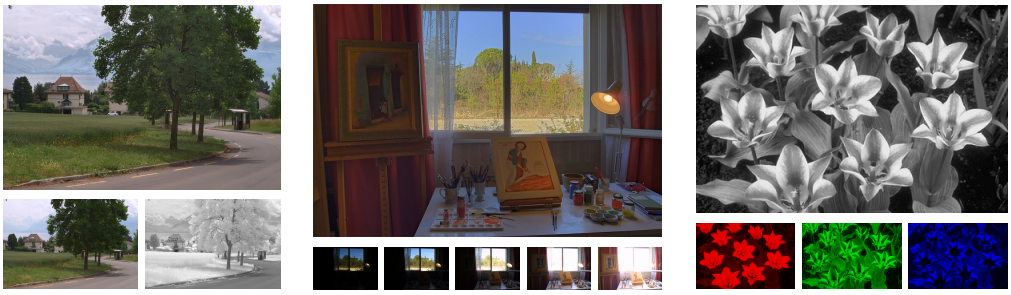
Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
hafner@mia.uni-saarland.de

Joachim Weickert

Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, Campus E1.7, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/



Multispectral imaging Exposure fusion Decolourisation

Figure 1: Exemplary applications of our general variational image fusion technique. Each of the resulting composite images (*top*) condenses the most important information from the input stack (*bottom*). (Input images: [14], J. Joffre [5], [17])

Abstract

In this paper, we present a general variational method for image fusion. In particular, we combine different images of the same subject to a single composite that offers optimal exposedness, saturation, and local contrast. Previous research approaches this task by first precomputing application specific weights based on the input, and then combining these weights with the images to the final composite later on. In contrast, we design our model assumptions directly on the fusion result. To this end, we formulate the output image as a convex combination of the input and incorporate concepts from perceptually inspired contrast enhancement such as a local and nonlinear response. This output-driven approach it the key to the versatility of our general image fusion model. In this regard, we demonstrate the performance of our fusion scheme with several applications such as exposure fusion, multispectral imaging, and decolourisation. For all application domains, we conduct thorough validations that illustrate the improvements compared to state-of-the-art approaches that are tailored to the individual tasks.

1 Introduction

The fusion of multiple images is a key component of many visual computing applications. Particularly, it is an essential tool when several photographs or sensors are required to capture all important structures of a scene. In this context, image fusion approaches aim at condensing the most important

information from the acquired image stack to single composite that is richer in details than any input image.
There exist many fusion approaches that are tailored to one specific application. Most of them pursue the following two-step pipeline: Based on application specific quality measures, they determine weights for each of the input images in a first step. Then, in a second step, these weights are combined with the input images to form the output image. Contrary to this two-stage approach, we propose a conceptually different idea in this paper. Instead of precomputing weights based on the input images, we present a variational method that directly aims for an output image with high quality. This has several advantages: First, all our model assumptions and parameters have an intuitive meaning and directly influence the fusion result in the desired way. Moreover, our approach produces images with visual phenomena such as a Cornsweet illusion that optimise the local contrast. This is not achievable with standard fusion methods. Most importantly, this output-driven idea is the key concept to refrain from an application specific weight precomputation and, in this way, to allow a general image fusion framework that performs well in many fusion applications.

Main Contributions. While the input data differ for every image fusion applications, most applications aim at the same goal: an output image with important perceptual qualities such as well-exposedness or a high local contrast. Based on this observation, we introduce a general variational framework for image fusion. We achieve this by refraining from an application specific precomputation of weights based on the input images. Instead, we formulate the fusion result as a convex combination of the input, and directly opt for an output that is optimal w.r.t. our energy functional. Here, we base our model assumptions on important perceptually inspired image enhancement concepts that account for the local contrast adaption of the human visual system. The minimisation of this energy yields to fusion results that capture the most important information from the input and feature important visual properties.
This paper is an extended version of our conference paper [34], where we introduced a variational exposure fusion method. Here, we show and discuss its general applicability to several fusion applications. In addition to exposure fusion, we demonstrate its performance for multispectral imaging and decolourisation. For all three main application domains, we conduct thorough evaluations on public image data sets and compare to previous work in the individual research areas. Last but not least, we improve the runtime of our algorithm by applying ideas of [10] and [32], in combination with a

parallel implementation on the graphics card.

Outline. First we discuss related work in Sec. 2. Afterwards we present our general variational image fusion model in Sec. 3. Its minimisation in Sec. 4 yields the desired composite image. In Sec. 5, we analyse our model and investigate its parameters. The evaluation in Sec. 6 demonstrates the versatility and performance of our technique for different image fusion applications. After discussing possible limitations of our method in Sec. 7, we conclude the paper with a summary and an outlook in Sec. 8.

2 Related Work

To explain how our model relates to previous work, let us first give a survey of our three main application domains. Furthermore, since we base our model assumptions on variational contrast enhancement concepts, we also review related work in this field.

2.1 Multispectral Imaging

In multispectral imaging, multiple sensors or filters are used to acquire a set of images. They all capture a different spectral range of the scene, e.g. the visible and the near-infrared band [29, 14]. The fusion of these images results in an image that offers more details than any of the photographs; see e.g. Fig. 1 (*left*). Based on brightness and saturation of the visible spectrum image, Zhang et al. [66] precompute a weight map to identify image regions that should be improved. Then, contrast and texture is transferred from the near-infrared to the visible spectrum. Lau et al. [38] and Eynard et al. [23] interpret the near-infrared as a fourth colour channel and regard multispectral image fusion as a colour transformation task. However, this has the following drawback: In some parts of the image the visible spectrum is more reliable than the near-infrared, and vice versa. Accordingly, it is not optimal to assume the same importance of all spectral ranges in every image region. Hence, contrary to [38] and [23], we interpret the near-infrared information as an additional spatial lightness information. As suggested in the psychophysical study of Fredembach and Süsstrunk [29], we thus explicitly model a spatially varying importance of the different spectral ranges.

2.2 Decolourisation

Although not immediately obvious, also decolourisation can be interpreted as an image fusion application. In general, decolourisation describes the conversion of a colour image to its greyscale representation, while preserving as much information as possible. For a detailed review and evaluation of decolourisation approaches before 2009, we refer to the paper of Čadík [17]. Thus, we mainly focus on more recent decolourisation methods here. A straightforward idea is to compute the greyscale image as the Y channel of the CIE XYZ colour space [25]. However, such a simple global mapping is not always sufficient to preserve the contrast. Hence, Smith et al. [59] first apply a global mapping, based on the Helmholtz-Kohlrausch effect. Then they locally add details with an unsharp masking technique. Gooch et al. [30] propose a fully local decolourisation method by forcing the differences between pixels in the greyscale image to resemble them from the colour image. Lu et al. [39] relax this constraint with a weak colour ordering, i.e. the sign of the colour difference is not assumed to be fixed. Also the structure preserving technique of Eynard et al. [23] is applicable for decolourisation. Another successful idea is to approach the decolourisation task by fusing the colour channels of the input image [6]. Generally, none of the colour channels or a simple global combination of them is sufficient to represent the colour image. However, a spatially varying channel fusion yields satisfying results; cf. Fig. 1 (*right*). In this paper, we also approach decolourisation by image fusion. However, contrary to [6], we do not precompute decolourisation specific fusion weights based on the input. Rather, we directly aim at an optimal output greyscale image. This not only affords the general applicability of our model to various fusion applications, but also allows to realise visual phenomena such as a Cornsweet illusion that optimise the local contrast. As it turns out, this is a very important feature to visually preserve the contrast if one considers the drastic intensity range restriction by a colour to greyscale transformation.

2.3 Exposure Fusion

Classical high dynamic range (HDR) methods combine several low dynamic range (LDR) images to one HDR image with the help of the exposure times and the camera response function; see e.g. [41, 21, 45, 62]. However, displaying those HDR results on standard monitors or printing them requires to compress the high dynamic range again. This process is called *tone mapping*; see [51] for a survey and [18] for a discussion and evaluation of various tone mapping operators. Since tone mapping is not the focus of this work, we restrict our discussion to the most related ones. In their gradient domain

tone mapper, Fattal et al. [27] account for the local contrast adaption of the visual system by attenuating large gradients, and maintaining or even enhancing the smaller ones. Similarly, Durand and Dorsey [22] decompose the HDR image into a base and a detail layer. Then they compress the base while keeping the details. Reinhard et al. [52] apply first a global transform, and locally increase the contrast afterwards. Also Mantiuk et al. [42] show and discuss the importance of the contrast adaption of the human visual system w.r.t. tone mapping. Most related to our work is the two-stage tone mapper of Ferradans et al. [28] that applies a variational contrast enhancement in the second stage.

However, if the main focus lies anyway on a displayable and well-exposed LDR image, there is a popular alternative to the described two-step procedure of HDR imaging and tone mapping, namely *exposure fusion* [44]. Here, the task is to skip the HDR image generation by a direct fusion of the differently exposed LDR images to an overall well-exposed composite. Such an exposure fusion approach has several advantages: First, there is no need to know the exposure times or the camera response function. It is even possible to include images that do not follow the HDR imaging model, e.g. flash and no flash photographs or images from different cameras. Second, this one-step approach allows a direct tuning of the final results without the detour via an intermediate HDR image. Obviously, exposure fusion is related to tone mapping. However, the different kind of input data ask for different algorithmic requirements and different model assumptions. In the meantime, exposure fusion has even developed to an own research area with various publications that we review next. Most existing exposure fusion methods pursue the following processing pipeline: In the first step, based on exposure fusion specific quality measures, weighting maps are determined for each of the input images. Such quality measures are for instance the magnitude of the Laplacian [13, 44], the entropy [31, 35], or the colour saturation [44, 57, 58]. Another idea, e.g. applied by Raman and Chaudhuri [50] or by Singh et al. [58], is to decompose the input images into base and detail layers. Then the amount of detail is considered as measure to determine the input image weights. In the second step, these weighting maps are combined with the input images to form the final composite. Here, the fusion strategies vary from region-based blending [31] and pixelwise weighted averaging [50, 35, 57, 56, 58] to gradient domain fusion [19, 60] and pyramid-based techniques [16, 13, 44]. Different to those two-step approaches, Raman and Chaudhuri [49] propose a variational method to directly compute the fused composite. However, this requires a smoothness constraint of the final image that may lead to over-smoothed blurry results. A more suitable idea by Kotwal and Chaudhuri [36] is to formulate the output image as a weighted

average of the input. Then, they design an energy on this composite. We follow a similar idea. However, we base our model assumptions on perceptually inspired contrast enhancement concepts. This allows to optimise the local contrast and to produce images with vivid colours. Moreover, we show that our general model is applicable to various fusion tasks.

2.4 Variational Contrast Enhancement

The discussions above show that there are many approaches that are specifically tailored to the individual tasks. However, all presented applications share a similar goal: The fusion of several images to one composite that offers optimal local contrast. We use this observation to present a general variational fusion approach. To this end, we profit from important findings in histogram modification and contrast enhancement that we review in this section. Based on the seminal work of Sapiro and Caselles [53] on histogram modification with differential equations, Bertalmío et al. [10] introduce a variational approach to locally increase the contrast of an image. In this context, Palma-Amestoy et al. [46] investigate several perceptually inspired energy terms. In his recent study, Bertalmío [8] shows connections to visual neuroscience. These contrast enhancement approaches have found first applications: Piella [47] incorporates a gradient domain term in the energy of [10]. This forces the similarity to a precomputed gradient field that combines the gradients from multiple images. However, the weights for the individual images are predetermined based on the input. In contrast, our energy minimisation that directly aims at an optimal composite is able to refrain from such an application specific weight precomputation. Recently, Bertalmío et al. [11] propose a gradient-based variational approach to fuse a pair of images with different exposure times. This method is specifically tailored to two input images and cannot be extended to multiple images in a straightforward way. Nevertheless, these applications motivate us to also base our model on those perceptually inspired contrast enhancement concepts. They clearly have demonstrated their usefulness, and their perceptual basis has been extensively discussed in various publications; see e.g. [10, 46, 9, 48, 8].

3 Variational Model

Our general goal is to fuse n input images $f_1, \dots, f_n$ to a single composite u that condenses the most important information from the image stack. To this end, we first formulate the output image as a convex combination of the

input:

$$u(\boldsymbol{x}) = \sum_{i=1}^{n} w_i(\boldsymbol{x}) \cdot f_i(\boldsymbol{x})\,, \tag{1}$$

with

$$w_i(\boldsymbol{x}) \geq 0 \qquad \text{and} \qquad \sum_{i=1}^{n} w_i(\boldsymbol{x}) = 1\,. \tag{2}$$

Here, $\boldsymbol{x} = (x_1, x_2)^\top$ denotes the position on the rectangular image domain $\Omega \subset \mathbb{R}^2$ and w_i the weight of the image f_i.
As discussed in Sec. 2, most previous research concentrates on determining weights based on application specific quality measures that are defined on the input images. In contrast to such a weight precomputation, we directly opt for an optimal fusion result u. To this end, we propose the following energy functional:

$$\begin{aligned} E(\boldsymbol{w}) = &\frac{1}{2} \int_{\Omega} \Big(\big(u(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\big)^2 + \delta \cdot (u(\boldsymbol{x}) - \mu)^2 \Big)\, \mathrm{d}\boldsymbol{x} \\ &- \frac{\gamma}{2} \iint_{\Omega}\int_{\Omega} g_\sigma(\boldsymbol{x}, \boldsymbol{y}) \cdot \Psi(u(\boldsymbol{x}) - u(\boldsymbol{y}))\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y} \\ &+ \frac{\alpha}{2} \int_{\Omega} \sum_{i=1}^{n} |\boldsymbol{\nabla} w_i(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x} \end{aligned} \tag{3}$$

subject to the constraints (1) and (2). The image weights $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ that follow from this energy optimisation can be seen as a side-product of our output-driven approach. In fact, we are mostly interested in the fused image u. However, formulating this image as a convex combination of the input allows to impose a smoothness constraint on the weights and not on the image itself. While the later is prone to cause over-smoothed blurry fusion results, the former is a much more intuitive and meaningful assumption. A further important advantage of this formulation is the inherent close attachment of u to the input data that prevents visual artefacts and an unrealistic appearance of the fusion results.
As discussed in Sec. 2, the energy functional in (3) is inspired by successful variational histogram modification and contrast enhancement techniques [53, 10, 46, 9]. In particular, these works discuss and analyse in which way the energy terms mimic important properties of the human visual system and e.g. relate to Land's Retinex theory [37].
Following [10], we model a so-called *dispersion term* (first line in (3)): The first part first part of this dispersion term forces u to resemble the attachment

image $\bar{f}$, which we choose as an average of the input images. As discussed in [46], this provides an attachment to the original data and accounts for the colour constancy assumption [37]. The second part implements the grey world principle [15, 55, 46]. It provides well-exposed images by keeping the solution close to the constant μ. Here, the influence of the second assumption can be steered with the positive parameter δ.
The second term, the *contrast term*, counteracts this dispersion term since it penalises uniform images more than images with a high local contrast. One should note the minus sign in front of the contrast term. Intuitively speaking, this energy term favours solutions that differ much from pixel to pixel. Here, the locality is introduced by the Gaussian weighting $g_\sigma(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-|\boldsymbol{x}-\boldsymbol{y}|^2}{2\sigma^2}\right)$. Furthermore, the function $\Psi_\lambda(z) = \sqrt{z^2 + \lambda^2}$ provides a nonlinear behaviour. In accordance with [10], its sigmoid-shaped derivative $\Psi'_\lambda(z) = \frac{z}{\sqrt{z^2+\lambda^2}}$ that appears in the algorithmic iteration (cf. Sec. 4) mimics the nonlinear response of the human visual system in the sense of a contrast transducer function [64, 43]. The parameter λ allows to tune this nonlinearity, and $\gamma \geq 0$ weights the influence of the contrast term.
The third term in our energy functional is a *regularisation term* that rewards smooth weight maps. It renders the assumption that neighbouring pixels in the fused composite should have similar weights. Here, $\boldsymbol{\nabla} := (\partial_{x_1}, \partial_{x_2})^\top$ denotes the gradient operator and $\alpha \geq 0$ steers the amount of smoothness. For some scenarios, more sophisticated edge-preserving smoothness terms might be beneficial. However, in all considered applications, they did not lead to significant improvements. In fact, mostly a smooth blending of the input images is desirable.
Last but not least, the simplex constraint (2) restricts the fusion result to pixelwise convex combinations of the input images. Hence, provides a close attachment to the input data. In combination with the smoothness constraint, this prevents an unnatural high amount of contrast and undesirable artefacts such as colour shifts or halos in the composite image.

Colour Image Processing. For the sake of simplicity, we have restricted ourselves to greyscale images so far. In case of colour images, we transform the input images from the RGB to the YCbCr colour space and define the dispersion and contrast term on the luminance channel only. To prevent a colour cast, we compute joint weight maps for all channels. Moreover, saturated colours make images to look vivid and expressive. To this end, we

Algorithm 1: Our general variational image fusion.

Input: input images $f_1, \ldots, f_n$
Output: fused composite u

set attachment image $\bar{f}$
repeat
    gradient descent step (Eq. 5)
    projection onto simplex (Algo. 2)
until *convergence*
assemble u (Eq. 1)

extend our energy (3) for colour images with the following *saturation term*:

$$-\frac{\beta}{2} \cdot \int_{\Omega} \Big(\big(u_{Cb} - 1/2\big)^2 + \big(u_{Cr} - 1/2\big)^2 \Big) \, \mathrm{d}\boldsymbol{x} \,, \tag{4}$$

where u_{Cb} and u_{Cr} denote the chroma channels of u. This term favours values different from grey, and thus images with vivid colours. The positive parameter β allows to control the amount of colour saturation. Once again, the minus sign should be noted.

4 Numerical Algorithm

To solve the discussed optimisation problem, we basically apply a gradient projection method; see e.g. Bertsekas [12]. Generally, each iteration consists of (i) a gradient descent step, followed by (ii) a projection onto the simplex; see Algorithm 1.

4.1 Gradient Descent Step

With iteration index k and time step size τ, the gradient descent of energy (3) with saturation term (4) reads

$$\begin{aligned} w_i^{k+1}(\boldsymbol{x}) = w_i^k(\boldsymbol{x}) - \tau \Big(& f_{Y_i}(\boldsymbol{x}) \Big(u_Y^k(\boldsymbol{x}) - \bar{f}_Y(\boldsymbol{x}) + \delta \, (u_Y^k(\boldsymbol{x}) - \mu) \\ & - \gamma \int_{\Omega} g_\sigma(\boldsymbol{x}, \boldsymbol{y}) \cdot \Psi_\lambda'(u_Y^k(\boldsymbol{x}) - u_Y^k(\boldsymbol{y})) \, \mathrm{d}\boldsymbol{y} \Big) \\ & - \beta \Big(f_{Cb_i}(\boldsymbol{x}) \big(u_{Cb}^k - 1/2\big) + f_{Cr_i}(\boldsymbol{x}) \big(u_{Cr}^k - 1/2\big) \Big) \\ & - \alpha \, \Delta w_i^k(\boldsymbol{x}) \Big) \,, \end{aligned} \tag{5}$$

Algorithm 2: Projection onto simplex [54].

Input: weights $\boldsymbol{w}$

Output: projected weights $\tilde{\boldsymbol{w}}$

$\boldsymbol{s} = \text{sort}(\boldsymbol{w})$ such that $s_1 \geq \ldots \geq s_n$

$m = \max \left\{ j \in \{1, \ldots, n\} \;\middle|\; s_j - \frac{1}{j} \left(\sum_{i=1}^{j} s_i - 1 \right) > 0 \right\}$

$\theta = \frac{1}{m} \left(\sum_{i=1}^{m} s_i - 1 \right)$

$\tilde{w}_i = \max \{ w_i - \theta, 0 \}$

for $i = 1, \ldots, n$ and with $(u_Y^k, u_{Cb}^k, u_{Cr}^k)^\top = \sum_{i=1}^{n} w_i^k(\boldsymbol{x}) \cdot (f_{Yi}, f_{Cbi}, f_{Cri})^\top$. We discretise Eq. 5 with finite differences on a rectangular grid with uniform grid sizes, and approximate the integral with the rectangle method. Furthermore, we initialise w_i with $1/n$, and assume the image to be mirrored at the boundaries.

4.2 Projection onto Simplex

After each gradient descent step, we account for the simplex constraint (2). We realise this by projecting the computed weights $\boldsymbol{w}^k$ onto the n-dimensional simplex with Algorithm 2 [54].

4.3 Algorithmic Speed-Up

Due to the integral term in (5), the computational complexity of a gradient descent step is $\mathcal{O}(N^2)$, where N denotes the number of pixels. As remedy, we apply the following numerical approximation [10]: First, we approximate the odd sigmoid-shaped function $\Psi'_\lambda(z)$ with the polynomial $\sum_{i=0}^{m} c_i z^i$ of degree m (here: $m = 7$). Specifically, we obtain the coefficients $c_0, \ldots, c_m$ by minimising the quadratic error function (assuming intensity values between 0 and 1)

$$F(c_0, \ldots, c_m) \;=\; \int_{-1}^{1} \Big(\sum_{i=0}^{m} c_i z^i - \Psi'_\lambda(z) \Big)^2 \, \mathrm{d}z \,. \tag{6}$$

This allows to approximate the term $\Psi'_\lambda(u(\boldsymbol{x}) - u(\boldsymbol{y}))$ by

$$\begin{aligned}\Psi'_\lambda(u(\boldsymbol{x}) - u(\boldsymbol{y})) &\approx \sum_{i=0}^{m} c_i \left(u(\boldsymbol{x}) - u(\boldsymbol{y})\right)^i \\ &= \sum_{i=0}^{m} c_i \left(\sum_{j=0}^{i} \binom{i}{j} u^{i-j}(\boldsymbol{x}) \, (-1)^j \, u^j(\boldsymbol{y}) \right) \\ &= \sum_{j=0}^{m} \left(\sum_{i=j}^{m} (-1)^j \, c_i \binom{i}{j} u^{i-j}(\boldsymbol{x}) \right) u^j(\boldsymbol{y}) \,,\end{aligned} \tag{7}$$

with the binomial coefficient $\binom{i}{j} = \frac{i!}{j!(i-j)!}$. Now we can write the integral term as

$$\sum_{j=0}^{m} \Bigg(\underbrace{\Big(\sum_{i=j}^{m} (-1)^j c_i \binom{i}{j} u^{i-j}(\boldsymbol{x}) \Big)}_{\text{(I)}} \underbrace{\int_\Omega g_\sigma(\boldsymbol{x}, \boldsymbol{y}) u^j(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}}_{\text{(II)}} \Bigg). \tag{8}$$

We see that (I) can easily be computed with powers of $u(\boldsymbol{x})$. Moreover, (II) is nothing else than a Gaussian convolution which we compute with a fast recursive algorithm [65]. In this way, the overall complexity reduces from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. To further speed up the algorithm, we exchange the gradient descent step by a Runge-Kutta scheme and accelerate it with a strategy that is similar to *Fast Explicit Diffusion* [32]. All this allows a fast parallel implementation on the graphics card.

5 Model and Parameter Analysis

5.1 Benefits of Local Contrast Term

Inspired by [10], let us first consider a variational histogram equalisation in Fig. 2. In our framework, this can be achieved by setting $\alpha = 0$ and replacing the simplex constraint (2) by $0 \leq w(\boldsymbol{x}) f(\boldsymbol{x}) \leq 1$. Applying a global contrast term, i.e. degrading $g_\sigma(\boldsymbol{x}, \boldsymbol{y})$ to the constant $1/|\Omega|$, yields a standard histogram equalisation (*middle*). On the other hand, a local contrast term allows to visually increase the contrast in the sense of a Cornsweet illusion [20, 55] (*right*). This illustrates the general advantages of a local contrast term compared to a global one.

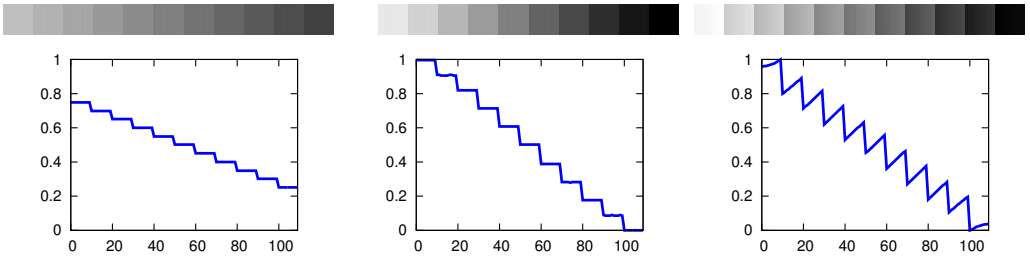


Figure 2: Global vs. local contrast term for histogram equalisation. *From left to right*: Input, with global contrast term, with local contrast term. *Top*: Intensity images. *Bottom*: Corresponding scanlines, where the x-axes represent the intensity values and the y-axes the pixels.

Figure 3: Benefits of output-driven optimisation. *From left to right:* Input colour image, result of Ancuti et al. [6], our result. Our output-driven optimisation implements a local contrast adaption that keeps the colour patches in the greyscale image distinguishable.

5.2 Benefits of Output-Driven Optimisation

Thanks to our output-driven optimisation, our fusion technique is capable of producing fusion results with the just discussed Cornsweet illusions. This is very hard to accomplish with standard fusion methods since they do not take into account the quality of the final output image when precomputing weights for the input images in advance. This is illustrated in Fig. 3 with a decolourisation example: Ancuti et al. [6] (*middle*) precompute decolourisation specific weights based on the input and fuse the images later on. In this way, it is not possible to create greyscale gradients in the individual patches. In contrast, our output-driven optimisation produces such gradients that optimise the local contrast (*right*). Furthermore, our direct optimisation w.r.t. the output image is the key concept that allows us to present a general technique for image fusion without tailoring it to the individual applications.

Figure 4: Influence of parameters. *From left to right*: μ (0.1, 0.9), σ (1%, 100% of image diagonal), γ (0, 0.5), β (0, 2).

5.3 Model Parameters

Let us now illustrate the influence of our main model parameters on the example of fusing an exposure set; see Fig. 1 (*middle*). In the *first column* of Fig. 4, we apply different values of μ. Since the dispersion term favours solutions that are close to μ, it is obvious that larger values lead to brighter results. We propose to compute μ automatically as the average of the input images. The *second column* depicts composite images for different scales σ of the Gaussian g_σ in the contrast term. We observe a larger local contrast with decreasing σ. There is a trade-off: A too large local contrast might be perceived as unnatural, a too small one as too flat. As a rule of thumb, we propose to set σ to 10% of the image diagonal. Similar observations apply to the contrast parameter γ (*third column*). Choosing it too small yields an image with low contrast, and choosing it too large gives unrealistic looking images. In general, setting it to 1/4 provides good results. Last but not least, the *fourth column* shows the effects of the proposed saturation term and its parameter β. Generally, a larger value of β leads to more saturated colours, and in this way to a more vivid appearance.
In all our experiments below, we apply the discussed procedure to determine μ and σ automatically. All other parameters are fixed for the individual applications. Table 1 shows the standard parameter setting. To conclude, all our model parameters have an intuitive meaning and are fixed or can be determined automatically. This allows an easy and straightforward use of our approach, also for non-experts.

Table 1: Default parameter setting.

α	β	γ	δ	λ	μ	σ
1	1	1/4	1	0.1	input average	10% of diagonal

Table 2: Runtimes on an *NVIDIA GeForce GTX 970*.

number of images	image size	runtime in seconds
3	640×480	4
3	1024×768	7
5	640×480	8
5	1024×768	18

5.4 Runtime Experiments

Our reference implementation is written in *CUDA* and runs on an *NVIDIA GeForce GTX 970* graphics card. We assume the algorithm to be converged if the root mean square difference of two fusion results between 100 iterations is less than $1 \cdot 10^{-4}$. Table 2 lists the runtimes for different number of images and different image sizes.

6 Evaluation

In this section, we demonstrate the performance of our method on different fusion applications. To evaluate our general technique, we compare to the best performing state-of-the-art approaches in all three main application areas, using public benchmark images. For fair comparisons, we either use the implementations of the authors with the default parameter settings or, if provided, their resulting images directly.

6.1 Multispectral Imaging

The idea of multispectral imaging is to capture different spectral ranges with two or more photographs of the same scene. Fusing them allows to produce an image that offers details that cannot be captured only within the visible spectrum. In this regard, Fig. 1 (*left*) illustrates the performance of our method for the enhancement of standard photographs (RGB) with an near-infrared image (NIR). We process the images in the following way: First we convert the RGB image to the YCbCr colour space. Then we apply the presented technique to fuse the luminance channel Y with the NIR image. Since we want to stay close to the visible spectrum image while adding details

Figure 5: Comparison of multispectral image fusion. *From left to right*: Lau et al. [38], Eynard et al. [23], and proposed. *Top*: Full images. *Bottom*: Zooms. Our fusion result features the most local details from both spectral ranges without appearing unrealistic. (Input images: [2])

from the near-infrared range, we choose the luminance channel as attachment image $\bar{f}$. Finally, we combine the fused luminance composite with the original chroma channels to form the output image.

In Fig. 5, we show our fusion result (*last column*) for a test image set used by Lau et al. [38] and Eynard et al. [23]. As mentioned in Sec. 2, contrary to Lau et al. (*third column*) and Eynard et al. (*fourth column*), we do not regard the near-infrared range as a fourth colour channel. Instead, inspired by psychophysical studies [29], we treat it as an additional spatial lightness information. As our result demonstrates, this is an important model assumption. Compared to both competing approaches, our fused image offers the most details from both spectral ranges; see for instance the trees and the mountains. Besides providing better quality, our approach has an additional important advantage compared to the approach of Eynard et al.: Their method requires the user to specify the output colour of some pixels explicitly. This is important for reasonable results. In contrast, our output-driven approach requires no such user interaction and works fully automatic.

To underline these findings, we conduct a detailed comparison to Eynard et al. [23] on further image sets. To this end, we use the public database by Brown and Süsstrunk [14], which contains RGB and NIR images for different real-world sceneries. Fig. 6 depicts for three of those image sets the input RGB images (*left*), the results of Eynard et al. (*middle*), and our results (*right*). Since the method of Eynard et al. requires user interaction by an expert, we compare to resulting images that are kindly provided by them.

First, the experiments show that both approaches add important structures to the final images that have not been visible in the input RGB images. However, in this respect, the zooms illustrate that our results are richer in details. Second, for the results of Eynard et al., one can observe undesirable colour casts compared to the RGB images; cf. for instance Fig. 6 (*last row*). This would be acceptable if the details would increase significantly in this way. However, this is not the case. Since we regard the NIR image as an additional spatial lightness information, our results do not suffer from such artefacts. To demonstrate the overall good performance of our general fusion approach for the multispectral imaging application, we include our resulting images for all 477 image sets from the mentioned database in the supplementary material (`http://www.mia.uni-saarland.de/Research/Image_Fusion`).
In our last multispectral experiment (Fig. 7), we deal with more than two input images or spectral ranges, respectively (*first four columns*). Also here, our fused image in the *last column* demonstrates the capability of our method to incorporate structures from all spectral ranges, without leading to an unrealistic appearance.

6.2 Decolourisation

We approach the decolourisation task in the following way: Basically we regard the RGB channels of the colour image as input of our fusion algorithm. Then we fuse all three channels to a grey-valued image that offers an optimal local contrast. Here we use the average over all colour channels as attachment image $\bar{f}$ and set δ to zero. Fig. 1 (*right*) illustrates this procedure and depicts our corresponding decolourisation result. While a single channel is not enough to preserve the details from the colour image, our fused composite is.
In [17], Čadík presents a decolourisation benchmark containing 24 images that represent various image classes. Based on this benchmark, we compare to different state-of-the-art decolourisation approaches. In Fig. 8, we depict three of those 24 images. Due to space restrictions, we decided to put the rest to the supplementary material. There we additionally compare to results from other decolourisation techniques. Here we restrict ourselves to the best performing ones, namely the method of Lu et al. [39] (*second column*) and Eynard et al. [23] (*third column*). All decolourisation results in Fig. 8 are of high quality. However, in some image regions, our greyscale images (*last column*) preserve more details from the colour image than the other ones.
Lu et al. [40] argue that the decolourisation benchmark by Čadík [17] is biased to synthetic images. Hence, they propose a new one with 250 colour images. Additionally, they introduce the so-called *E-score* quality measure to quan-

titatively judge different decolourisation results. Unfortunately, as it turns out, this measure does not take the local contrast adaption of the human visual system into consideration. Rather, it even penalises effects such as a Cornsweet illusion with high errors. This is illustrated in Fig. 9: Although, the right image offers a better visual quality, it receives a significantly worse E-score. Here, a higher E-score means a better quality, and τ defines a colour difference that is not perceivable by the human visual system; cf. [40]. Hence we decided to apply a global contrast term for the E-score evaluation. The corresponding E-score graphs for the benchmarks of Čadík [17] and Lu et al. [40] are depicted in Fig. 10. Considering the benchmark of Čadík (*left*), we obtain the best quality measures for almost all τ values. For the decolourisation benchmark of Lu et al. (*right*), we get results of similar quality w.r.t. their E-score measure. We even outperform all other methods for larger τ. First, this shows that our general image fusion approach performs well for the decolourisation task. It produces greyscale images that are of similar or even superior quality than state-of-the-art approaches which are tailored to decolourisation. Second, as Fig. 9 illustrates, our results could be even improved with a local contrast term. In this regard, we believe that an extension of a decolourisation quality measure to incorporate such properties of the visual system is of great importance. We plan this in future work.
colour

6.3 Exposure Fusion

The task of exposure fusion is to combine differently exposed images to a single composite that offers optimal exposedness and local contrast. This is illustrated in Fig. 1 (*middle*). Here, our fused image contains details from all input images, is well-exposed, and features a good contrast.
To evaluate our general image fusion method for the task of exposure fusion, we conduct a thorough comparison to two state-of-the-art approaches. First, we consider the popular exposure fusion method by Mertens et al. [44]. It builds the basis of many exposure fusion approaches and provides still top results compared to more recent techniques. Second, we consider the method of Singh et al. [58]. It is a very recent exposure fusion method that provides high quality results and compares favourably to other state-of-the-art techniques. As mentioned in Sec. 2, both methods precompute weights based on the input images, and fuse them later on. Here we will show that our idea, which directly aims for an optimal output image, is able to outperform those input based approaches. In agreement with Singh et al., we selected twelve representative exposure sets from the HDR photographic survey of Fairchild [24]. Each set consists of nine differently exposed LDR images.

Due to space constraints, we depict in Fig. 11 results for three exposure sets. A comparison for the other sets can be found in the supplementary material. There, we additionally include our fusion results for all images sets (105) by Fairchild to show the generally good performance of our method for exposure fusion. Especially in the zooms in Fig. 11, the higher amount of local contrast provided our approach is obvious, both in dark and bright image regions. Moreover, our images do not suffer from halos or blurring effects that can be observed for the method of Singh et al. [58].

Unfortunately, no meaningful perceptually motivated quality measures exist to objectively evaluate the exposure fusion results based on the given input LDR images. However, in a related context, Aydin et al. [7] introduced the so-called *Dynamic Range Independent Metric* (DRIM). This metric bases on properties of the human visual system and can be applied to compare images with a different dynamic range, e.g. an HDR reference image and an LDR representative of it. To have such a ground truth reference image, we use publicly available HDR images [1] and create sets of LDR images with different exposure times from it. To ensure that those images are representatives of real LDR images, we apply appropriate exposure times and a common camera response function. Specifically, we compute for each HDR image five LDR images separated by one exposure value. These images serve as input for the exposure fusion techniques. Finally, we apply DRIM to compare the reference (ground truth) HDR image with the fused LDR results. DRIM uses the following colour code: Green indicates a loss of visible contrast, blue an amplification of invisible contrast, and red a reversal of visible contrast. In addition, the colour saturation is proportional to the amount of distortion. Fig. 12 depicts the fusion results and the corresponding DRIM distortion maps for two image sets produced by the method of Mertens et al. (*left*), by Singh et al. [58] (*middle*), and our approach (*right*). As the DRIM measure indicates, our results show less distortions. This illustrates that our model assumptions, which are based on perceptually inspired concepts from variational contrast enhancement, are well-suited to create exposure fusion results with a high perceptual quality.

In our last exposure fusion experiment in Fig. 13, we additionally compare to the method of Kotwal and Chaudhuri [36] since it is a related exposure fusion approach. The higher local contrast provided by our approach compared to all other methods is visible. Furthermore, in comparison to Kotwal and Chaudhuri, our method does not suffer from a colour cast and provides a vivid colour impression. We achieve this by the proposed saturation term and the coupled handling of all colour channels. Considering the result of Singh et al., our result does not only feature more local contrast, but is also free from an unrealistic and undesirable detail enhancement; see e.g. the

wall.

6.4 Multilight Image Collection

In contrast to the classical two-stage pipeline of high dynamic range imaging and tone mapping, our approach is completely independent of the knowledge of the exposure times and the camera response function. On top of that, also images that do not follow the HDR imaging model, e.g. images from different cameras or images captured under non-constant lightning conditions, can be included easily into the input stack. We illustrate this by fusing a flash and no-flash image set in Fig. 14, and by fusing images with changing illumination in Fig. 15.

7 Limitations

Our model is tailored to static scenes, i.e. the input images should be aligned. Obviously, this is no problem for decolourisation. However, exposure series and multispectral acquisition techniques may suffer from either camera or object motion. This results in so-called *ghosting* artefacts in the fused image; cf. Fig. 16. To overcome this problem, a registration of the input images in a preprocessing step is required; see e.g. [33] for the alignment of exposure series. Moreover, our variational approach generally allows to directly incorporate specific deghosting strategies (cf. [61, 63] and references therein) in the energy functional. Such an adaptation of our fusion technique to unaligned images is part of our future work.

8 Conclusions

In this paper, we have presented a general variational method for image fusion. The main difference to previous research is that we intentionally refrain from precomputing application specific weights based on the input images, and combining the images with those predetermined weights later on. Instead, we model an energy functional that directly opts for an optimal composite image. This output-driven idea is the key concept why our method works that well in various application areas. We demonstrate this by means of thorough evaluations in all three main application areas, namely multispectral imaging, decolourisation, and exposure fusion. We compare to the state-of-the-art and best performing methods of each field. As it turns out, our general approach produces results of similar high quality, and even outperforms competing methods for various example images. This shows the

generally good performance and versatility of our technique. On top of that, all components of our variational model have an intuitive meaning and allow a direct manipulation of desired properties of the output. In this regard, all parameters can be fixed or determined automatically in a straightforward way. This allows an easy use, even by non-experts. To conclude, we believe that our presented technique is generally suitable for applications that require the fusion of multiple images. It provides a composite that condenses the most important information from the input images in an adequate way, and offers important features such a local contrast adaption.
As discussed in Sec. 7, the extension of our method to the case of unaligned input images is part of our future work. Additionally, we plan to test and evaluate our general approach for even more fusion applications, e.g. focus fusion. First steps in this direction show very promising results.

Acknowledgments. We would like to thank Prof. Mark Fairchild for providing the LDR exposure sets, and Harbinder Singh as well as Dr. Davide Eynard for executing their algorithms on the validation images. Our research has been partially funded by the Deutsche Forschungsgemeinschaft (DFG) through a Gottfried Wilhelm Leibniz Prize for Joachim Weickert. This is gratefully acknowledged.

References

[1] `resources.mpi-inf.mpg.de/hdr/gallery.html`.

[2] `www.comp.nus.edu.sg/~photo/projects/nir.html`.

[3] `www.cs.columbia.edu/CAVE/software/rascal/rrslrr.php`.

[4] `www.gigapan.com/galleries/10008/gigapans`.

[5] `www.hdrsoft.com/examples2.html`.

[6] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert. Image and video decolorization by fusion. In R. Kimmel, R. Klette, and A. Sugimoto, editors, *Computer Vision — ACCV 2010*, volume 6492 of *Lecture Notes in Computer Science*, pages 79–82. Springer, Berlin, 2011.

[7] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27(3):Article No. 69, Aug. 2008.

[8] M. Bertalmío. From image processing to computational neuroscience: A neural model based on histogram equalization. *Frontiers in Neuroscience*, 8:Article No. 71, June 2014.

[9] M. Bertalmío, V. Caselles, and E. Provenzi. Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision*, 83(1):101–119, June 2009.

[10] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. *IEEE Transactions on Image Processing*, 16(4):1058–1072, Apr. 2007.

[11] M. Bertalmío and S. Levine. Variational approach for the fusion of exposure bracketed pairs. *IEEE Transactions on Image Processing*, 22(2):712–723, Feb. 2013.

[12] D. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. *IEEE Transactions on Automatic Control*, 21(2):174–184, Apr. 1976.

[13] L. Bogoni. Extending dynamic range of monochrome and color images through fusion. In *Proc. International Conference on Pattern Recognition*, volume 3, pages 7–12, Barcelona, Spain, Sept. 2000.

[14] M. Brown and S. Süsstrunk. Multi-spectral SIFT for scene category recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1063–6919, Colorado Springs, CO, June 2011. `ivrg.epfl.ch/supplementary_material/cvpr11/`.

[15] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, July 1980.

[16] P. Burt and R. Kolczynski. Enhanced image capture through fusion. In *Proc. International Conference on Computer Vision*, pages 173–182, Berlin, Germany, May 1993.

[17] M. Čadík. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum*, 27(7):1745–1754, Oct. 2008. `cadik.posvete.cz/color_to_gray_evaluation/`.

[18] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics*, 32(3):330–349, June 2008. `dcgi.felk.cvut.cz/home/cadikm/tmo/`.

[19] W.-H. Cho and K.-S. Hong. Extending dynamic range of two color images under different exposures. In *Proc. International Conference on Pattern Recognition*, volume 4, pages 853–856, Cambridge, UK, Aug. 2004.

[20] T. N. Cornsweet. *Visual Perception*. Harcourt College Publishers, Fort Worth, 1970.

[21] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH 1997*, pages 369–378, Los Angeles, CA, Aug. 1997.

[22] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, July 2002.

[23] D. Eynard, A. Kovnatsky, and M. M. Bronstein. Laplacian colormaps: a framework for structure-preserving color transformations. *Computer Graphics Forum*, 33(2):215–224, May 2014.

[24] M. D. Fairchild. The HDR photographic survey. In *Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 6, pages 233–238, Albuquerque, NM, Nov. 2007.

[25] M. D. Fairchild. *Color Appearance Models*. Wiley-IS&T, Chichester, 3 edition, 2013.

[26] R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics*, 26(3):Article No. 51, July 2007.

[27] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, July 2002.

[28] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. An analysis of visual adaptation and contrast perception for tone mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, Oct. 2011.

[29] C. Fredembach and S. Süsstrunk. Colouring the near-infrared. In *Proc. IS&T/SID Color Imaging Conference*, pages 176–182, Portland, OR, Nov. 2008.

[30] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch. Color2Gray: Salience-preserving color removal. *ACM Transactions on Graphics*, 24(3):634–639, July 2005.

[31] A. A. Goshtasby. Fusion of multi-exposure images. *Image and Vision Computing*, 23(6):611–618, June 2005.

[32] S. Grewenig, J. Weickert, and A. Bruhn. From box filtering to fast explicit diffusion. In M. Goesele, S. Roth, A. Kuijper, B. Schiele, and K. Schindler, editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 533–542. Springer, Berlin, 2010.

[33] D. Hafner, O. Demetz, and J. Weickert. Simultaneous HDR and optic flow computation. In *Proc. International Conference on Pattern Recognition*, pages 2065–2070, Stockholm, Sweden, Aug. 2014.

[34] D. Hafner and J. Weickert. Variational exposure fusion with optimal local contrast. In J.-F. Aujol, M. Nikolova, and N. Papadakis, editors, *Scale-Space and Variational Methods in Computer Vision*, Lecture Notes in Computer Science. Springer, Berlin, 2015. to appear.

[35] J. Herwig and J. Pauli. An information-theoretic approach to multi-exposure fusion via statistical filtering using local entropy. In *Proc. International Conference on Signal Processing, Pattern Recognition and Applications*, pages 50–57, Innsbruck, Austria, Feb. 2010.

[36] K. Kotwal and S. Chaudhuri. An optimization-based approach to fusion of multi-exposure, low dynamic range images. In *Proc. International Conference on Information Fusion*, pages 1942–1948, Chicago, IL, July 2011.

[37] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, Jan. 1971.

[38] C. Lau, W. Heidrich, and R. Mantiuk. Cluster-based color space optimizations. In *Proc. International Conference on Computer Vision*, pages 1172–1179, Barcelona, Spain, Nov. 2011.

[39] C. Lu, L. Xu, and J. Jia. Contrast preserving decolorization. In *Proc. IEEE International Conference on Computational Photography*, pages 1–7, Seattle, WA, Apr. 2012.

[40] C. Lu, L. Xu, and J. Jia. Contrast preserving decolorization with perception-based quality metrics. *International Journal of Computer Vision*, 110(2):222–239, Nov. 2014.

[41] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T Annual Conference*, pages 442–448, Springfield, VA, May 1995.

[42] R. Mantiuk, K. Myszkowski, and H.-P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception*, 3(3):286–308, July 2006.

[43] J. J. McCann. The role of simple nonlinear operations in modeling human lightness and color sensations. In B. E. Rogowitz, editor, *Human Vision, Visual Processing, and Digital Display*, volume 1077 of *Proc. SPIE*, pages 355–363. SPIE Press, Bellingham, 1989.

[44] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28(1):161–171, Mar. 2009. `research.edm.uhasselt.be/~tmertens/`.

[45] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 374–380, Fort Collins, CO, June 1999.

[46] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles. A perceptually inspired variational framework for color enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):458–474, Mar. 2009.

[47] G. Piella. Image fusion for enhanced visualization: A variational approach. *International Journal of Computer Vision*, 83(1):1–11, June 2009.

[48] E. Provenzi and V. Caselles. A wavelet perspective on variational perceptually-inspired color enhancement. *International Journal of Computer Vision*, 106(2):153–171, Aug. 2014.

[49] S. Raman and S. Chaudhuri. A matte-less, variational approach to automatic scene compositing. In *Proc. International Conference on Computer Vision*, pages 574–579, Rio de Janeiro, Brazil, Oct. 2007.

[50] S. Raman and S. Chaudhuri. Bilateral filter based compositing for variable exposure photography. In *Proc. EUROGRAPHICS 2009 (Short Papers)*, pages 369–378, Munich, Germany, Mar. 2009.

[51] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Elsevier, Oxford, 2 edition, 2010.

[52] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, July 2002.

[53] G. Sapiro and V. Caselles. Histogram modification via differential equations. *Journal of Differential Equations*, 135(2):238–268, Apr. 1997.

[54] S. Shalev-Shwartz and Y. Singer. Efficient learning of label ranking by soft projections onto polyhedra. *Journal of Machine Learning Research*, 7:1567–1599, July 2006.

[55] R. Shapley and C. Enroth-Cugell. Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:1567–1599, Mar. 1984.

[56] R. Shen, I. Cheng, and A. Basu. QoE-based multi-exposure fusion in hierarchical multivariate Gaussian CRF. *IEEE Transactions on Image Processing*, 22(6):2469–2478, June 2013.

[57] R. Shen, I. Cheng, J. Shi, and A. Basu. Generalized random walks for fusion of multi-exposure images. *IEEE Transactions on Image Processing*, 20(12):3634–3646, Dec. 2011.

[58] H. Singh, V. Kumar, and S. Bhooshan. Weighted least squares based detail enhanced exposure fusion. *ISNR Signal Processing*, 2014:Article No. 498762, Feb. 2014.

[59] 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*, 27(2):193–200, Apr. 2008.

[60] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. Probabilistic exposure fusion. *IEEE Transactions on Image Processing*, 21(1):341–357, Jan. 2012.

[61] A. Srikantha and D. Sidibé. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 27(6):650–662, July 2012.

[62] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile HDR video production system. In *Proc. ACM SIGGRAPH*, pages 41:1–41:10, Vancouver, Canada, 2011.

[63] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem. The state of the art in HDR deghosting: A survey and evaluation. *Computer Graphics Forum*, 2015. to appear.

[64] H. R. Wilson. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics*, 38(3):171–178, Oct. 1980.

[65] I. T. Young and L. J. van Vliet. Recursive implementation of the Gaussian filter. *Signal Processing*, 44(2):139–151, June 1995.

[66] X. Zhang, T. Sim, and X. Miao. Enhancing photographs with near infra-red images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Anchorage, AK, June 2008.

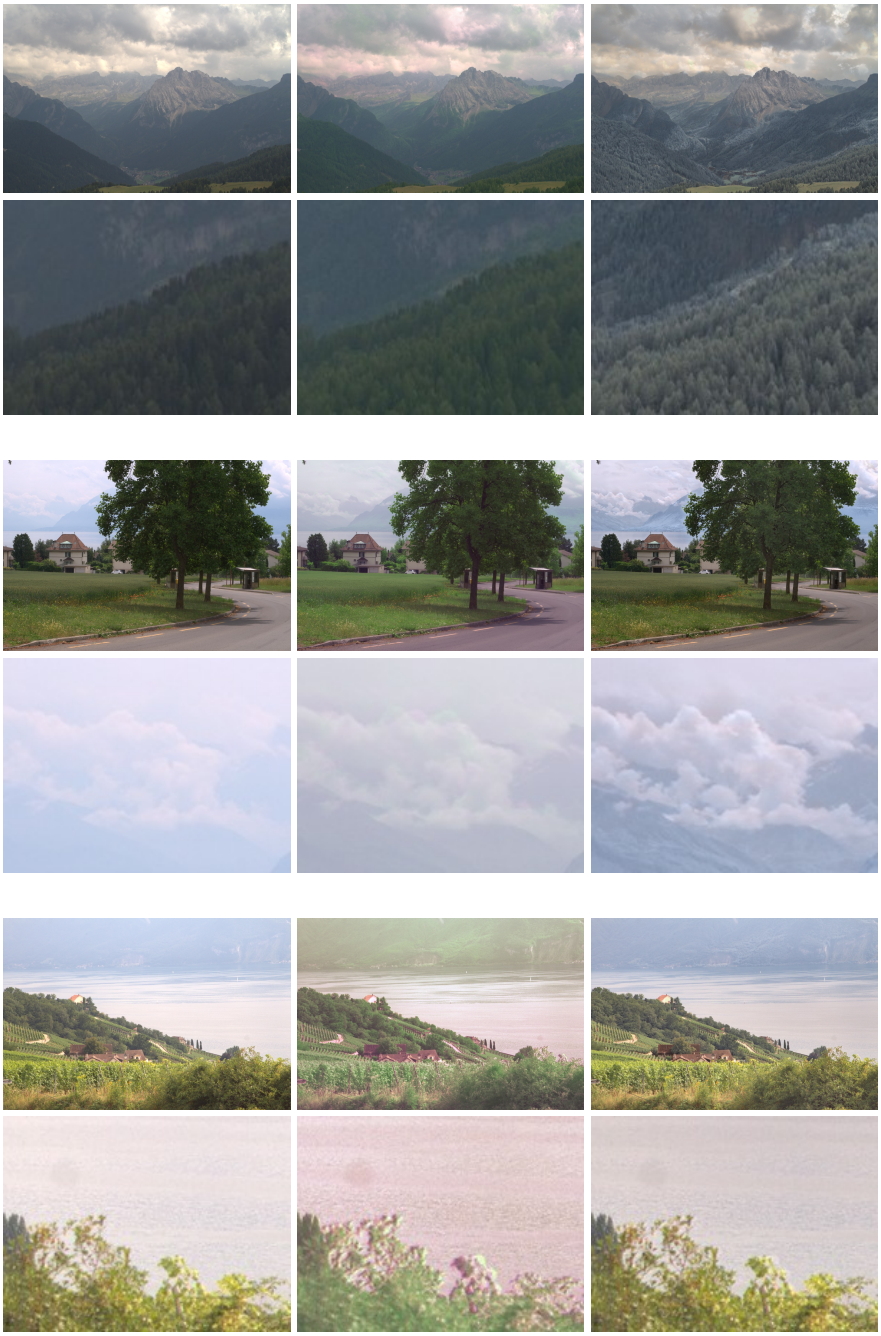


Figure 6: RGB and NIR fusion. Our fused images (*right*) show more details than the standard photographs on the *left*, and also than the resulting images of Eynard et al. [23] in the *middle*. *Odd rows*: Full images. *Even rows*: Zooms. (Input images: [14])

Figure 7: Multispectral imaging. *From left to right*: Visible spectrum, raking light, infrared, ultraviolet fluorescence, and our fused composite. Our fusion scheme condenses the information from all spectral ranges to one visually pleasant composite. (Input images: [4])

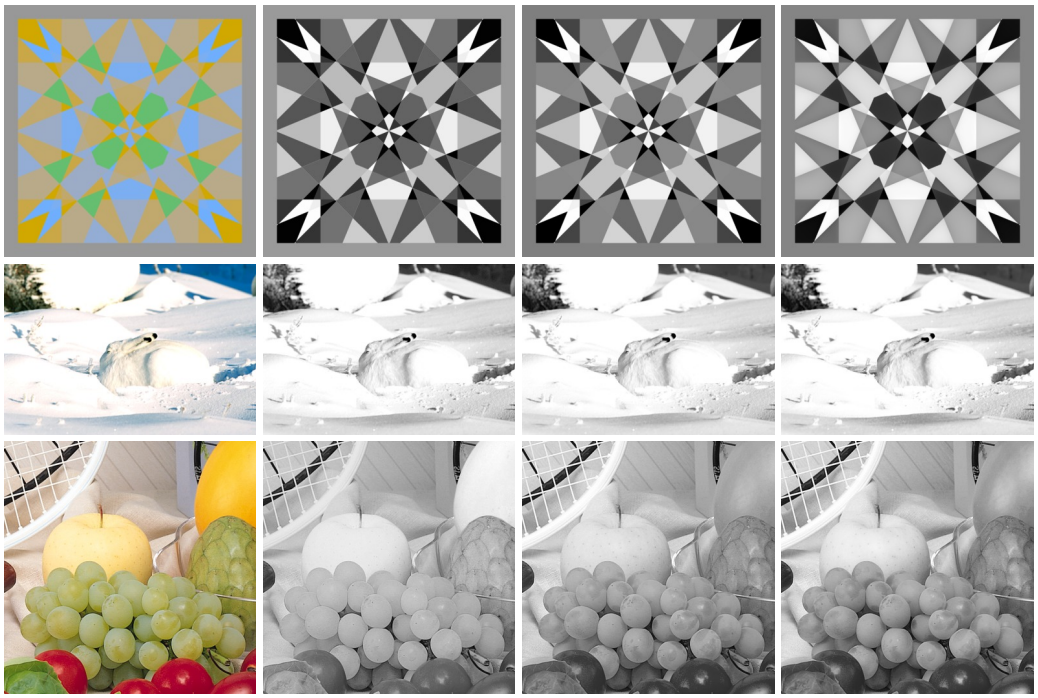


Figure 8: Decolourisation result for the benchmark of [17]. *From left to right*: Input colour image, Lu et al. [39], Eynard et al. [23], and our result.

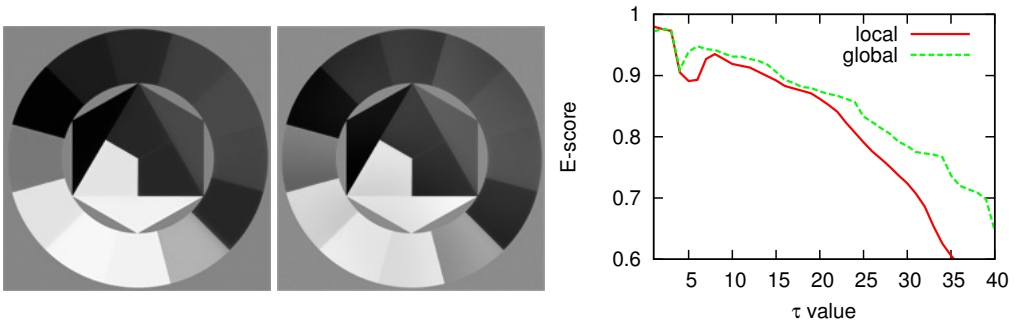


Figure 9: Effect of global (*left*) and local (*middle*) contrast term on the E-score measure of Lu et al. [40] (*right*). Unfortunately, the E-score measure does not account for the local contrast adaption of the visual system. It even penalises it with much smaller E-score values.

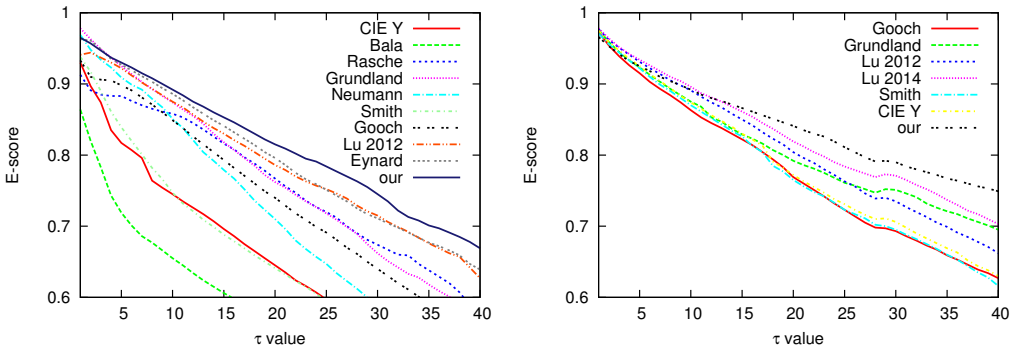


Figure 10: E-score evaluation [40]. The higher, the better. *Left*: Benchmark of Čadík [17]. *Right*: Benchmark of Lu et al. [40].

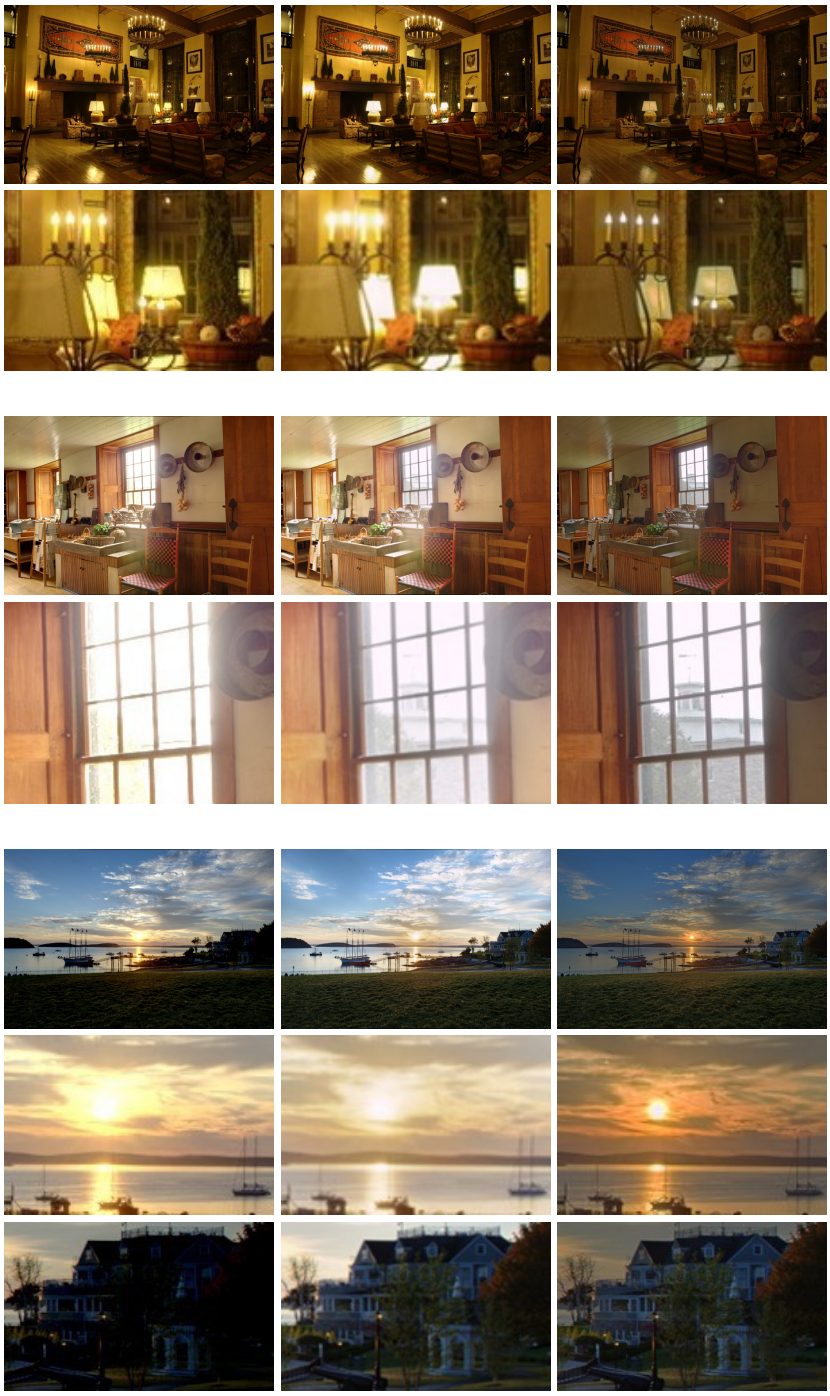


Figure 11: Exposure fusion results (with zooms) for LDR image sets provided by Fairchild [24]. *From left to right*: Result of Mertens et al. [44], Singh et al. [58], and our result. High resolution images and results for further exposure sets can be found in the supplementary material.

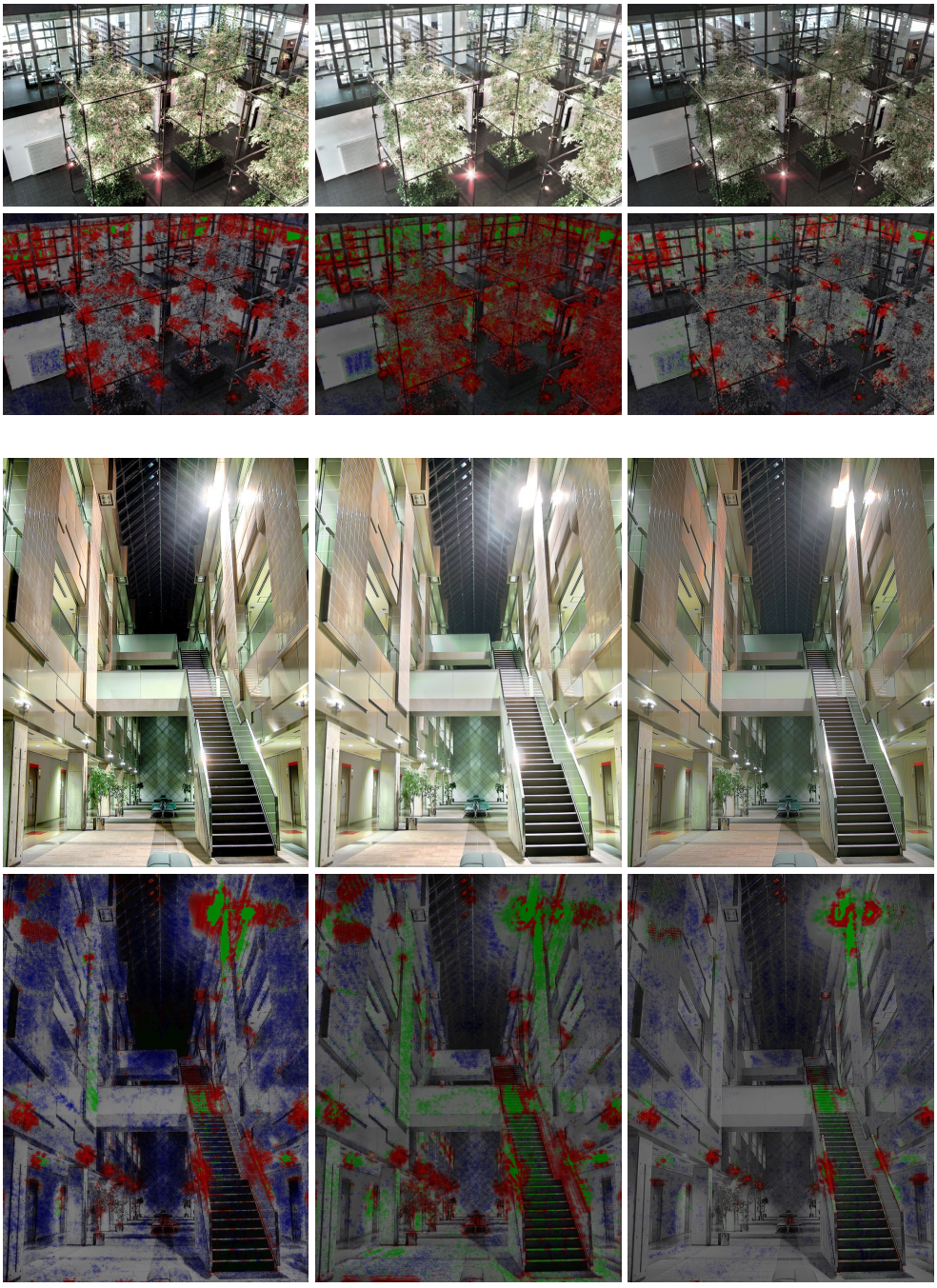


Figure 12: *From left to right*: Exposure fusion of Mertens et al. [44], Singh et al. [58], and our composite image. *Odd rows*: Resulting images. *Even rows*: Corresponding DRIM distortion maps. The more colours, the more visible distortions.

Figure 13: Exposure fusion results of Kotwal and Chaudhuri [36], Mertens et al. [44], Singh et al. [58], and our result. Particularly at the lamps, the better local contrast provided by our method is obvious. This is achieved without producing artefacts; compare for instance the wall in the result of Singh et al. and in the proposed one. (Input images: [3])

Figure 14: Fusion of flash and no-flash photographs. *From left to right*: Ambient, flash, and our fused image.

Figure 15: Fusion of images captured under varying light conditions. *First three columns*: Multilight image collection [26]. *Last column*: Our fused result. It contains details from all input images where the amount of details can be steered with the contrast parameters γ and σ.

Figure 16: Moving objects may lead to ghosting artefacts in the fused composite. *First three*: Input images from [61]. *Right*: Fused image with ghosts.