

How to make a small phone camera shoot like a big DSLR: creating and fusing multi-modal exposure series

Thomas Binder^b and Florian Kriener^b and Christian Wichner^b and Manuel Wille^b and Mike Wellner^c and Thomas Kaester^{a,c} and Thomas Martinetz^a and Erhardt Barth^{a,c}

^aInstitute for Neuro- and Bioinformatics, University of Lübeck, Ratzeburger Allee 160,
D-23538 Lübeck, Germany;

^bNik Software GmbH, Germany;

^cPattern Recognition Company GmbH, Germany.

ABSTRACT

The present work aims at improving the image quality of low-cost cameras based on multiple exposures, machine learning, and a perceptual quality measure. The particular implementation consists of two cameras, one being a high-quality DSLR, the other part of a cell phone. The cameras are connected via USB. Since the system is designed to take many exposures of the same scene, a stable mechanical coupling of the cameras and the use of a tripod are required. Details on the following issues are presented: design aspects of the mechanical coupling of the cameras, camera control via FCam and the Picture Transfer Protocol (PTP), further aspects of the design of the control software, and post processing of the exposures from both cameras. The cell phone images are taken with different exposure times and different focus settings and are simultaneously fused. By using the DSLR image as a reference, the parameters of the fusion scheme are learned from examples and can be used to optimize the design of the cell phone. First results show that the depth of field can be extended, the dynamic range can be improved and the noise can be reduced.

Keywords: computational photography, image fusion, perceptual image quality, optimization

Research is supported by the *German Federal Ministry of Education and Research* within the *KMU Innovativ* program. The grant number is 01IS10009.

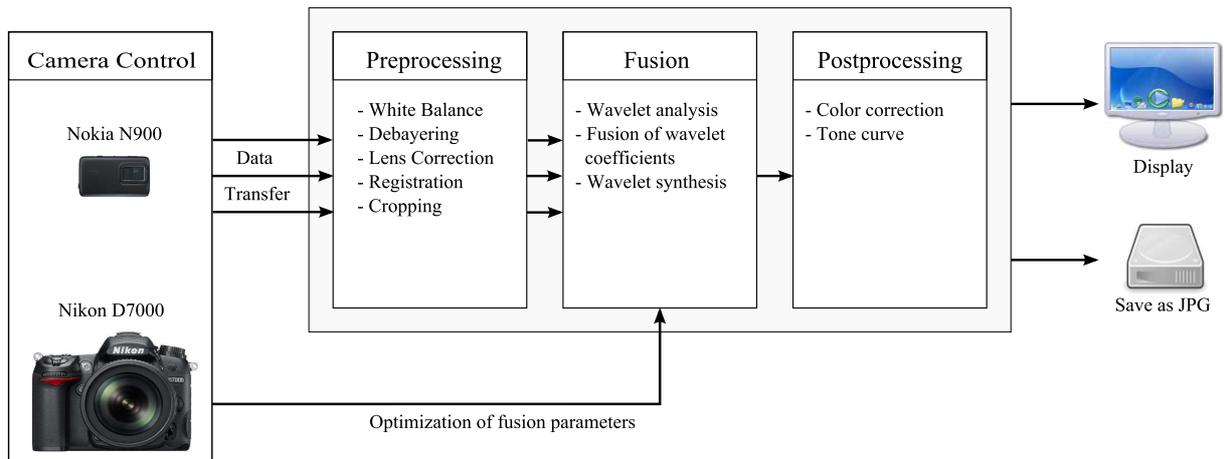
1. INTRODUCTION

Current digital cameras do not fully use their technological potential that would differentiate them from their analog ancestors. For example, fast multiple exposures are possible even with low-cost cameras, and parameters like exposure time and focus can be adjusted quickly. Computational photography is a strong trend that will probably change the way pictures will be taken in the future.¹⁻³ However, the focus of Computational Photography is on topics that require non-standard hardware like for example coded apertures and arrays of micro-lenses that can capture light fields. These computational cameras⁴ are fascinating devices and we are looking forward to working with them in the near future. The focus of our current work, however, is on exploring ideas from computational photography that can be implemented with current off-the-shelf hardware.

The goal is to obtain high-quality images with low-cost hardware by taking a series of images that are fused to a single high-quality image. Image fusion is already in use, for example to expand the dynamic range of an image by fusing images taken with different exposures⁵ or to extend the depth of field by fusing images taken

Further author information: (Send correspondence to Erhardt Barth) Erhardt Barth: barth@inb.uni-luebeck.de
Cite as: T. Binder, F. Kriener, C. Wichner, M. Wille, M. Wellner, T. Kaester, T. Martinetz, E. Barth, "How to make a small phone camera shoot like a big DSLR: creating and fusing multi-modal exposure series," Proceedings SPIE Vol. 8291 (Human Vision and Electronic Imaging XVII), 829106 (2012). DOI: <http://dx.doi.org/10.1117/12.905560>
Copyright 2012 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Figure 1. Top-level overview of the suggested workflow. Once the fusion parameters have been optimized and fixed, the DSLR is no longer required.



with different focus settings.⁶ It has also been shown that, at equal depth of field, the signal-to-noise ratio of the image can be improved by fusing images taken with a large aperture and different focus settings instead of taking just one image with a smaller aperture.⁷ Our approach extends previous work in several ways. First, we vary not only one parameter when taking a series of images: currently focus and exposure time are varied. Second, we jointly optimize the fusion parameters. Third, we use a reference image for optimization, i.e., we apply supervised machine-learning techniques.

To add practical value, we implemented a system with a low-cost camera that must be optimized and a high-end camera that serves as a reference. The first camera is a Nokia N900 mobile phone, which benefits from the development of the Frankencamera.² The second is a Nikon D7000, which has much better dynamic range and signal-to-noise ratio, and offers a much more flexible depth of field. In this paper we describe the system and the fusion algorithms, and present first results.

Note that the idea of optimizing low cost imaging hardware with respect to image quality by the use of a high quality reference system poses new challenges regarding the modeling of subjective image quality. Currently we only consider this problem by using the structural similarity measure (see below) but we believe that models of image quality need to be further developed for these kinds of applications.

2. SYSTEM ARCHITECTURE, PRE- AND POSTPROCESSING

While the fusion of exposures in the wavelet domain is the key idea, technical effort has to be made in order to try out the outlined ideas. This section describes a system, consisting of hard- and software, designed to provide aligned exposures resp. exposure series with varied parameters.

2.1 Mechanical setup

The initial idea requires a facility to shoot exposures of the same scene simultaneously with two different cameras, which by itself is an easy task. However, it becomes incomparably harder when requiring a correspondence of pixels in all exposures, such that two corresponding pixels in different exposures belong to the same spatial position in the scene. This correspondence requirement forced us to abandon all approaches for a mechanical setup in which the cameras are mounted against each other with a parallax, such as attaching the cell phone to the hot shoe or the tripod socket of the DLSR. In addition, the hot shoe mounting variant usually involves having some play, which makes it unusable in the present setting.

Figure 3. Camera hardware and mounting used to produce image material. The accessory arm is turned by 180 degrees after the first camera has completed shooting. The second camera can start shooting once it has taken the position of the first camera. Bottom to top, the setup consists of: a sturdy tripod, a panorama head, an accessory arm with two heads, a micropositioning plate for the DSLR, a cell phone car mount and standard small ironware for mounting the cell phone on the accessory arm. The panorama head clicks into place at the opposite positions. All non-camera parts were chosen based on their ruggedness and high load capacity. For example, the tripod is designed to carry weights up to 26 1/2 lbs (12kg).



The no parallax requirement implies that the second camera must take the position of the first when shooting. A practical solution offering high precision is shown in Figure 3. By rotating the accessory arm by 180 degrees one camera can take the position of the other. In the case of ideal camera adjustment the entry pupils of both cameras are matched by the rotation.

2.2 Camera hardware

For the prototype we had to choose particular makes of DSLR and cell phone. The existence of the FCam API² made us pick the Nokia N900 as low-cost camera. The N900 is one of Nokia's 2009/2010 top-of-the-line smart phones. Its operating system, Maemo, is from the Linux family. The FCam API offers an easy and flexible way to control the camera ingredients of the N900. The Nikon D7000 was chosen as a current upper middle class DSLR make; it offers DX format and good interoperability with the Picture Transfer Protocol (PTP).

Both cameras are able to produce RAW files and were configured to do so. The D7000 produces NEF files (Nikon's proprietary RAW format), while the FCam API allows to save the sensor contents of the N900 as DNG (Adobe's non-proprietary RAW format). In the setting of this work, shooting RAW is essential since it offers the greatest flexibility for any kind of processing: the sensor data is roughly proportional to the amount of incident light; it is linear data. The linearity is preserved during the preprocessing step such that the fusion step can rely on it. For this reason we were happy about the DNG feature of the FCam API.

The camera of the N900 is limited to changes of exposure time, focus, and ISO sensitivity; there is no way to change zoom or aperture. Therefore we restrict ourselves to changes of exposure time and focus. The effect of focus changes is pronounced and obvious for small distance or near focus, respectively. For larger distances, the N900 depth of field is

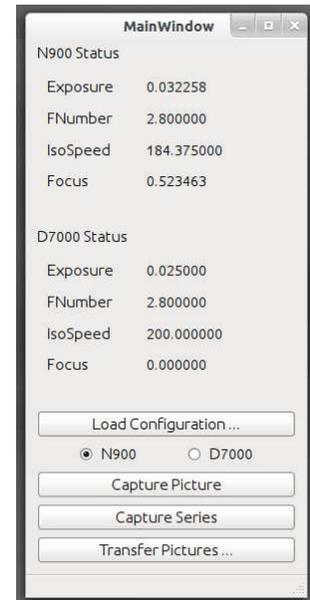


Figure 2. Exemplary view of the UI of the client running on a desktop machine. The client controls the N900 server. The window displays the current exposure parameters for both N900 and D7000, and has buttons to initiate shooting and downloading. Independently for both cameras the exposure count and parameters for each exposure are given as so-called configurations (XML).

so large such that these changes are difficult to spot. Therefore, we want to explore the near range in particular, and the DSLR lens was chosen appropriately. In addition, the lens must deliver images with a similar angle of view as the cell phone. A fast, wide-angle lens fulfills these requirements; we used a 20mm/f2.8 lens.

2.3 Capturing software

Due to the possibility of accidental shaking and movements, touching the tripod with the mounted camera equipment during the shooting process is risky and must be avoided. On the other hand, parameters must be changed between exposures. It is therefore suggesting to use tethered shooting and tethered parameter changes for both the N900 and the D7000. Both are connected via a USB 2.0 cable which is visible in Figure 3. In order to make this work and to let the N900 act as a USB master, both the N900's operating system and libusb had to be patched.

A server software was created for the N900, which is in charge of (a) controlling the N900's camera via FCam, (b) controlling the D7000 via PTP, and (c) communication with a client software running on a desktop machine. Although controlling the server via the N900's touch screen is conceivable, the operator interacts with the server only through the client software to avoid shaking. Figure 2 shows the main windows of the client software. The communication between server and client is done via Wireless LAN. The sustainable rate for continuous shooting is approximately two seconds per exposure, where we have taken up to 500 exposures in a single run. In order to achieve said rate, we had to optimize memory strategies of the server software.

2.4 Pre- and postprocessing

As both cameras are configured to produce RAW files but we want to look at and evaluate the result of the fusion, full raw development must take place. Apart from certain cropping and calibration steps the fusion module was integrated in the raw development pipeline as shown in Figure 1.

2.4.1 Preprocessing related to raw development

The parts of the preprocessing step related to raw development include adjusting white balance, debayering, and lens correction, cf. Figure 1. White balance, the scaling of the red and blue channels with respect to the green channel, is determined by two constants. We use the as-shot white balance stored in the RAW files. For both NEF and DNG file formats care must be taken to extract the correct white balance coefficients from the metadata.

Since the sensor data carries the Bayer pattern, debayering must take place to produce linear RGB data. Many debayering algorithms are available and offer sufficient quality. The algorithm due to Hirakawa and Parks is one such example.⁸ Differences between various debayering algorithms can be told only from fine detail in sharp images; in other areas such algorithms will produce more or less the same results.

The lens correction applied corrects distortion only; we do not use it to correct vignetting and chromatic aberration. The distortion correction is responsible for re-establishing lines in the images, i.e. straight lines in the scene should be reproduced as straight lines in the image after the lens correction step. After accurate lens correction it can be expected that a line-preserving transformation with a larger number of free parameters, such as a perspective transformation, is suitable for aligning the images. We used the lens correction algorithms contained in OpenCV.^{9,10}

2.4.2 Preprocessing due to many exposures

Spatial standardization of the exposures is required; therefore we perform camera calibration, alignment, and cropping. To fulfill the correspondence requirement, all clamps and screws are tightened, and the cameras are calibrated mechanically against each other. Moreover, in order to make up for image differences caused by different focus distances, all N900 exposures are calibrated against each other. To this end we use the calibration of OpenCV once more. As with the lens correction, printouts of a checkerboard are photographed and processed; the result of the calibration is a perspective transformation which transforms the 2D exposure coordinate systems

into each other. The transformation includes a central shrinking of about 1/2 of the D7000 exposure which takes into account the different dimensions of DSLR and cell phone exposures.

To further standardize the image material, any exposure is cropped from 2584x1960 or 4928x3264 (for N900 and D7000, respectively) to 1024x1024. First, the rectangular shape of each exposure following the transformation of the calibration steps must be re-established. Second, since the N900 produces heavy vignetting which remains unchanged by the lens correction step, a desired side effect of the cropping is that the vignette is more or less removed. Finally, exposures from both cameras have different aspect ratios; which is neutralized by the cropping step.

2.4.3 Postprocessing

The main purpose of the postprocessing step is to transform the RGB data from the linear domain to a domain which is visually pleasing to the human eye. This is done by applying a color transformation and a tone curve. The resulting RGB data is assumed to be in a certain output color space, for example sRGB. Of course the color transformation depends on the choice of output color space. Since currently we restrict to black and white images the color transformation is of no importance.

3. IMAGE FUSION AND OPTIMIZATION OF PARAMETERS

Computing a single improved image from a series of exposures with lower quality is a task which is well-known from focus stacking or high dynamic range (HDR) imaging. In the following we will describe our new fusion framework, which is able to perform both tasks (and more) at once. Since even with low parameter counts manual parameter tuning is not an option, we perform automatic parameter optimization by means of a reference-based approach using the acquisition system described in Section 2.

3.1 Image representation

In image processing it is often advantageous to transform an image to a new basis. In this work we chose the 2D dual-tree discrete wavelet transform (DDWT), denoted by W , which is due to Selesnick et al.¹¹ The DDWT is two times expansive; we use the filtering coefficients introduced by Farras, Selesnick and Kingsbury.^{12,13} Moreover, the DDWT is insensitive to shifts, directionally selective in two and higher dimensions and was already applied successfully for focus stacking.⁶

Suppose the exposure stack from the cell phone is given by I_n where $n = 1, \dots, N$. Let I_R denote the reference exposure from the DSLR. The wavelet transform W maps any image I to $(a_p)_{p \in P} \in \mathbb{C}^{\#P}$, which is a vector of dimension $\#P$ of complex wavelet coefficients. We assume that the number M of resolutions of any wavelet transform is the same; we use $m = 1, \dots, M$ to denote available resolutions. The band-pass wavelet coefficients within each scale are enumerated by the sets

$$Q_m = \{m\} \times \{1, 2, 3\} \times S_m \subseteq \mathbb{N}^4, \quad m = 1, \dots, M,$$

where $k = 1, 2, 3$ denotes one of three directional subbands and S_m is the set of all spatial positions (u, v) within subband k of resolution m . After preprocessing we may assume that the dimensions of all exposures coincide such that also the dimension of the wavelet coefficient space is the same for all transforms. Finally, the set $Q = \cup_{m=1}^M Q_m$ indexes all wavelet high-pass coefficients. Joining $P = Q \cup R$ we get the whole wavelet domain, where $R = S_M$ represents the DC component of the wavelet transform.

3.2 Fusion algorithms

During image fusion the images of an image stack have to be combined in that way, that the resulting image reaches application specific quality requirements. In HDR imaging we are looking originally for images with optimal dynamic range between lightest and darkest areas preserving all the relevant image details. Focus stacking is aiming at enhancing the depth of field to visualize all relevant information in one image, which cannot be achieved by means of a digital camera and a certain lens.⁶

Our fusion scheme works as follows: In a first step, each cell phone image is converted to the wavelet domain: $a^{(n)} := W(I_n)$, the fusion then produces a single set of wavelet coefficients by $b = F(a^{(1)}, \dots, a^{(N)})$, where F is one of the fusion rules defined in what follows. The reconstructed image J — the result of the fusion — is then obtained by $J := W^{-1}b$, wherein W^{-1} denotes the inverse wavelet transform.

To define a fusion variant, values b_p have to be prescribed for all $p \in P = Q \cup R$. For $r \in R$ we invariably use averaging: $b_r = \frac{1}{N} \sum_{n=1}^N a_r^{(n)}$. To fully define a fusion variant, it therefore suffices to prescribe b_q for $q \in Q$.

An example of a simple fusion is the maximum rule. For fixed $q \in Q$ it is given by

$$b_q = a_q^{(l)}, \quad \text{where } l := \operatorname{argmax} \left\{ |a_q^{(n)}| \mid n = 1, \dots, N \right\}. \quad (1)$$

The maximum rule works well with focus stacks; its goal is to create a depth of field which is as large as possible. For arbitrary $z \in \mathbb{C}$ we define hard thresholding independently for each resolution $m = 1, \dots, M$ by

$$T_m(z) := \begin{cases} z, & \text{if } |z| > \epsilon_m, \\ 0, & \text{otherwise,} \end{cases}$$

where $\epsilon_m \geq 0$ are suitably chosen real numbers. Then a second fusion example, governed by the mean, is given by

$$b_{mkuv} = \frac{1}{w_m} \sum_{n=1}^N T_m(a_{mkuv}^{(n)}), \quad \text{where } w_m := \# \bigcup_{n=1}^N \left\{ q \in Q_m \mid T_m(a_q^{(n)}) \neq 0 \right\}. \quad (2)$$

Differently set, the fusion result is the mean of wavelet coefficients not touched by the thresholding. Although simple fusion rules like (1) or (2) produce satisfactory result, they are likely to fail if several camera parameters are varied during image acquisition. Therefore, let us consider a further fusion variant given by

$$b_{mkuv} = \sum_{n=1}^N \frac{e^{-\beta_m |a_{mkuv}^{(n)}|}}{\sum_{\tilde{n}=1}^N e^{-\beta_m |a_{mkuv}^{(\tilde{n})}|}} T_m(a_{mkuv}^{(n)}), \quad (3)$$

where $\beta_m \in \mathbb{R}$, $m = 1, \dots, M$ are the parameters to be optimized. Formula (3) was chosen since it parametrizes the transition between the fusion rules (1) for $\beta_m \ll 0$, and (2) for $\beta_m = 0$. Again, the advantage of (3) is that its derivative with respect to β_m can be given in a closed form. Note also that (3) is a pointwise operation; it does not involve any neighborhoods.

3.3 Parameter Optimization

Due to the aforementioned disadvantages of (1) and (2) we turn to (3) and use it to model the fusion b of exposures where both the focus and the shutter speed are varied simultaneously. Since manual parameter tuning is inefficient, we need to be able to measure distances between images, more precisely, between the result of the fusion and the reference exposure. Since fusion is conducted in the wavelet domain, for performance reasons it is suggesting to have an image similarity measure defined in the wavelet domain as well. We introduce two different similarity measures.

The first is the usual vector space norm, applied to a fixed resolution m . On this complex subspace $\langle a, b \rangle_m := \sum_{q \in Q_m} a_q \overline{b_q}$ and $\|a\|_m := \sqrt{|\langle a, a \rangle_m|}$ denote the associated scalar product and induced norm, respectively. The wavelet coefficient space is a complex $\#Q$ -dimensional vector space with scalar product $\langle a, b \rangle := \sum_{q \in Q} a_q \overline{b_q}$ and induced norm $\|a\| := \sqrt{|\langle a, a \rangle|}$, where $\|a\|^2 = \sum_{m=1}^M \|a\|_m^2$ by Parseval's identity. Then we can perform a 1D search to minimize

$$E_m(\beta_m) = \|a^R - b\|_m^2 = \sum_{k=1}^3 \sum_{(u,v) \in S_m} |a_{mkuv}^R - b_{mkuv}|^2$$

for each $m = 1, \dots, M$; this minimizes also

$$E(\beta_1, \dots, \beta_M) = \|a^R - b\|^2 = E_1(\beta_1) + \dots + E_M(\beta_M)$$

in M dimensions. Note that minimizing E is equivalent to minimizing the mean squared error of the wavelet high-pass coefficients. Recall that b is defined as in (3).

The second approach is to use the structural similarity (SSI) due to Wang and Simoncelli,^{14,15} applied in the wavelet coefficient space:

$$C(\beta_1, \dots, \beta_M) = \text{CWSSIM}(a^R, b) = \frac{2|\langle a^R, b \rangle| + \kappa}{\|a^R\|^2 + \|b\|^2 + \kappa}. \quad (4)$$

Here $\kappa > 0$ is a scalar that stabilizes the result of the division when the denominator is small. Again, since we prefer to optimize in 1D only, maximizing

$$C_m(\beta_m) = \text{CWSSIM}_m(a^R, b) = \frac{2|\langle a^R, b \rangle_m| + \kappa}{\|a^R\|_m^2 + \|b\|_m^2 + \kappa}$$

for each $m = 1, \dots, M$ gives optimized values for β_m . Also in this formula b is given by (3). Note however that this time solving the 1D problem M times is not equivalent to maximizing (4).

4. RESULTS

In the following we present first results obtained with our fusion scheme. We are focussing on the following two questions: (i) does fusion by means of wavelet coefficients and (3) work when both exposure and focus are varied, and (ii) can parameter optimization be further improved using a perceptual quality measure instead of the traditional mean squared error?

The results have been obtained with a resolution count of $M = 3$ for the wavelet transform. Therefore, since the input images have dimension 1024x1024, the remaining DC component has dimension 128x128. To parametrize hard thresholding by T_m , histograms of absolute values of wavelet coefficients of average images were analyzed. A suitable ϵ_m was obtained by taking the 70–75th percentile values of such per-resolution histograms. Images containing much detail result in higher values of ϵ_m , while images with a lot of homogeneous areas give lower values.

To illustrate our fusion scheme we set up a table-top scene, took one reference image and four low-quality input images with two different values of exposure and focus, see Figure 4. Note that the fused image and the reference image are quite similar and that the quality of the fused image is comparable to that of the reference image and better than that of the input images. Also note that the depth of field and the dynamic range are both larger in the fused image than in the single input images. Moreover, the noise level is reduced. For better illustration, in Figure 5 we show details cropped from the images in Figure 4. Note that, in some aspects, the fused image is even better than the reference image, for example the toy pupil differentiates better from the iris.

However, in Figure 5 we illustrate two further important results. First, we show that the fusion parameters obtained with the CWSSIM measure yield better results than those obtained with the mean squared error (compare Figs. 5(f) and (g)). This is due to the fact that CWSSIM yields higher values of β_m thus enhancing the adaptive effect described below.

Second, we show that the fusion scheme is locally adaptive (Figure 5(e)). With the *fusion entropy* we characterize the distribution of the fusion weights $\exp(-\beta_m |a_{mkuv}^{(n)}|) / \sum_{\tilde{n}=1}^N \exp(-\beta_m |a_{mkuv}^{(\tilde{n})}|)$ in (3). Maximum entropy then corresponds to an equal distribution and this, in turn, to a fusion scheme in which the coefficients are averaged as in (2). In case of low entropy, the larger coefficients are weighted stronger, which, in the extreme case, corresponds to a maximum operation as in (1). Therefore, our scheme defined by (3) performs an adaptive fusion such that uniform regions are averaged and structured regions are dominated by the most prominent structure. This seems to work well for focus stacking but also implements a noise reduction.



(a) I_1 dark, focus far



(b) I_2 dark, focus near



(c) I_3 bright, focus far



(d) I_4 bright, focus near



(e) Fused image J , CWSSIM



(f) Reference image I_R

Figure 4. Results obtained for an image stack of $N = 4$ images, taken with two different exposure times and two different focus settings for each exposure time. Note the improved image quality and the extended depth of focus in the fused image

5. DISCUSSION

We have presented a method that can be used to improve the image quality of low-cost cameras. The current hardware consists of a cell phone camera and a high quality digital camera, a DSLR. The cell phone can shoot an image stack with different exposure values and focus settings and the DSLR takes a reference image of the same scene. The image stack is fused with a novel fusion scheme that is locally adaptive and has parameters that are optimized by using the reference image. The fusion is performed in the wavelet domain and the cost function for the optimizer is based on the difference between the wavelet coefficients of the reference image and those of the fused image. The difference is either the mean squared error or the perceptually motivated CWSSIM measure.

We did not present a systematic evaluation of the method but showed some interesting results. First, the image quality of the fused image is much better than the quality of the single images: the depth of focus is extended, the dynamic range is increased, and the noise level is lower. Second, the results are better when using the perceptual CWSSIM measure.

Open questions remain, for example we are still working on a general solution for the problem of registering the images within the stack and with respect to the reference image. This problem is encountered in many other tasks and we have not elaborated on it here. However, we consider the approach presented in this paper as first step towards more complex fusion schemes that can provide further image enhancements based on, for example,

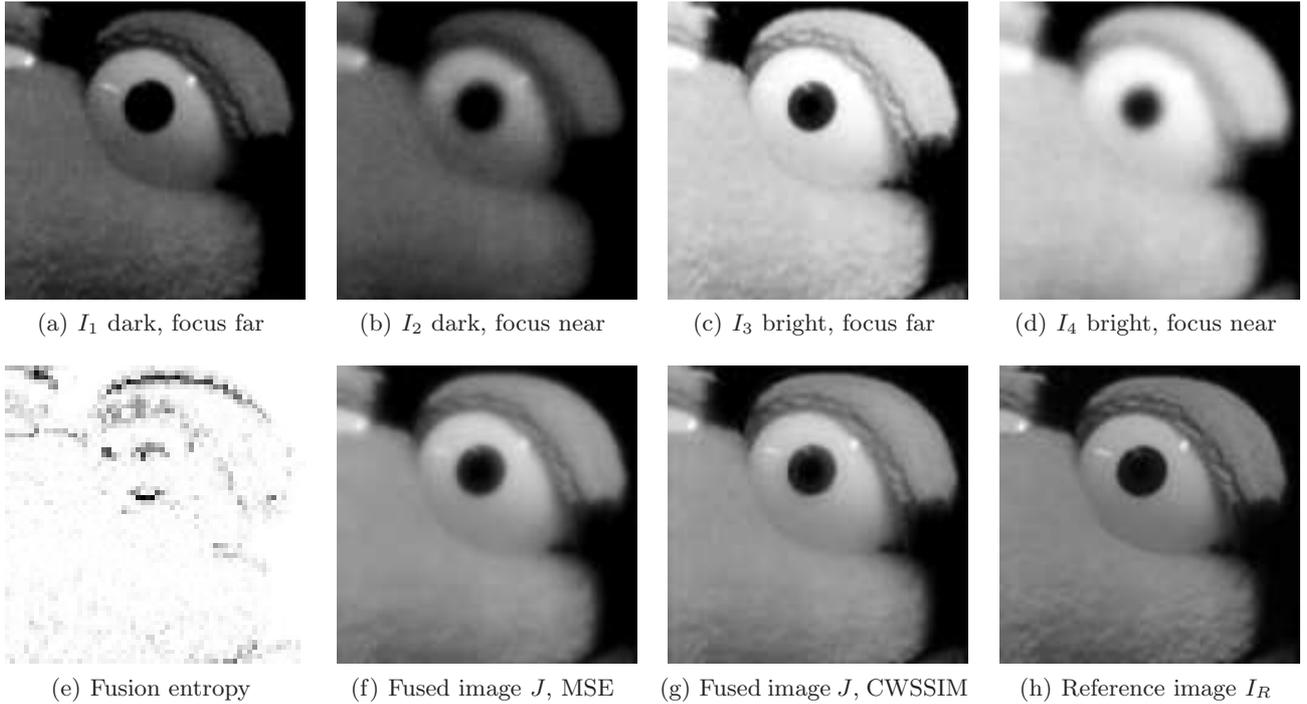


Figure 5. Same results as in Figure 4 shown for a detail cropped from the images in Figure 4. In addition, in (e) we show the entropy of the fused coefficients (for the CWSSIM case). Note that our fusion algorithm tends to pick the max rule at edges and the mean rule at uniform regions (MSE case is similar but less pronounced). In (f) we show the result obtained for the same stack, which is here fused with parameters learned by optimizing the mean squared error instead of the CWSSIM.

spatial filtering and local gain control. Also, the way the image stack is obtained can be extended, for example by using a larger number of different exposures or by varying other parameters like the gain. In a next step, not only the fusion parameters but also the imaging parameters will be optimized.

The benefits of a method for optimizing the results obtained by low-cost hardware by using high-end hardware as a reference seems obvious. But the method could be used for other applications, for example to create images that just look like some reference images. By exploring various fusion schemes and parameter values, images with a novel look could also be obtained. In such cases one would like to use no-reference models of perceptual image quality. However, while models that can predict the visibility of artifacts are quite mature, the theory of what makes a cool image is still missing.

REFERENCES

- [1] Nayar, S. K., “Computational cameras: redefining the image,” *Computer* **39**, 30–38 (2006).
- [2] Adams, A., Talvala, E.-V., Park, S. H., Jacobs, D. E., Ajdin, B., Gelfand, N., Dolson, J., Vaquero, D., Baek, J., Tico, M., Lensch, H. P., Matusik, W., Pulli, K., Horowitz, M., and Levoy, M., “The Frankencamera: an experimental platform for computational photography,” in [*ACM SIGGRAPH 2010 Papers (Los Angeles, California, July 26–30, 2010)*], 1–12 (2010).
- [3] Raskar, R. and Tumblin, J., [*Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors*], A. K. Peters, Ltd., Natick, MA, USA (2009).
- [4] Zhou, C. and Nayar, S., “Computational cameras: convergence of optics and processing,” *IEEE Transactions on Image Processing* **20**(12), 3322–33409 (2011).

- [5] Debevec, P. E. and Malik, J., “Recovering high dynamic range radiance maps from photographs,” in [*SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*], 369–378 (1997).
- [6] Forster, B., Van De Ville, D., Berent, J., Sage, D., and Unser, M., “Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images,” *Microscopy Research and Technique* **65**, 33–42 (2004).
- [7] Hasinoff, S. W., Kutulakos, K. N., Durand, F., and Freeman, W. T., “Time-constrained photography,” in [*Proc. 12th IEEE International Conference on Computer Vision, ICCV*], 333–340 (2009).
- [8] Hirakawa, K. and Parks, T. W., “Adaptive homogeneity-directed demosaicing algorithm,” *IEEE Transactions on Image Processing* **14**(3), 360–369 (2005).
- [9] Bouguet, J.-Y., “Camera calibration toolbox for MATLAB,” see http://www.vision.caltech.edu/bouguetj/calib_doc (2010).
- [10] Zhang, Z., “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334 (2000).
- [11] Selesnick, I. W., Baraniuk, R., and Kingsbury, N., “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine* **22**(6), 123–151 (2005).
- [12] Abdelnour, A. F. and Selesnick, I. W., “Nearly symmetric orthogonal wavelet bases,” in [*Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*], (May 2001).
- [13] Kingsbury, N., “Image processing with complex wavelets,” *Phil. Trans. R. Soc. Lond. A* **357**(1760), 2543–2560 (1999).
- [14] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).
- [15] Wang, Z. and Simoncelli, E. P., “Translation insensitive image similarity in complex wavelet domain,” in [*Proc. IEEE Inter. Conf. Acoustic, Speech & Signal Processing Volume II*], 573–576 (2005).