HAL
open science

Robust Joint Image Reconstruction from Color and Monochrome Cameras

Muxingzi Li, Peihan Tu, Wolfgang Heidrich

To cite this version:

Muxingzi Li, Peihan Tu, Wolfgang Heidrich. Robust Joint Image Reconstruction from Color and Monochrome Cameras. BMVC 2019 - 30th British Machine Vision Conference, Sep 2019, Cardiff, United Kingdom. hal-02370233

HAL Id: hal-02370233

https://inria.hal.science/hal-02370233v1

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust Joint Image Reconstruction from Color and Monochrome Cameras

Muxingzi Li*
muxingzi.li@inria.fr
Inria Sophia Antipolis - Méditerranée
Valbonne, France

Peihan Tu*
phtu@umd.edu
University of Maryland
College Park, United States

Wolfgang Heidrich
wolfgang.heidrich@kaust.edu.sa
King Abdullah University of Science and Technology (KAUST)
Thuwal, Saudi Arabia

Abstract

Recent years have seen an explosion of the number of camera modules integrated into individual consumer mobile devices, including configurations that contain multiple different types of image sensors. One popular configuration is to combine an RGB camera for color imaging with a monochrome camera that has improved performance in low-light settings, as well as some sensitivity in the infrared. In this work we introduce a method to combine simultaneously captured images from such a two-camera stereo system to generate a high-quality, noise reduced color image. To do so, pixel-to-pixel alignment has to be constructed between the two captured monochrome and color images, which however, is prone to artifacts due to parallax. The joint image reconstruction is made robust by introducing a novel artifact-robust optimization formulation. We provide extensive experimental results based on the two-camera configuration of a commercially available cell phone.

1 Introduction

With the ever decreasing cost of camera modules, we see more and more consumer devices with multiple cameras, including configurations where multiple cameras point into the same direction. Although systems like the L16 camera from Light[1] are not yet widely adopted, systems with two front facing cameras are already quite common. This includes systems such as the Huawei P9 phone, which contains two rectified cameras with different sensor types, namely an RGB camera for color imaging, and a black-and-white (monochrome) camera for improved low-light performance as well as extra sensitivity in the UV and IR part of the spectrum. Figure 1 shows the relative image quality of the color and monochrome cameras of the P9 phone under low-light situations, for the same exposure time. This difference in quality is explained by the drastically improved sensitivity of the monochrome sensor for wavelengths from the UV to the near IR spectrum, as indicated by the response curves of the color and monochrome version of the SONY IMX287 sensor [1].

© 2019. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

*Both authors contributed equally. The work was done while the authors were at KAUST.

[1] https://light.co

We present a joint demoisaicking and denoising approach specially designed to improve images from the color camera utilizing the light-efficient monochrome camera. This requires two steps: aligning the image pair, and merging photometric details while simultaneously suppressing noise. Aligning a monochrome and a color image is a challenging problem. Even state-of-the-art alignment methods are artifact-prone. Alignment is complicated by three facts. First, our algorithm targets low-light situations. The captured images, especially color images, are noisy, which compromises image features. Second, while the spacing between the two cameras is small, the parallax is not negligible for closeby objects, and does result in depth-dependent disparities, with all the well-known consequences, such as occlusion regions etc. Third, because of different photometric response of the two sensors, textures may occur in only one of the two cameras but not in the other. For example, different RGB colors may map to the same grayscale value, resulting in a loss of detail in the monochrome image. On the other hand, the monochrome camera may be able to see additional texture in the UV and IR part of the spectrum, and small-scale texture may simply fall below the noise threshold in the RGB image. This difference in photometric characteristics also presents a challenge for the final merging of the image pair.

Figure 2 shows an outline of our pipeline. The alignment step (flow estimation) exploits a state-of-the art optical flow approach [32] which allows for displacements in both axes, instead of assuming a perfectly rectified stereo pair, as the latter can be hard to maintain in mobile devices over extended periods. Reliable flow is detected and saved in a weight matrix, which is fed to the optimization step. We propose an image formation model that integrates a reference channel into joint demosaicking and denoising of CFA images. By solving with the proximal algorithm for a piecewise smooth relation between color channels and the monochrome reference channel, we are able to transfer structural details across the image pair while preserving the underlying color.

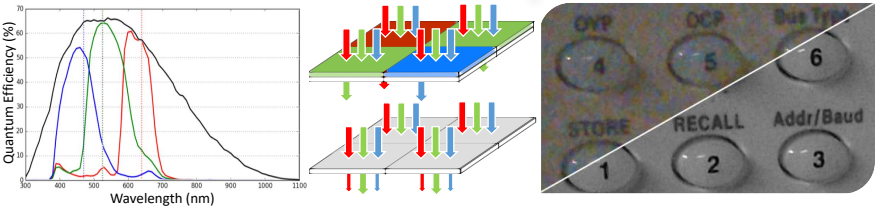

Figure 1: Due to the use of color filter arrays, RGB sensors have significantly reduced light sensitivity compared to monochromatic cameras with the same sensor design. Same scene shot by color camera and monochrome cameras with the same exposure time exhibits different levels of noise and sharpness.

2 Related Work

Single-image denoising and demosaicking. BM3D [7] has been the state-of-the-art denoising algorithm for a decade and continues to outperform learning-based methods on real photographs [31]. Representative works on Bayer image demosaicking include residual interpolation [18], adaptive homogeneity-directed interpolation [14], self-similarity demosaicking [4] etc. Joint demosaicking and denoising has also been studied [3, 12, 19], including its specific application to low-light images [5].

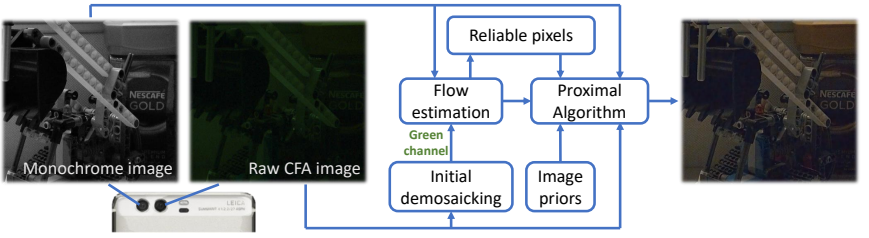

Figure 2: Our proposed pipeline takes a pair of one color filter array (CFA) image and one monochrome image, aligns them, detects visual artifacts in the warped monochrome image and outputs a jointly demosaicked and denoised color image.

Alternative sensor designs. Low-light imaging can be improved with novel sensor designs (i.e. alternatives to the Bayer filter), for instance, RGB-IR [35, 37] and RGBW sensors [16]. However, RGB-IR sensors have issues with color fidelity: most color cameras use quite thick glass filters to eliminate IR light across the whole sensor. Bayer-style per-pixel IR filters are much less effective, and contaminate the RGB image with leftover IR components. For the same reason, RGBW sensors (Sony, Canon) require a global IR filter, therefore lose IR information for low light sensitivity. Also it is hard to find good exposure settings for RGBW sensors without either RGB being underexposed or W being overexposed. Both sensor patterns suffer from spatial resolution degradation owing to the sensor composition having even more color components than the Bayer pattern. Our method avoids these degradations by using two cameras.

Alignment-free multi-image processing. Several works produce images with reduced noise and rich details while preserving color fidelity [2, 9, 10, 30, 38], by merging images taken under different lighting conditions or camera settings. Zhang et al [39] and Luo et al [28] formulate a statistical model to perform non-local denoising for noisy multi-view images, similar in spirit to video denoising [22]. However, these two approaches do not provide precise pixel-to-pixel alignment and are unable to transfer details from one image to another. Zhuo et al [40] propose to use near-infrared images to enhance noisy low-light images with a complex, hybrid two-camera imaging system, which is similar to our goal. They bypass the alignment problem with a hardware solution (aligning the optical axis of the two cameras using a beam splitter, which further reduces light efficiency and are hard to implement in a compact form, e.g. in cell phones), whereas our approach works simply with stereo cameras.

Alignment-guided multi-image processing. Single-chip array cameras provide thin designs with per-lens color filters. Registration of sub-images of different modalities are needed, which is a challenging task when parallax and occlusions arise. Heide et al [12] optimize camera array images within a proximal framework using optical flow and cross-channel correlation to align image edges. However, alignment artifacts in close shots are not considered. Liu et al [25] propose a burst-image denoising algorithm by selecting consistent pixels from a sequence of aligned images. The camera motion they deal with is much smaller compared to our stereo baseline. Shibata et al. [33] utilize joint filter with aligned guidance using weight volume. Jeon et al. [17] propose a stereo matching algorithm with color and monochrome cameras and recover a color image by colorizing the monochrome image. Limitations of colorization become obvious when some colorful structures are occluded in the color image, which can happen in close-range shots. Holloway et al. [15] and

Li [20] are the closest to our work, but require many cameras to avoid disocclusion artifacts, whereas we only need two cameras and deal with disocclusion in software.

3 Method

We formulate an optimization problem for jointly demosaicking and denoising the CFA image using detail information from the monochrome camera next to the color camera (Fig. 2). Our model works on raw sensor data extracted from cameras, however, if desired it can also apply to images preprocessed by the supplied image processing software of the device.

Our optimization step requires pixel-level alignment between the monochrome and RGB image pair. However, the alignment produced by state-of-the-art methods still contains artifacts, as discussed in the Section 1. The objective function is therefore designed to suppress artifacts in reconstructed images by encouraging agreement between reliably aligned pixels and refering to the noisy CFA image in badly aligned regions. This is achieved with the aid of a weight mask that assigns weight to pixels according to the alignment quality.

3.1 Image Alignment

In order to align the image pair, the raw CFA image $\mathbf{I}^{\text{cfa}}$ is initially demosaicked using DCRaw [6]. The alignment is performed between the green channel and the monochrome image, as the green channel is sharper and usually contains less noise. We have compared several different optical flow algorithms for alignment [23, 24, 27, 32, 36] and selected EpicFlow [32] because of its robustness to illumination changes which is crucial for cross-channel alignment. After alignment, the image from the monochrome camera is warped to match the raw CFA image. The warped monochrome image serves as a reference channel for the reconstruction step.

3.2 Joint Image Reconstruction

3.2.1 Image Formation Model

We denote the unknown reconstructed image as $\mathbf{I}^{\text{out}} \in \mathbb{R}^N$, the aligned reference monochrome image as $\mathbf{I}^{\text{ref}} \in \mathbb{R}^N$ and the input CFA image $\mathbf{I}^{\text{cfa}} \in \mathbb{R}^n$ in a vectorized form, where $N = 3n$ with n being the total number of pixels.

Recent methods for cross-channel and cross-image information transfer typically consist in a per-patch parameter evaluation step followed by a heuristic patch blending step [11, 34]. These algorithms require a pre-specified patch size, which is a parameter that heavily influences the quality of the outcome. To our knowledge, there exists no optimal way for automatic selection of the patch size. To avoid such complications, we propose a patch-free detail transfer model. We model the reconstructed image as an element-wise linear combination of the reference image $\mathbf{I}^{\text{ref}}$ and its reverse image $\mathbf{1} - \mathbf{I}^{\text{ref}}$:

$$\mathbf{I}^{\text{out}} = \mathbf{A} \odot \mathbf{I}^{\text{ref}} + \mathbf{B} \odot (\mathbf{1} - \mathbf{I}^{\text{ref}}) \tag{1}$$

where the correlation between sensor responses of the monochrome channel and individual color channels is modeled by two maps $\mathbf{A}$ and $\mathbf{B}$ with a piecewise smoothness constraint. The piecewise smoothness constraint is motivated by the observation that the correlation is approximately constant in a uniformly-colored region and piecewise linear near image edges

where discontinuities appear. The reverse image and its associated correlation map $\mathbf{B}$ are introduced for the purpose of decoupling low and high pixel values. For instance, when the reference image contains very dark pixels in some region, the reverse image can contribute to approximating the reconstructed image. Without the reverse image, the correlation map $\mathbf{A}$ would have to take extremely large values in such a region, which violates the imposed smoothness constraint. In our optimization problem, we exploit a cross-channel prior and a total variation prior as the smoothness constraint.

3.2.2 Optimization

Our reconstruction algorithm performs simultaneous demosaicking and denoising of a CFA image by optimizing the following objective function:

$$\min_{\mathbf{A},\mathbf{B}} \left\| \mathbf{I}^{\text{cfa}} - \mathbf{D}\big(\mathbf{A} \odot \mathbf{I}^{\text{ref}} + \mathbf{B} \odot (\mathbf{1} - \mathbf{I}^{\text{ref}})\big) \right\|_2^2 + \phi(\mathbf{A},\mathbf{B};\mathbf{W}) \tag{2}$$

where $\mathbf{D}$ denotes the CFA downsampling matrix. $\phi(\mathbf{A},\mathbf{B};\mathbf{W})$ represents the regularization term, where $\mathbf{W}$ is a precomputed weight matrix that reflects the reliability of the alignment at each pixel location (see Sec. 3.2.3). Pixels where $\mathbf{I}^{\text{ref}}$ and $\mathbf{I}^{\text{cfa}}$ are well-aligned are given high weight. Robustness to underlying artifacts in the reference image is achieved by weighting of the regularization term pixel by pixel.

Many existing natural image priors assume piecewise smoothness, which does not necessarily hold when the image contains lots of higher-order gradients and high-frequency contents. These smoothness priors, such as total variation (TV), often over-smooth textured regions. Such concern is alleviated in our image model since maps $\mathbf{A}$ and $\mathbf{B}$ are not expected to carry rich texture details, but model the correlation between channels instead. Texture and edge information is already encoded in the reference image $\mathbf{I}^{\text{ref}}$. Therefore strong smoothness conditions can be applied to $\mathbf{A}$ and $\mathbf{B}$ while automatically preserving texture and edges. Cross-channel gradient correlation [13] can be applied to enforce similarity between gradients of different chromatic channels l,k assuming $\nabla\mathbf{A}_l/\mathbf{A}_l \approx \nabla\mathbf{A}_k/\mathbf{A}_k$ and $\nabla\mathbf{B}_l \approx \nabla\mathbf{B}_k$ where ∇ denotes the image gradient operator. The latter gradients are not scaled since $\mathbf{B}$ in general takes low values, which is an observation in our experiments. We propose the following regularization term

$$\begin{aligned}\phi(\mathbf{A},\mathbf{B};\mathbf{W}) =& \lambda_1 \|\mathbf{W} \odot \nabla\mathbf{A}\|_1 + \lambda_2 \|\mathbf{W} \odot \nabla\mathbf{B}\|_1 \\ &+ \gamma_1 \sum_{l \neq k} \|\mathbf{W} \odot \big(\nabla\mathbf{A}_l \odot \mathbf{A}_k - \nabla\mathbf{A}_k \odot \mathbf{A}_l\big)\|_1 \\ &+ \gamma_2 \sum_{l \neq k} \|\mathbf{W} \odot \big(\nabla\mathbf{B}_l - \nabla\mathbf{B}_k\big)\|_1\end{aligned} \tag{3}$$

where λ_1, λ_2 are weights of the total variation prior on $\mathbf{A}$ and $\mathbf{B}$ respectively, and γ_1, γ_2 are weights of the cross-channel prior. We fix $\lambda_1 = 0.4, \lambda_2 = 0.6, \gamma_1 = 0.1, \gamma_2 = 1.1$ in all experiments. The weight matrix $\mathbf{W}$ forces strong smoothness and cross-channel similarity regularization when $\mathbf{I}^{\text{ref}}$ and $\mathbf{I}^{\text{cfa}}$ are well-aligned. $\mathbf{W}$ takes low values in the remaining pixels so that they are encouraged to match the observed CFA image while maintaining weak smoothness. The reconstructed image can be blended around the boundary between reliable and unreliable pixels with the denoised and demosaicked color image [8] to remove possible discontinuities. Notice that when no reference image is provided, simply taking $\mathbf{I}^{\text{ref}} = \mathbf{0}$, $\mathbf{W} = \mathbf{1}$ reduces the above formulation to a standard single-image demosaicing and denoising

problem [12]. Therefore the above optimization formulation can be seen as a generalization of optimization-based image demosaicking and denoising.

The proposed optimization problem (2) with regularizers defined in (3) can be handled by a number of convex non-linear solvers. We solve it with linearized ADMM [29], which in practice takes 25 iterations to obtain results later shown in the paper. Each iteration (see appendix) has O(n) complexity and all operations are parallelizable.

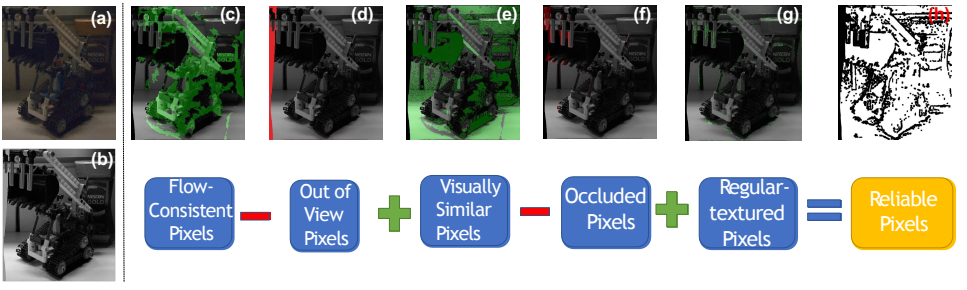

Figure 3: Classification of reliable pixels. Green and red indicates reliable and unreliable pixels, respectively. (a) Noisy RGB image, only the green channel is used in alignment. (b) Artifact-corrupted reference image. (c) Flow-consistent regions in green. A large portion of visually consistent pixels are excluded by the flow consistency check. (d) Pixels that flow out of view. (e) Visually similar pixels not satisfying flow consistency marked in green. (f) Occluded pixels in red. (g) Additional regularly textured pixels that are not identified to be visually similar in (e). The example image does not contain many textured regions that are not yet included. However this step clearly does not bring false positives when regular texture does not appear. (h) Reliable pixels shown in white.

3.2.3 W-Matrix Construction

In Eq. 2, **W** is introduced as an indicator of the quality of alignment. Our reconstruction model tolerates false negatives more easily than false positives. The false positive case means misclassification of badly aligned pixels that could violate the image formation model (Eq. 1). We hope to minimize the false negative cases (maximize “useful" pixels) while striving to avoid false positives.

The complete process of detecting well aligned pixels is outlined in Fig. 3. Two simple tests are first conducted to filter out a large proportion of badly aligned pixels. The first test measures the flow consistency by computing the L2 distance between forward and backward flows, followed by a thresholding step to identify reliable flows (Fig. 3(c)). The second test labels pixels that are mapped out of the image boundary as unreliable pixels (Fig. 3(d)). The above criteria are often used as simple tricks to detect reliable pixels in image alignment. However we observe that many visually consistent regions are excluded by these criteria.

Despite inaccurate flow, visually similar regions can still be exploited in our reconstruction algorithm as the monochrome image in general contains less noise. We measure visual similarity between the reference image and the green channel of color image in terms of texture similarity. Texture similarity can be identified by comparing the SIFT descriptor [26] on gradient images. SIFT behaves nicely in textured regions but its performance drops dramatically in textureless regions due to the existence of noise. Fig. 4(a) shows that similar

textured regions have small SIFT difference, which is not the case for visually similar textureless regions. We detect textureless regions by directly examining image gradients and mark them as visually similar. Fig. 3(e) shows an example of visually similar regions.

So far, reliable pixels consist of two classes identified using (i) flow consistency and (ii) visual similarity. However, there is an additional class of unreliable pixels that could be wrongly included by the above visual similarity group: occluded pixels. We rely on two heuristic cues to detect occluded regions: (i) the differences between image edges, because the occlusions usually result in double-edge artifacts in the warped reference image, and (ii) flow gradients, because occlusions appear near object boundaries each side of which has different scene depths and thus disparities/flows. Fig. 4(b) illustrates the detection of occluded region via image edges. Fig. 4(c) shows an example of the detection of a large occluded region. Some regularly textured regions, although visually consistent, could be confused with occlusion because of similar flow discontinuities when the alignment algorithm gets stuck in a local minimum. These pixels are identified using both SIFT difference and flow gradients and added back to the set of reliable pixels.

For the proposed detection process, a set of universal parameters is chosen so as to work across the whole set of images used in our experiments. Implementation details and the choice of parameters can be found in the supplementary material.

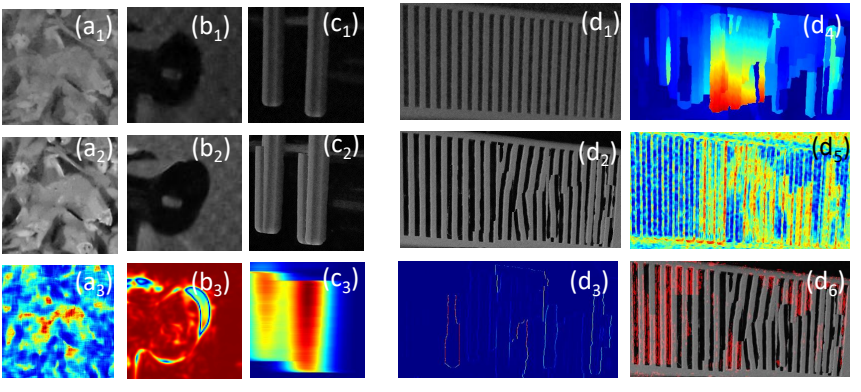

Figure 4: Illustration of reliable and unreliable pixels detection. Subscript “1" and “2" in the above notations indicates green-channel color and monochrome images. Red indicates high values in heatmaps. Left side: the first column (a) illustrates the detection of visually similar textured region via SIFT differences (a_3); The second column (b) illustrates the detection of occluded regions using image edges; (b_3) shows SSIM values computed for comparing edges; the third column (c) illustrates the detection of occluded regions using flow gradients. Right side: (d) illustrates the detection of regularly textured regions, (d_3) (d_4) (d_5)(d_6) represents absolute flow gradients, flow consistency, SIFT differences and detected well-aligned regular regions respectively.

4 Results

All images in the experiments were captured in low-light environments using a Huawei P9 smartphone with the same exposure time for both color and monochrome cameras. The ISO

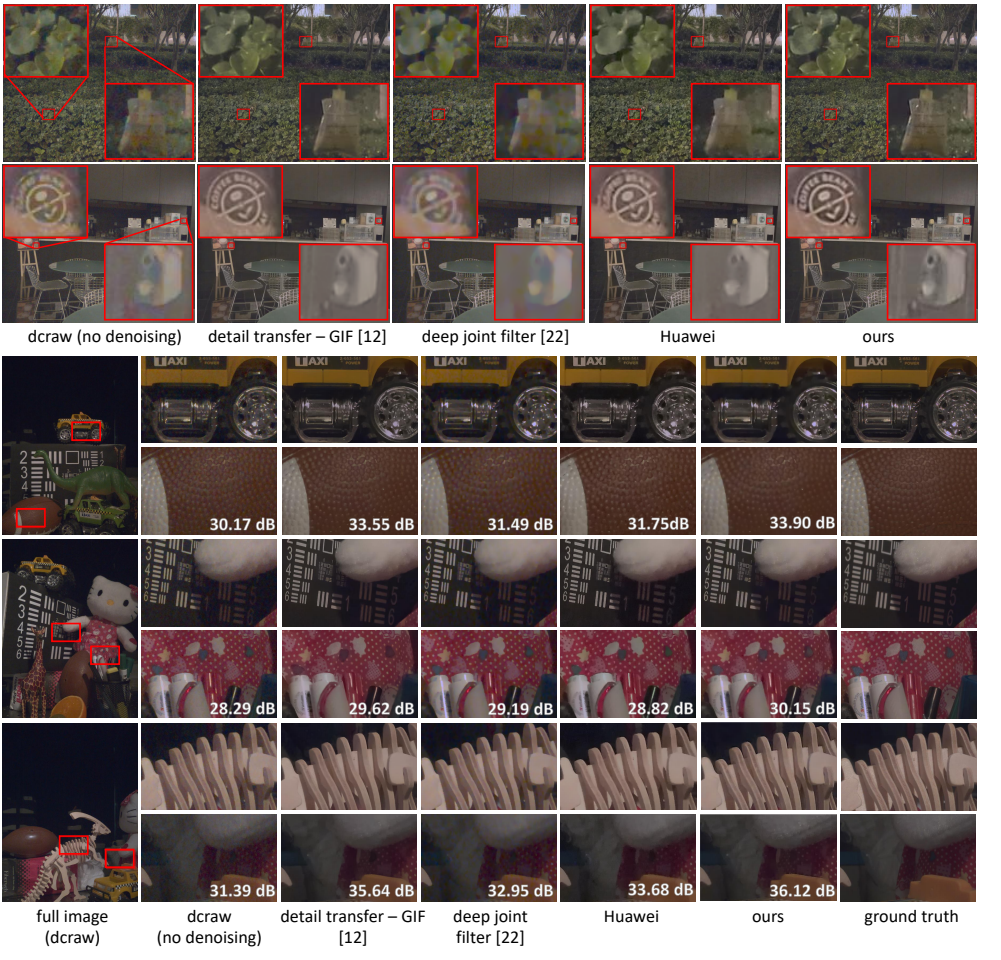

Figure 5: Qualitative and quantitative comparison on a few examples. Displayed PSNR are computed over the whole images. To view full images generated by all methods, please refer to the supplementary material.

setting is left in auto mode. We increase brightness of all displayed examples for better visualization of dark regions. We capture both faraway scenes and closeby objects, which lead to a wide range of disparities. Ground truth is obtained for a small number of sample scenes by extending the exposure time and lowering ISO accordingly. A more sophisticated way of capturing ground truth images can be found in [31].

Our algorithm works with a single setting of parameters. In other words, no image-dependent tuning is necessary (see supplementary material for the choice of parameters). We compare our method against Huawei's black-box software and two state-of-the-art multi-image denoising methods: a guided image filter version of the detail transfer approach based on layer decomposition [11, 30] and a learning-based deep joint filter [21]. The detail transfer algorithm is originally proposed for denoising flash/no-flash image pairs. It is adapted to pre-aligned monochrome/RGB pairs by treating the monochrome reference as the flash

image and combining their shadow and specular mask with the **W**-matrix obtained by our algorithm. For the deep joint filter algorithm, we provide a modified reference image by replacing unreliable pixels with pixels from the green channel of the demosaicked image by BM3D [7].

Figure 5 demonstrates that our approach outperforms state-of-the-art methods as well as the commercial software. Noise in real photography differs from naive synthetic noise [31], especially in low-light scenarios. Therefore the deep joint filter trained on synthetic data failed to remove color noise on real data. As compared to others, our method transfers most details from the monochrome camera and reconstructs sharper images (see zoomed-in views). Ground truth captured in a controlled environment is provided for the bottom examples in Figure 5. PSNR is calculated as a quantitative measure.

Misaligned pixels are sometimes not identified by **W**-matrix. Figure 6 shows that filter-based detail transfer algorithm transfers artifacts from reference to the color image whereas our optimization-based approach alleviates the effect of incorrect alignment. The learning-based algorithm does not generalize well on real data so that it cannot transfer details which include artifacts [71]. Unlike previous detail transfer algorithm [11, 30], our image formation model does not assume perfectly aligned reference image, which is the reason why our method is more robust to undetected artifacts.

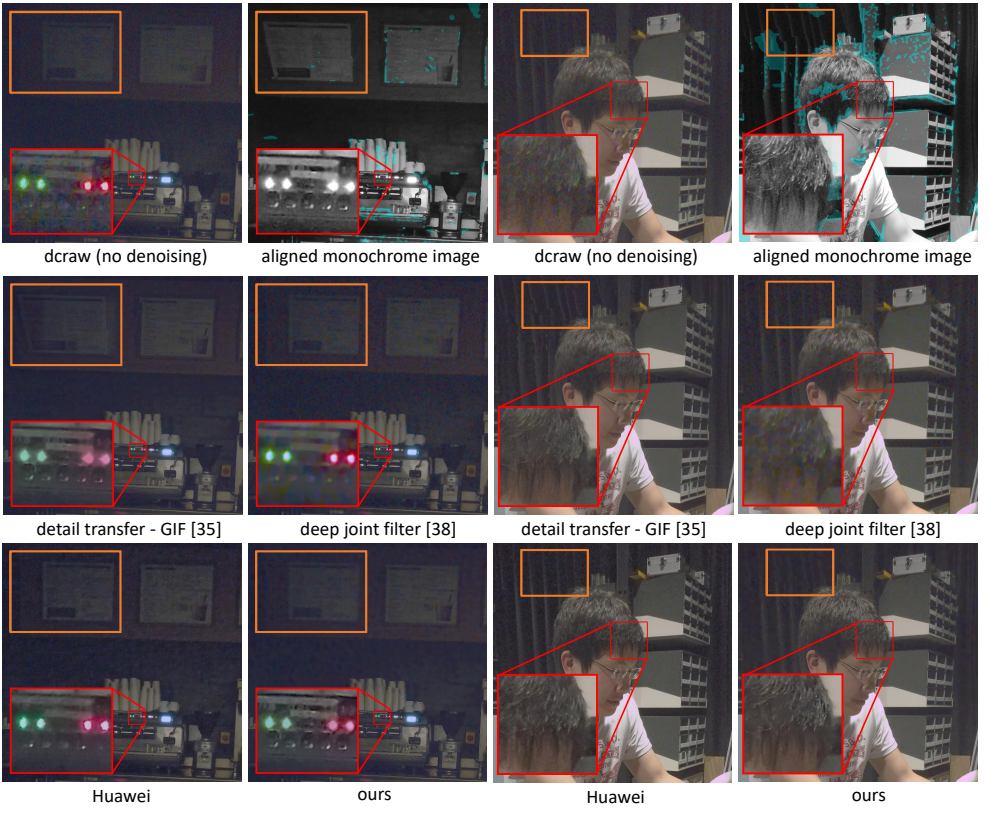

Figure 6: Unreliable pixel alignments detected by our **W**-matrix construction algorithm are highlighted in monochrome images. Misclassification of reliable pixels is shown in orange boxes. We achieve good balance between removal of undetected artifacts and detail transfer.

5 Conclusion

We have presented an end-to-end method that reconstructs high-quality color images from noisy CFA data with the aid of a monochrome camera in low-light scenarios. The joint demosaicking and denoising formulation exploits the similarity between the monochrome reference image and the latent color image. The formulation is robust to underlying artifacts brought by imperfect image alignment. This is achieved (i) by introducing a novel image formation model which allows imperfect alignment between two images, and (ii) by robustly and conservatively pixel-wise weighting of the regularization term to balance image quality improvement and artifact suppression. Our method can also transfer more details than previous state-of-the-art detail transfer algorithms between images.

References

[1] BFS-U3-04S2 EMVA imaging performance specification. https://www.ptgrey.com/support/downloads/10917. Accessed: 2018-03-13.

[2] Amit Agrawal, Ramesh Raskar, Shree K Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics (Proceddings SIGGRAPH 2005)*, 24(3):828–835, 2005.

[3] Hiroki Akiyama, Masayuki Tanaka, and Masatoshi Okutomi. Pseudo four-channel image denoising for noisy cfa raw data. In *IEEE International Conference on mage Processing (ICIP)*, pages 4778–4782. IEEE, 2015.

[4] Antoni Buades, Bartomeu Coll, Jean-Michel Morel, and Catalina Sbert. Self-similarity driven color demosaicking. *IEEE Transactions on Image Processing*, 18(6):1192–1202, 2009.

[5] Priyam Chatterjee, Neel Joshi, Sing Bing Kang, and Yasuyuki Matsushita. Noise suppression in low-light images through joint denoising and demosaicing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 321–328. IEEE, 2011.

[6] Dave Coffin. Decoding raw digital photos in linux. https://www.cybercom.net/~dcoffin/dcraw/. Accessed: 2018-03-14.

[7] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[8] Aram Danielyan, Markku Vehvilainen, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Cross-color bm3d filtering of noisy raw data. In *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, pages 125–129, 2009.

[9] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. *ACM transactions on graphics (Proceddings SIGGRAPH 2004)*, 23(3):673–678, 2004.

[10] Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics (Proceddings SIGGRAPH 2007)*, 26(3):51, 2007.

[11] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(6):1397–1409, 2013.

[12] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajkak, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, Jan Kautz, and Kari Pulli. Flexisp a flexible camera image processing framework. *ACM Transaction on Graphics (Proceddings SIGGRAPH 2014)*, 23(3):664–672, 2004.

[13] Felix Heide, Mushfiqur Rouf, Matthias B. Hullin, Bjorn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2013)*, 32(5):149:1–149:14, 2013.

[14] Keigo Hirakawa and Thomas W Parks. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005.

[15] Jason Holloway, Kaushik Mitra, Sanjeev Koppal, and Ashok N. Veeraraghavan. Generalized assorted camera arrays: Robust cross-channel registration and applications. *IEEE Transactions on Image Processing*, 24(3):823–835, 2015.

[16] Seunghoon Jee, Ki Sun Song, and Moon Gi Kang. Sensitivity and resolution improvement in rgbw color filter array sensor. *Sensors*, 18(5):1647, 2018.

[17] Hae-Gon Jeon, Joon-Young Lee, Sunghoon Im, Hyowon Ha, and In So Kweon. Stereo matching with color and monochrome cameras in low-light conditions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4086–4094, 2016.

[18] Daisuke Kiku, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Residual interpolation for color image demosaicking. In *IEEE International Conference on Image Processing (ICIP)*, pages 2304–2308. IEEE, 2013.

[19] Teresa Klatzer, Kerstin Hammernik, Patrick Knobelreiter, and Thomas Pock. Learning joint demosaicing and denoising based on sequential energy minimization. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2016.

[20] Feng Li. *A Hybrid Camera System for Low-Light Imaging*. PhD thesis, University of Delaware, 2011.

[21] Yijun Li, Jia-Bin Huang, Ahuja Narendra, and Ming-Hsuan Yang. Deep joint image filtering. In *European Conference on Computer Vision (ECCV)*, 2016.

[22] Ce Liu and William T Freeman. A high-quality video denoising algorithm based on reliable motion estimation. In *European Conference on Computer Vision*, pages 706–719. Springer, 2010.

[23] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(5): 978–994, 2011.

[24] Ce Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[25] Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, and Jian Sun. Fast burst images denoising. *ACM Transaction on Graphics (Proceedings SIGGRAPH Asia 2014)*, 33(6):232:1–232:9, 2014.

[26] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[27] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 1981 DARPA Image Understanding Workshop*, 1981.

[28] Enming Luo, Stanley H Chan, Shengjun Pan, and Truong Q Nguyen. Adaptive non-local means for multiview image denoising: Searching for the right patches via a statistical approach. In *IEEE International Conference on Image Processing (ICIP)*, pages 543–547. IEEE, 2013.

[29] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.

[30] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Transaction on Graphics (Proceddings SIGGRAPH 2004)*, 23(3):664–672, 2004.

[31] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2750–2759, 2017.

[32] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[33] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi. Misalignment-robust joint filter for cross-modal image pairs. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3315–3324, 2017.

[34] Tiancheng Sun, Yifan Peng, and Wolfgang Heidrich. Revisiting cross-channel information transfer for chromatic aberration correction. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[35] Huixuan Tang, Xiaopeng Zhang, Shaojie Zhuo, Feng Chen, Kiriakos N. Kutulakos, and Liang Shen. High resolution photography with an rgb-infrared camera. In *2015 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10, 2015.

[36] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. DeepFlow: Large displacement optical flow with deep matching. In *IEEE Intenational Conference on Computer Vision (ICCV)*, 2013.

[37] Hiroki Yamashita, Daisuke Sugimura, and Takayuki Hamamoto. Enhancing low-light color images using an rgb-nir single sensor. In *Visual Communications and Image Processing (VCIP)*, pages 1–4, 2015.

[38] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. *ACM transactions on graphics (Proceedings SIGGRAPH 2007)*, 26(3), 2007.

[39] Li Zhang, Sundeep Vaddadi, Hailin Jin, and Shree K Nayar. Multiple view image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1542–1549. IEEE, 2009.

[40] Shaojie Zhuo, Xiaopeng Zhang, Xiaoping Miao, and Terence Sim. Enhancing low light images using near infrared flash images. In *IEEE International Conference on Image Processing (ICIP)*, pages 2537–2540. IEEE, 2010.