ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/327283017

Color Image Demosaicking Using a 3-Stage Convolutional Neural Network Structure

Conference Paper · October 2018

DOI: 10.1109/ICIP.2018.8451020

CITATIONS
19

READS
1,534

3 authors:

Kai Cui
Technische Universität München
12 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Zhi Jin
Sun Yat-Sen University
50 PUBLICATIONS **429** CITATIONS
SEE PROFILE

Eckehard Steinbach
Technische Universität München
467 PUBLICATIONS **7,470** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Deep learning based image/video quality enhancement View project

Project Multimodal Teleoperation systems View project

All content following this page was uploaded by Kai Cui on 01 September 2018.

The user has requested enhancement of the downloaded file.

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

COLOR IMAGE DEMOSAICKING USING A 3-STAGE CONVOLUTIONAL NEURAL NETWORK STRUCTURE

Kai Cui[⋆] *Zhi Jin*[†⋆] *Eckehard Steinbach*[⋆]

[⋆] Technical University of Munich, Chair of Media Technology, Munich, Germany
[†] Shenzhen University, College of Information Engineering, Shenzhen, P.R. China

ABSTRACT

Color demosaicking (CDM) is a critical first step for the acquisition of high-quality RGB images with single chip cameras. Conventional CDM approaches are mostly based on interpolation schemes and hand-crafted image priors, which result in unpleasant visual artifacts in some cases. Motivated by the special characteristics of inter-channel correlations (higher correlations for R/G and G/B channels than that for R/B), in this paper, a 3-stage convolutional neural network (CNN) structure for CDM is proposed. In the first stage, the G channel is reconstructed independently. Then, by using the reconstructed G channel as guidance, the R and B channels are recovered in the second stage. Finally, high-quality RGB color images are reconstructed in the third stage. The objective and visual quality evaluation results show that the proposed structure achieves noticeable quality improvements in comparison to the state-of-the-art approaches.

Index Terms— Bayer color filter array (CFA), demosaicking, convolutional neural networks, residual learning

1. INTRODUCTION

In the majority of the modern digital cameras, for simplicity and low cost, a single image sensor with a color filter array (CFA) is used to acquire color images. With this kind of camera, only one intensity value (R, G, or B) can be recorded for a pixel at a time. In order to obtain the full-color images, an interpolation process is necessary to reconstruct the missing color components, which is usually called color demosaicking (CDM). The most commonly used CFA in modern cameras is based on the Bayer pattern [1]. Bayer pattern CFA based CDM has been extensively studied [2]. The simplest CDM method is to use bilinear or spline interpolation for each color channel independently. In these approaches, however, the correlation between different color channels is not exploited. Therefore, a key problem for CDM is how to exploit and make the most of the inter-channel correlations.

This work has been supported by a PhD grant from the China Scholarship Council for Kai Cui, a TUM University Foundation Fellowship and the National Natural Science Foundation of China (61701313) for Zhi Jin.

Many image priors have been introduced to model the inter-channel correlations (e.g., homogeneity [3], sparsity [4], local and non-local similarity [5, 6], integrated gradients [7]). However, these priors are mostly hand-crafted and lead to unpleasant visual artifacts in specific cases. Recently, residual interpolation (RI) [8, 9] based approaches were proposed to perform the CDM in the residual domain. The residual is defined as the difference between the original samples and a prediction of these samples using data from other color channels. The guided filter from [10] is adopted to minimize the residual in order to achieve a better estimation of the missing color components. Later, the RI method was extended to minimized-laplacian residual interpolation (MLRI) [11], which minimizes the Laplacian energy of the residual, iterative residual interpolation (IRI) [12], which performs the RI in an iterative manner, and adaptive residual interpolation (ARI) [13, 14], which adaptively chooses the number of iterations for each pixel. These RI based methods improve the results of CDM significantly, but in some cases, the performance is still not satisfactory.

With the success of convolutional neural networks (CNN) in image processing, CNN based CDM algorithms have also been proposed. In [15], a deep joint demosaicking and denoising structure was proposed, which can solve the two problems with a single network. In [16], a 2-stage network was proposed, which reconstructs the G channel in the first stage and then reconstructs the R/G/B channels jointly in the second stage. This approach achieves a significant PSNR improvement in comparison to previous approaches and represents the state-of-the-art performance.

$$r = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(A_{ij}-\bar{A})(B_{ij}-\bar{B})}{\sqrt{(\sum_{i=1}^{M}\sum_{j=1}^{N}(A_{ij}-\bar{A})^2)(\sum_{i=1}^{M}\sum_{j=1}^{N}(B_{ij}-\bar{B})^2)}} \quad (1)$$

$$\bar{A} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}A_{ij}, \bar{B} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}B_{ij}$$

Since exploiting the inter-channel correlation is the key point for CDM, we first focus on this part and measure the

inter-channel correlation for a large set of test images. 4744 images from the Waterloo Exploration Database (WED) [17] are used to calculate the correlation coefficients between R/G, G/B and R/B, respectively. The correlation coefficient is defined in Equation 1, where A and B are the considered color components of an image, $\bar{A}$ and $\bar{B}$ are the corresponding mean values, M and N are the width and height of the image. As shown in Table 1, the correlation coefficients for R/G and G/B are obviously higher than those for R/B. Also for R/G and G/B, even though the mean values are close, the variances are quite different. These tests show that the correlation between different channels is quite different.

Table 1. Inter-channel Correlations

Channels	Mean	Variance
RG	0.9010	0.0149
GB	0.9149	0.0105
RB	0.7892	0.0407

Inspired by these observations, in this paper, we propose a 3-stage CNN structure for color demosaicking. Since the G channel has twice as many samples compared to the R and B, in the first stage, the G channel samples are reconstructed independently. Because of the high inter-channel correlations between R and G, as well as G and B, in the second stage, we use two separate networks and the reconstructed G channel from the first stage, to guide the reconstruction of the R and B channels, respectively. In the third stage, the intermediate R, G, B obtained from the previous stage are concatenated as input, to exploit the correlations further. Finally, the high-quality demosaicked images are obtained.

The main contributions of the proposed approach are: (i) a 3-stage CNN, in which the performance of CDM is enhanced stage by stage. (ii) Two separate networks are used to reconstruct R and B in the second stage, which makes the most of the inter-channel correlations between R/G and G/B. (iii) An extensive evaluation and comparison with existing approaches on different datasets has been done, which proves the superiority and effectiveness of the proposed scheme.

2. PROPOSED 3-STAGE DEMOSAICKING SCHEME

In CDM with Bayer CFA, half of the pixels belong to the G channel, the other half is equally split between R and B. The three channels exhibit strong inter-channel correlation, both structurally and spectrally, which means that the samples from other channels can be used to enhance the quality of the current channel. Based on these characteristics, we propose the 3-stage CNN structure shown in Fig. 1 for CDM.

First, bilinear interpolation is used for initialization. Appropriate initialization makes the networks stable and easier to train. The first stage is designed to reconstruct the G channel. The $Input$ is split into $InputR$, $InputG$ and $InputB$, and the $InputG$ is fed into the first stage. Then, the output of the first stage $IntermediateG$ is concatenated with $InputR$ and $InputB$, respectively, and fed into the second stage. The second stage is designed to explore the correlations between R/G and G/B, using the high-quality $IntermediateG$ to guide the reconstruction of R and B. Using two separate networks in the second stage to reconstruct R/G and G/B is motivated by the aforementioned differences in the inter-channel correlation of R/G and G/B. Two separate networks can better model and make the most of the inter-channel correlations. In the third stage, we concatenate the obtained intermediate R, G, B data, as the input of the third stage, where the inter-channel correlations are further exploited. Finally, the demosaicking results are obtained from the third stage. The residual learning structure from [18] is used for each stage to boost the learning process.

Fig. 2 shows the detailed structure of the network unit for each stage. In the first layer, 128 filters of size $3 \times 3 \times d$ are used to generate feature maps, the last convolutional layer adopts d filters of size $3 \times 3 \times 128$ to generate the corresponding output. For the hidden layers, 128 filters of size $3 \times 3 \times 128$ are adopted. The number of the layers in each unit K is set to 5 and d is set to 1, 2, 3 in the three stages, respectively. Stride is set to 1, and zero-padding of size 1 is used to ensure that each feature map has the same size as the input.

Consider the training dataset $(\mathbf{X}_i, \mathbf{Y}_i)_{i=1}^N$, where $\mathbf{X}_i$ is the i-th Bayer pattern image, $\mathbf{Y}_i$ is the corresponding ground-truth RGB image, and N is the number of the images in the training data. During training, a loss function is defined to optimize the parameters of the networks. As shown in Fig. 1, four losses are defined for the proposed scheme. In the first stage, L_G is defined for the G channel. In the second stage, two loss functions L_{RG} and L_{GB} are defined since R/G and G/B are processed separately. In the third stage, L_{RGB} is defined as the loss for all three channels. The mean squared error (MSE) function is used as the loss function and the overall loss function used during training is defined as follows.

$$\begin{aligned} L(\omega_1, \omega_{21}, \omega_{22}, \omega_3) &= \frac{1}{4}(L_G(\omega_1) + L_{RG}(\omega_1, \omega_{21}) + \\ &\quad L_{GB}(\omega_1, \omega_{22}) + L_{RGB}(\omega_1, \omega_{21}, \omega_{22}, \omega_3)) \\ L(\omega) &= \frac{1}{N}\sum_{i=1}^{N}(\|\mathcal{F}(I_i;\omega) - O_i\|^2) \end{aligned} \tag{2}$$

where ω_j are the corresponding network parameters of the j-th stage. I_i and $\mathcal{F}(I_i;\omega)$ are the i-th input and output of each stage, and O_i is the corresponding ground-truth.

3. EXPERIMENTS AND RESULTS

The WED database promoted in [17] is adopted in our experiments as training data. In this dataset, there are 4744 full-color high-quality natural images of various scenes. We

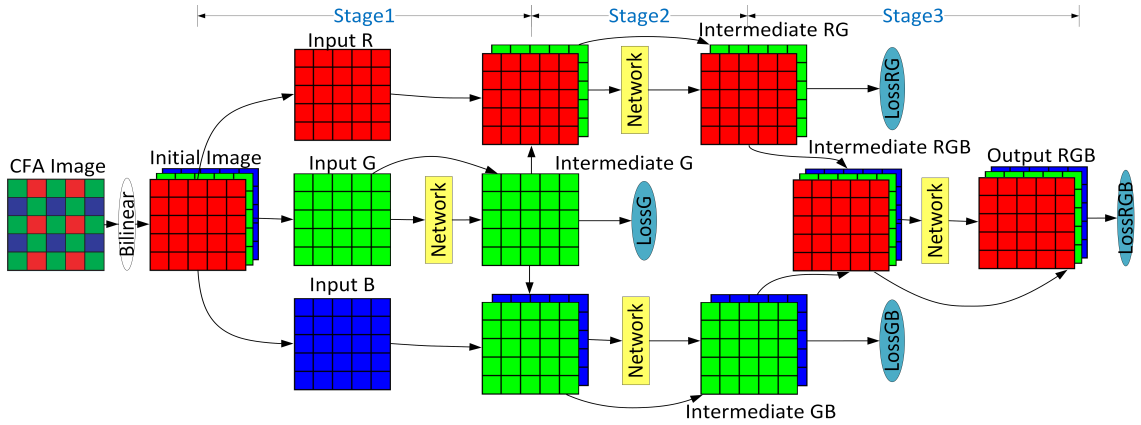

Fig. 1. Structure of the proposed 3-stage CNN scheme

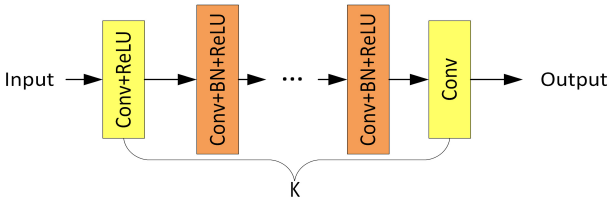

Fig. 2. Structure of Network Unit

randomly pick 4644 images for the training dataset, and the remaining 100 images are used as test dataset. The patch size is set to 50×50, and the patches are non-overlapped. There are 361728 patch-pairs generated from the training dataset. The mini-batch size is set to 64. The weights of the networks are initialized according to [18] and the Adam solver is used to optimize the parameters. The starting learning rate is 0.001, and divided by 5 every 20 epochs. There are 80 epochs in total. Other hyper-parameters are using the default settings from [19]. All the experiments are performed using Matlab(2017b) with the Matconvnet [20] toolbox.

There are two commonly used datasets for CDM evaluation, the Kodak [21] and the McMaster [5] dataset. However, for the Kodak dataset, the images have relatively low resolution and high spectral correlation, which is not optimal for the evaluation of CDM algorithms for modern digital cameras [16]. The McMaster dataset also has limitations concerning scene variety since it contains only 18 images. In [16], the authors proposed a new dataset WED-CDM, including 100 images from the WED dataset, which is composed of various scenes and color gradations, but currently this dataset is not publicly available. In order to evaluate the proposed scheme comprehensively, we also use the remaining 100 images from the WED dataset, named as WED-NEW dataset to evaluate the performance of the different CDM algorithms.

First, an example is presented in Fig. 3 to show the visual quality of the proposed method in comparison to existing algorithms. Usually the texture-rich and sharp color transition area is the challenging case for CDM. We zoom in the window part in the kodim19 to show the details. The 2-Stage-R refers to our re-implementation of the 2-Stage algorithm[1]. It can be seen that for previously proposed interpolation approaches, unpleasant visual artifacts can be observed along the edges of the window. With the proposed method, these artifacts can be well eliminated and the visual quality is improved.

The average Peak Signal-to-Noise Ratio (PSNR) and composite PSNR (CPSNR) are adopted to evaluate the objective quality of different approaches. Ten pixels along the border are cropped because some algorithms suffer from border effects. The results are listed in Table 2. The 2-Stage and the 2-Stage-R refer to the original results reported in [16] and our re-implemented version, respectively. There are PSNR differences between the results reported in [16] and those obtained with our re-implementation, which is possibly caused by randomly choosing the training data from the WED dataset. From the results, it can be seen that the proposed method leads to a 0.4–0.8dB PSNR improvement on different datasets in comparison to the existing algorithms.

We also adopt the structural similarity (SSIM) [22] as the evaluation metric on different datasets, as it takes structural information into account. Table 3 shows the average SSIM of different approaches on three datasets. The results are similar as for the PSNR evaluation. The proposed method outperforms the state-of-the-art.

In order to demonstrate the effectiveness of the proposed 3-stage scheme, Table 4 shows the intermediate results of each stage for the WED-NEW dataset. It can be seen that the quality is enhanced stage by stage. Especially in the second stage, the quality of all three channels is enhanced, which proves that not only G channel guides the reconstruction of R and B channels, but the reconstruction of G also benefits from the other two channels.

[1]The source code and the trained model of [16] are not publicly available, the codes and the trained model of our approach are available at `https://amnesiack.github.io/ICIP2018CDM/`

(a) Ground Truth (b) AHD (35.12dB) (c) DLMMSE (38.52dB) (d) GBTF (39.62dB) (e) LDI-NAT (35.29dB) (f) RI (35.57dB)

(g) MLRI (36.80dB) (h) ARI (38.84dB) (i) RI_new (36.33dB) (j) ARI_new (38.84dB) (k) 2-Stage-R(40.62dB) (l) Proposed (41.93dB)

Fig. 3. Visual Quality Comparison on kodim19 of Kodak dataset (Best seen on a computer monitor).

Table 2. Average PSNR and CPSNR results (in dB) for three datasets, the best performance is marked in bold face

Dataset	Kodak				McMaster				WED-NEW			
Method	R	G	B	RGB	R	G	B	RGB	R	G	B	RGB
AHD [3]	37.00	39.64	37.31	37.77	33.00	36.98	32.16	33.49	34.20	37.78	34.56	35.12
DLMMSE [6]	39.18	42.63	39.58	40.11	34.03	37.99	33.04	34.47	35.56	39.57	35.91	36.55
GBTF [7]	39.68	43.34	40.01	40.62	33.98	37.34	33.07	34.38	35.84	39.73	36.12	36.81
LDI-NAT [5]	37.14	39.48	37.01	37.71	36.19	39.52	34.37	36.12	35.62	38.69	35.71	36.37
RI [8]	37.94	41.00	37.82	38.61	36.10	39.99	35.38	36.50	36.00	39.52	36.41	36.93
MLRI [11]	38.87	41.83	38.86	39.58	36.35	39.90	35.36	36.62	36.53	39.93	36.82	37.42
ARI [13]	39.10	42.31	38.90	39.79	37.41	40.72	36.05	37.52	36.71	40.17	36.91	37.59
RI_new [9]	38.62	41.18	38.49	39.21	36.72	40.23	35.59	36.91	36.49	39.64	36.76	37.32
ARI_new [14]	39.27	42.43	39.10	39.95	37.45	40.68	36.21	37.60	36.73	40.20	36.93	37.58
2-Stage [16]	41.38	44.85	41.04	42.04	39.14	42.10	37.31	38.98	-	-	-	-
2-Stage-R	41.36	44.31	40.31	41.64	38.85	42.04	37.05	38.74	38.52	42.30	38.54	39.39
Ours	**42.07**	**45.18**	**41.09**	**42.39**	**39.60**	**42.60**	**37.68**	**39.39**	**39.32**	**43.04**	**39.37**	**40.19**

Table 3. SSIM results for three datasets

Dataset	Kodak	McMaster	WED-NEW
AHD [3]	0.9798	0.9573	0.9705
DLMMSE [6]	0.9866	0.9645	0.9777
GBTF [7]	0.9873	0.9637	0.9785
LDI-NAT [5]	0.9727	0.9690	0.9707
RI [8]	0.9826	0.9735	0.9776
MLRI [11]	0.9846	0.9729	0.9793
ARI [13]	0.9833	0.9760	0.9788
RI_new [9]	0.9835	0.9744	0.9789
ARI_new [14]	0.9840	0.9771	0.9793
2-Stage-R	0.9876	0.9793	0.9832
Ours	**0.9941**	**0.9802**	**0.9851**

Table 4. Intermediate PSNR and CPSNR (in dB) results for the proposed 3-stage approach on the WED-NEW dataset

PSNR	R	G	B	RGB
Input (Bilinear)	28.82	33.16	29.07	29.90
Stage1	-	36.52	-	-
Stage1+2	34.61	38.61	37.09	36.35
Stage1+2+3	39.32	43.04	39.37	40.19

4. CONCLUSION

This paper presents a 3-stage CNN-based color demosaicking scheme. The first stage is used to reconstruct the G channel. Then, in the second stage, with the guidance of the reconstructed G channel, R and B can be reconstructed and enhanced separately. In the third stage, all intermediate R, G, B results are concatenated as the input, and a high-quality color demosaicking image can be obtained. The experimental results on different datasets show that the proposed scheme leads to better performance than the state-of-the-art CDM algorithms. Also, the intermediate results of each stage show that the quality of the images is enhanced stage by stage, which proves the rationality and effectiveness of the proposed network.

5. REFERENCES

[1] B. E. Bayer, "Color imaging array," July 20 1976, US Patent 3,971,065.

[2] D. Menon and G. Calvagno, "Color image demosaicking: An overview," *Signal Processing: Image Communication*, vol. 26, no. 8, pp. 518–533, Oct. 2011.

[3] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 360–369, Mar. 2005.

[4] X. Wu, D. Gao, G. Shi, and D. Liu, "Color demosaicking with sparse representations," in *2010 IEEE International Conference on Image Processing (ICIP)*, Sept. 2010, pp. 1645–1648.

[5] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic imaging*, vol. 20, no. 2, pp. 023016, Apr. 2011.

[6] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2167–2178, Dec. 2005.

[7] I. Pekkucuksen and Y. Altunbasak, "Gradient based threshold free color filter array interpolation," in *2010 IEEE International Conference on Image Processing (ICIP)*, Sept. 2010, pp. 137–140.

[8] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Residual interpolation for color image demosaicking," in *2013 IEEE International Conference on Image Processing (ICIP)*, Sept. 2013, pp. 2304–2308.

[9] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1288–1300, Mar. 2016.

[10] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[11] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Minimized-laplacian residual interpolation for color image demosaicking," in *Proceedings of SPIE*, Mar. 2014, vol. 9023, p. 90230L.

[12] W. Ye and K. K. Ma, "Color image demosaicing using iterative residual interpolation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5879–5891, Dec. 2015.

[13] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, "Adaptive residual interpolation for color image demosaicking," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept. 2015, pp. 3861–3865.

[14] Y. Monno, D. Kiku, M. Tanaka, and M. Okutomi, "Adaptive residual interpolation for color and multispectral image demosaicking," *Sensors*, vol. 17, no. 12, pp. 2787, Dec. 2017.

[15] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 191:1–191:12, Nov. 2016.

[16] R. Tan, K. Zhang, W. Zuo, and L. Zhang, "Color image demosaicking via deep residual learning," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 793–798.

[17] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, Feb. 2017.

[18] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1646–1654.

[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, May 2015.

[20] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 689–692.

[21] "Kodak lossless true color image suite," `http://http://r0k.us/graphics/kodak/`.

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

View publication stats