International Journal of Computer Science and Business Informatics

IJCSBI.ORG

Demosaicing and Super-resolution for Color Filter Array via Residual Image Reconstruction and Sparse Representation

Jie Yin
School of Communication and Information Engineering
Shanghai University, Shanghai, China

Guangling Sun
School of Communication and Information Engineering
Shanghai University, Shanghai, China

Xiaofei Zhou
School of Communication and Information Engineering
Shanghai University, Shanghai, China

Abstract
A novel approach of demosaicing and super-resolution for Color Filter Array (CFA) based on residual image reconstruction and sparse representation is proposed. Given an intermediate image produced by certain demosaicing and super-resolution, a residual image between a final reconstruction image and the intermediate image is reconstructed using sparse representation. Richer edges and details are found in the final reconstruction image. Specifically, a generic dictionary is learned from a large set of composite training data composed of intermediate data and residual data. The learned dictionary implies a mapping between the two data. A specific dictionary adaptive to the input CFA is learned thereafter. Using the adaptive dictionary, the sparse coefficients of intermediate data are computed and transformed to predict residual image. The residual image is added back into the intermediate image to obtain the final reconstruction image. Experimental results confirm the state-of-the-art performance in terms of PSNR and subjective visual perception.

Keywords
Demosaicing; Super-resolution; Residual image reconstruction; Sparse representation

1. INTRODUCTION
Single chip named color filter array (CFA) is used in most resource constrained digital image/video capture devices [1]. The most popular Bayer pattern is illustrated in figure 1. Often, a full color and enlarged image produced from a low spatial resolution CFA are both required. Demosaicing is executed to get a full color image and super-resolution (SR) is executed to

get an enlarged spatial resolution image. Generally, there are two categories of schemes to achieve this goal in literature: The first is to demosaic CFA then superresolve the demosaiced CFA [2].Obviously, any approach for general image SR can be used in SR step; The second is to superresolve CFA then demosaic the superresolved CFA[3,4]. The major drawback of this method is that good methods for superresolving general image are not suitable for CFA image. Moreover, it is very difficult to design an appropriate SR solution for complex CFA pattern and has to be design different SR solution for different CFA pattern. The comparison indicates the more feasibility and flexibility of the former scheme.

While plenty of both demosaicing and SR techniques have been investigated and their combinations have obtained satisfied results [2, 5, 6], there are still much improvement work to do. For instance, as stated in [2], after demosaicing, if SR is implemented in multi-spectral color space individually, the color artefacts introduced by demosaicing will be worse during SR. As chromaticity channel is much smoother than intensity channel [2], SR in intensity channel and chromaticity channel will get better performance. We will also study the problem along with this direction.

In this work, a full color and enlarged image as an intermediate result is first obtained by using certain demosaicing and SR method, and then relying on it, a residual image making use of sparse representation is found to complement edges and details being lack of in the intermediate image. Two aspects are distinguished from our previous work [7]: One is a specific dictionary adaptive to the current image is further learned to improve the residual image reconstruction quality. The other is sparse coefficients of intermediate data are transformed to obtain residual image instead of a simple strategy of scaled residual image. It is necessary to point that in essence, arbitrary demosaicing and SR techniques or their combinations are allowed to get the intermediate image; however, better demosaicing and SR techniques are still expected to achieve more satisfied results. Hence in this work, an intermediate image is obtained by using methods proposed in [2].

The rest of the paper is structured as follows: section 2 outlines and discusses the proposed method in detail; section 3 provides experimental results; section 4 concludes the paper.

2. THE PROPOSED METHOD

2.1. Framework of proposed method

Framework of proposed method has been outlined in table 1.As mentioned above, since it is favoured that intensity and chromaticity channel are processed respectively and the intermediate image is acquired by method presented in [2], the intermediate result will contain three such channels: Green channel, RG difference and BG difference channels. The two difference channels are $R-G$ and $B-G$.It is admitted that the Green channel corresponds to intensity channel and the two difference channels correspond to chromaticity channel. Generally, structural and texture information is contained in intensity channel and chromaticity channel is merely related with chromatic information. Therefore, the residual image is only reconstructed for Green channel and nothing done for two difference channels without degrading visual quality. Another important issue must be emphasized that the color space of Green and two differences certainly could be transformed into other color space such as YCbCr; nevertheless, the artefacts involved in Green and two difference channels brought by demosaicing and SR will be accumulated in Y channel. That is the reason why we maintain Green channel, RG difference and BG difference channels.

2.2. Dictionary learning for residual image reconstruction

In recent years, sparse representation based on dictionary learned from data has been applied successfully to cope with image restoration tasks such as image denoising, deblurring and SR et al [8]. Inspired by [9], we also use sparse representation and dictionary learning to address our problem. Different from [9], in addition to learning a generic dictionary, we also learn a specific dictionary adaptive to the input CFA image content characteristics.

2.2.1 Generic dictionary learning

Given a training image set, the generic dictionary is learned as follows: First, it is simulated that an original image m is down sampled to a low spatial resolution image and further down sampled to a CFA image with Bayer pattern m' (other patterns easily can be extended). Second, m' is demosaiced and superresolved with certain techniques to get $\hat{m}$. This procedure is performed for each training image. Third, from m and $\hat{m}$ (in fact the Green channel of them), numerous image patch pairs in which two patches have same sizes and position at same locations are extracted; from two patches p and $\hat{p}$, residual patch p_r is produced by $p_r = p - \hat{p}$ and all such residual patches compose a residual image. Finally, p_r and edges of $\hat{p}$ are connected to form training data. To detect edges from $\hat{p}$, first-order edge extraction operators with horizontal, vertical, diagonal and anti-diagonal directions are convolved with $\hat{m}$. The four operators have been shown in figure 2. A target function of dictionary and sparse representation is designed in model (1) constrained by unit vector of atom:

$$\{\hat{\boldsymbol{D}},\hat{\boldsymbol{S}}\} = \arg\min_{\boldsymbol{D},\boldsymbol{S}} \left(\frac{1}{2}\|\boldsymbol{DS}-\boldsymbol{X}\|_F^2 + \lambda\|\boldsymbol{S}\|_{1,1}\right) \quad s.t.\ \boldsymbol{d}_j^T\boldsymbol{d}_j = 1, j = 1,2,\ldots K \qquad (1)$$

Where D, $\boldsymbol{S}$ are dictionary and representation coefficients matrix respectively, and X, K and λ denote training data matrix, the number of dictionary atom and regularization factor respectively. Alternative scheme is utilized to solve the function containing l1,1-norm regularization of representation coefficient matrix: Given dictionary, sparse coefficients are calculated and given sparse coefficients, dictionary is learned. Algorithms used for obtaining dictionary is the one introduced in [10] and algorithms used for obtaining sparse coefficients of each data is coordinate decent introduced in [11].

2.2.2 Adaptive dictionary learning

To make the generic dictionary more adaptive to the content of input CFA image, the generic dictionary is further modified according to the input CFA image. Specifically, the input CFA is demosaiced first and the filled Green channel is used to generate training data with the same manner as the generic dictionary learning does. Also, the learning process is driven with the same methods as the generic dictionary learning. Only difference lies in the ways of dictionary initialization: random training data compose initial dictionary for generic dictionary learning, and the learned generic dictionary compose initial dictionary for adaptive dictionary learning.

2.2.3 Separated data representation

Once $\hat{\boldsymbol{D}}$ is obtained, it is separated into residual dictionary and edge dictionary denoted as $\boldsymbol{D}^r$ and $\boldsymbol{D}^e$ respectively. Meanwhile, any training data x could be approximated well by $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{s}}$ as follows:

$$\boldsymbol{x} \approx \hat{\boldsymbol{D}}\hat{\boldsymbol{s}}, \quad \hat{\boldsymbol{d}}_j^T\hat{\boldsymbol{d}}_j = 1, j = 1,2,\ldots,K \tag{2}$$

Equation (2) could also been rewritten as residual part and edge part:

$$\begin{aligned}
&\boldsymbol{x}^r \approx \boldsymbol{D}^r\hat{\boldsymbol{s}} \\
&\boldsymbol{x}^e \approx \boldsymbol{D}^e\hat{\boldsymbol{s}} \\
&\quad = \left(\bar{\boldsymbol{d}}_1^e \cdot \left\|\boldsymbol{d}_1^e\right\|, \bar{\boldsymbol{d}}_2^e \cdot \left\|\boldsymbol{d}_2^e\right\|, \ldots \bar{\boldsymbol{d}}_K^e \cdot \left\|\boldsymbol{d}_K^e\right\|\right)\hat{\boldsymbol{s}} = \left(\bar{\boldsymbol{d}}_1^e, \bar{\boldsymbol{d}}_2^e, \ldots \bar{\boldsymbol{d}}_K^e\right)\begin{pmatrix} \hat{s}_1 \cdot \left\|\boldsymbol{d}_1^e\right\| \\ \hat{s}_2 \cdot \left\|\boldsymbol{d}_2^e\right\| \\ \ldots \\ \hat{s}_K \cdot \left\|\boldsymbol{d}_K^e\right\| \end{pmatrix}, \\
&\quad = \bar{\boldsymbol{D}}^e\hat{\boldsymbol{s}}' \\
&\left(\bar{\boldsymbol{d}}_j^e\right)^T\left(\bar{\boldsymbol{d}}_j^e\right) = 1, j = 1,2,\ldots K
\end{aligned} \tag{3}$$

Where $\bar{\boldsymbol{D}}^e$ denotes a normalized dictionary of which each atom is a unit vector and $\|\bullet\|$ denotes l_2-norm of a vector. Equation (3) implies that the sparse representation coefficients obtained from normalized edge dictionary could be used to approximate the residual data by relying on the relation between $\hat{\boldsymbol{s}}'$ and $\hat{\boldsymbol{s}}$.

2.3. Residual Green channel reconstruction

After an initial enlarged and full Green channel $\hat{G}$ is obtained, it will be convolved with four direction edge extraction operators. Then the convolved result of Green channel is partitioned into overlapping patches and each patch is represented by vector $\boldsymbol{y}^e$ composed of four direction edge features. The optimal sparse linear combinations of atoms of edge dictionary for $\boldsymbol{y}^e$ are searched through minimizing following formula:

$$s^* = \arg\min_s \left(\frac{1}{2}\left\|\bar{\boldsymbol{D}}^e \boldsymbol{s} - \boldsymbol{y}^e\right\|_2^2 + \lambda\|\boldsymbol{s}\|_1\right) \quad (4)$$

Similar to a problem contained in (1), a convex optimization regularized by l_1-norm is solved by coordinate decent algorithm [11]. Obviously, $\boldsymbol{s}^*$ plays the same role as $\hat{\boldsymbol{s}}'$ shown in equation (3) so that residual patch could be predicted as follows:

$$\boldsymbol{y}_r \approx \boldsymbol{D}^r \begin{pmatrix} s_1^* / \|\boldsymbol{d}_1^e\| \\ s_2^* / \|\boldsymbol{d}_2^e\| \\ \cdots \\ s_K^* / \|\boldsymbol{d}_K^e\| \end{pmatrix} \quad (5)$$

Once each residual patch $\boldsymbol{y}_r$ is reconstructed sequentially, the whole residual Green channel G_r is reconstructed. Consequently, the final reconstructed Green channel G is obtained as follows:

$$G = \hat{G} + G_r \quad (6)$$

Finally, B channel and R channel are reconstructed as follows:

$$\begin{aligned} R &= \hat{G} + R(\\ B &= \hat{G} + B(\end{aligned} \quad (7)$$

3. EXPERIMENTAL RESULTS

3.1. Training set option and parameter setting

An image set provided by [9] is selected as training set. A patch size of 6*6 is selected so that the dimension of edge component is 36*4=144 and the dimension of residual component is 36. Thus the total dimension of complete dictionary atom is 180. The number of dictionary atom is chosen as 1024 and the regularization factor λ is set as 0.5. To make a comparison with [2], enlarging factor 2 is tested.

3.2. Testing image and parameter setting

Kodak database is chosen as testing image set, see figure 3. Overlapping pixel number between adjacent patches is 2.

3.3. Evaluation by PSNR and subjective visual perception

The proposed method is compared with Zhang's scheme proposed in [2] both in terms of PSNR and subjective visual perception. Table 2 has listed average PSNR of three channels. For all testing images, higher performance has been achieved by the proposed method. From subjective visual perception criterion, enhancements of edges and details can be observed obviously in figure 4. In this experiment, three images are tested to demonstrate effects. Image (a) is input CFA image; Image (b) is a result of Zhang's method and image (d) is a result of our method. To show the quality of residual image reconstruction, residual images of the testing image have been shown in image (c). We can see that from large to small, multi-scale residual edges and details have been found by our method.

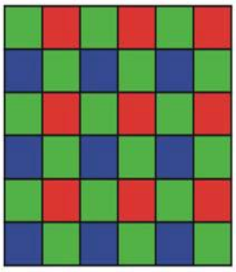


Figure 1.Bayer pattern of CFA

$$\begin{bmatrix}-1 & 0 & 1\end{bmatrix} \qquad \begin{bmatrix}-1\\ 0\\ 1\end{bmatrix} \qquad \begin{bmatrix}-1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix} \qquad \begin{bmatrix}0 & 0 & -1\\ 0 & 0 & 0\\ 1 & 0 & 0\end{bmatrix}$$

horizontal vertical anti-diagonal diagonal

Figure 2. Edge extraction operators

Figure 3. Twenty-four testing images from Kodak PhotoCD (referred to as image 1 to image 24, enumerated from left to right and top to bottom).

Table 1. Framework of proposed method

1. **Input**: A low spatial resolution CFA.
2. **Main Steps**:

   2.1 The input CFA is demosaiced and superresolved to produce $\hat{G}$, *RG* and *BG*.

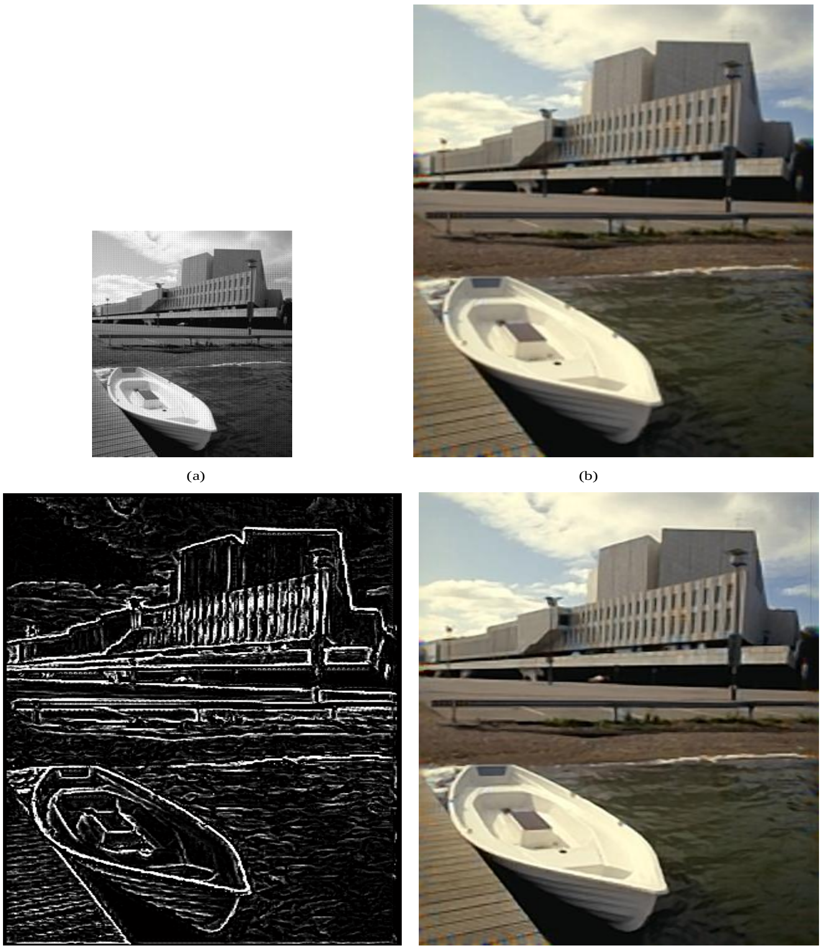
   2.2 A residual image of $\hat{G}$ is reconstructed by using an adaptive dictionary and sparse representation.

   2.3 The residual green channel is added back into $\hat{G}$ and the final green channel *G* is obtained accordingly.
3. **Output**: A high spatial resolution and full color image by changing from *G*, *RG* difference and *BG* difference into RGB.

Table 2. PSNR (dB) of proposed and Zhang's scheme

Image name	Zhang's	proposed	Image name	Zhang's	proposed
Kodim01	24.82	25.22	Kodim13	22.49	22.68
Kodim02	31.16	31.45	Kodim14	26.18	26.41
Kodim03	31.99	32.29	Kodim15	30.97	31.19
Kodim04	31.08	31.28	Kodim16	29.88	30.07
Kodim05	24.42	24.91	Kodim17	30.36	30.72
Kodim06	26.28	26.52	Kodim18	26.61	26.97
Kodim07	30.90	31.08	Kodim19	26.70	27.21
Kodim08	22.12	22.46	Kodim20	29.79	30.42
Kodim09	30.49	30.98	Kodim21	26.87	27.15
Kodim10	30.64	31.00	Kodim22	28.25	28.58
Kodim11	27.42	27.77	Kodim23	31.92	32.31
Kodim12	31.57	31.94	Kodim24	25.26	25.51

Figure 4-1



(a) (b)

(c) (d)

Figure 4-2

International Journal of Computer Science and Business Informatics

IJCSBI.ORG

Figure 4-3

Figure 4. Comparison of two methods for the purpose of subjective visual perception

4. CONCLUSION

In this paper, a novel scheme of demosaicing and SR for CFA via residual image reconstruction and sparse representation is presented. Using a training image set, a mapping between edge of filled and superresolved Green channel and corresponding residual image is obtained by dictionary learning. Given an intermediate Green channel, edges are extracted from it and sparse coefficients are searched using edge part in dictionary. The transformed sparse coefficients are utilized to linearly combine residual part of the dictionary to generate the residual image. Finally, the residual image is added back into the intermediate Green channel to produce a final reconstruction Green channel. Intermediate RG and BG channels are retained. The proposed scheme is capable of improving reconstruction quality of arbitrary demosaicing and SR method. The experimental results have demonstrated the state-of-the-art results in both PSNR and subjective visual perception.

5. ACKNOWLEDGMENTS

The authors are grateful to Professor Shuozhong Wang for his assistance in improving the language usage.

REFERENCES

[1] R. Lukac, K.N. Plataniotis. Color filter arrays: design and performance analysis, IEEE Transactions on Consumer electronics, 51 (4) (2005) 1260–1267.

[2] L. Zhang, David Zhang. A joint demosaicing–zooming scheme for single chip digital color cameras, Computer Vision and Image Understanding 107 (2007) 14–25.

[3] K.-L. Chung et al. New joint demosaicing and zooming algorithm for color filter array, IEEE Transactions on Consumer electronics, 55 (3) (2009) 1477–1486.

[4] R. Lukac, K.N. Plataniotis, D. Hatzinakos. Color image zooming on Bayer pattern, IEEE Transactions on Circuits and Systems for Video Technology 15 (2005) 1475–1492.

[5] X.Li, Bahadir Gunturk, L.Zhang. Image demosaicing: a systematic survey, Proceedings of the SPIE, Vol 6822(2008) 68221J-68221J-15.

[6] G. Crist′obal et al. Superresolution imaging: a survey of current techniques. Proceedings of the SPIE, Vol 7074(2008) 70740C.

[7] G.Sun, Y.Chen, Z.Shen.Demosaicking and zooming for Color Filter Array via residual image reconstruction. Proceedings of the Second International Conference on Internet Multimedia Computing and Service, 2010, 139–142.

[8] Michael Elad, M´ario A.T. Figueiredo. On the role of sparse and redundant representations in image processing, Invited Paper, Proceedings of the IEEE – Special Issue on Applications of Sparse Representation and Compressive Sensing, 2010,972–982.

[9] J.Yang, J.Wright, T.Huang, Y.Ma. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 19(11) (2010) 2861–2873.

[10] M. Yang, L. Zhang, J. Yang and D. Zhang. Metaface learning for sparse representation based face recognition. In ICIP, 2010.

[11] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1) (2010) 1–22.