Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XI về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR); Hà Nội, ngày 09-10/8/2018
DOI: 10.15625/vap.2018.00014

DEMOSAICING ALGORITHM - HOW TO EVALUATE IT?

Bui Duc Tho, Ho Phuoc Tien, Nguyen Tan Khoi

The University of Danang - University of Science and Technology

ABSTRACT: *Demosaicing plays an important role in generating color images in digital cameras. Yet, almost all demosaicing algorithms create artifacts, which reduce visual quality of a reconstructed color image. Objective evaluation of demosaiced images has therefore been of great interest and is essential to devising an efficient demosaicing method. Instead of focusing on one or two criteria as often carried out in the literature, the present paper will consider various criteria, and a set of demosaicing algorithms, for a thorough examination of quality of demosaiced color images. Our results show that the performance of these algorithms generally reveals a similar tendency over different criteria. Besides, we also observe that it is necessary to combine objective criteria to evaluate demosaicing algorithms effectively.*

Keywords: *demosaicing, objective evaluation, artifact.*

I. INTRODUCTION

Color demosaicing is an important operation in a digital camera and, hence, determines color quality of images generated by it. Usually, digital cameras capture only one color at each pixel location using a Color Filter Array (CFA); the most widely used CFA is the Bayer CFA pattern [1]. Demosaicing aims at reconstructing two missing colors at each pixel location in order to provide a full color image.

There have been a large number of demosaicing algorithms in the literature. They have treated the problem under various points of view: spatial domain or frequency domain, low or high complexity, optimization-based method or human-vision-inspired algorithm, etc. The readers can get an overview of demosaicing methods in [2-6].

In spite of much effort for solving the problem, a perfect demosaicing algorithm still remains unanswered. In fact, several issues need to be resolved before we expect a fully satisfied method. The most popular difficulties or artifacts that arise the most often are blurring, false color, and zipper effect (see more in section III). This leads to a natural question: how to evaluate a demosaicing method properly? The answer to this question will undoubtedly be essential to devising a high performance demosaicing algorithm, which would be very interesting in the industry of digital cameras or smartphones.

Evaluating image quality, in particular color quality, has long been of great interest in image processing. A first approach is naturally to ask humans to compare quality between two given images. While this subjective approach is still the most consistent because humans are the final judge in appreciating image quality, it becomes complicated when it comes to a large number of images. Therefore, an alternative approach of evaluation should be objective and based on mathematical formulas. Such a method allows us to evaluate images massively and easily. This is even more significant as there are, nowadays, abundant images from social networks or Internet. While human evaluation is reference, objective criteria have been carefully designed to approximate subjective ones.

The simplest objective criterion is undoubtedly PNSR. Since the last decade, SSIM has also attracted much interest. When it comes to demosaicing, some specific criteria are required to measure artifacts such as false color and zipper effect. While one or more of the above criteria has been used for demosaicing evaluation, according to our best knowledge, there has been no work that compares at the same time all these criteria. In this paper, we will focus on objective evaluation in order to try to answer the following questions:

- Can these criteria evaluate demosaiced color images consistently?

- Does a simple criterion, e.g. PSNR, provide results that are in line with those generated by complex criteria such as false color or zipper effect?

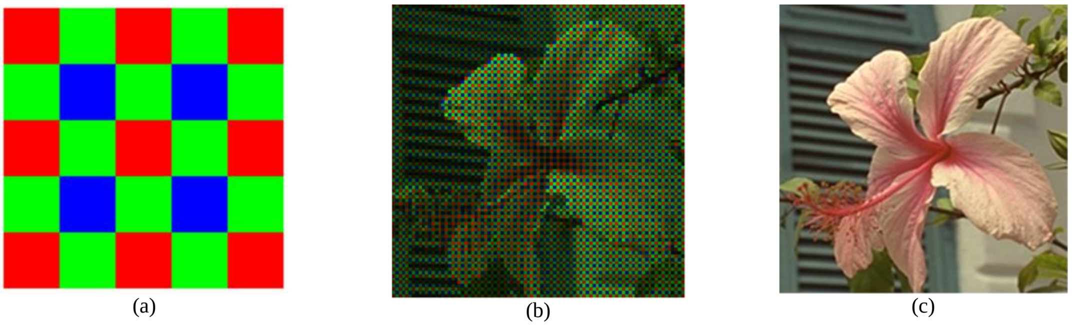
- What is/ are the best criterion/criteria for demosaicing evaluation?

We hope that the answers to these questions would eventually lead to an effective method to quantitatively measure demosaiced images in particular and image quality in general.

The remaining of this paper is described as follows. In section II, we revise demosaicing methods. In section III, we consider criteria that will be used to evaluate these demosaicing algorithms. Experiment and results will be presented in section IV. Finally, section V summaries some conclusions and perspective.

II. DEMOSAICING ALGORITHMS

We will review popular demosaicing methods. Some of them are of much interest in industry; others represent the state-of-the-art. Generally, we are interested in algorithms with low computational complexity but high performance. For all these algorithms, the objective is to reconstruct a full color image (Fig. 1c) from a mosaic image (Fig. 1b). The used CFA matrix is presented in Fig. 1a.



(a) (b) (c)

Fig. 1. (a) Bayer CFA pattern, (b) Example of a mosaic image, (c) Full color image reconstructed from (b) [6]

In the following, we summarize the algorithms that will be used for comparison in section IV. More details about these methods can be found in [6].

Bilinear interpolation: this method can be considered as the first one proposed for demosaicing. To reconstruct a color image, we interpolate, by convolutional masks, three channels Red, Green, Blue independently.

Laroche's method [7] belongs to the edge-based category of demosaicing. In fact, it takes into account edge direction in order to reduce interpolation error, i.e. interpolation is carried out along edge. Such interpolation is first applied to the Green channel. Once the Green channel is completed, the Red/ Blue channel is reconstructed based on the constant-color-difference assumption, which stipulates that color difference (e.g. between Green and Red or Green and Blue) of an object is constant.

Hamilton's method [8] improves Laroche's method by using the Laplacian (second order derivative) for the R and B pixels to correct the simple average interpolation for the G values. The advantage of this correction is to reduce aliasing in the reconstructed G channel.

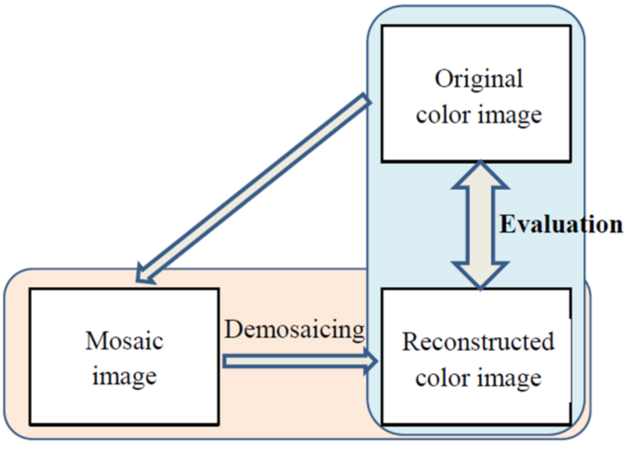


Fig 2. Evaluation procedure of demosaicing algorithms

Alleysson's method [9] provides a different approach to demosaicing when trying to exploit the spectrum of a mosaic image and the human visual system's characteristic that is the separation between luminance and chrominance. In spite of an analysis in the frequency domain, the algorithm is carried out in the spatial domain.

A. Kiku's method [10] is a state-of-the-art demosaicing algorithm. Its novelty comes from the residual interpolation combined with optimization.

B. Ho-Phuoc's method [6] makes an improvement to Alleysson's method. While keeping the separation between luminance and chrominance, the authors proposed to use edge-based interpolation for chrominance channels. Consequently, this algorithm reduces artifacts in highly detailed regions.

While Ho-Phuoc's method showed high performance with PSNR [6], we would like to confirm that with other objective criteria in the present work.

We implemented all of these algorithms except Kiku's method, which is available online [11].

III. PERFORMANCE EVALUATION

While subjective evaluation is always appreciated, it is not always easy to be carried out in practice. Another approach is to seek an objective criterion that can be as close to human evaluation as possible.

It is important to notice that image quality or color evaluation is subject to a great number of researches since more than a century. The main difficulty is to find out a space in which difference in colors can be measured in the same way as human perception of color. At this stage, since no single criterion has proved to be perfect, we will instead consider a set of widely used criteria and examine their eventual relations.

Fig. 2 describes a schema for evaluating demosaicing algorithms. In reality, we do not know the original color image: we only have the mosaic image, from which we will reconstruct a full color image. However, for the purpose of evaluation, we suppose that we also know the original image.

In the following, let us denote that I_1 is the original color image and I_2 is the reconstructed color image obtained from a demosaicing algorithm. M and N is the height and width, respectively, of the image. The objective criteria are described as follows.

A. PSNR

PSNR (Peak Signal-to-Noise Ratio) may be the most often used method to evaluate image quality due to its simplicity: it measures the pixel-wise difference between two images. PSNR can be adapted to color images by average the three color plans. Concretely, for demosaicing evaluation, PSNR is computed as:

$$PSNR(I_1, I_2) = 10 \log \frac{255^2}{MSE(I_1, I_2)} \tag{1}$$

$$MSE(I_1, I_2) = \frac{1}{3MN} \sum_{c=1}^{3} \sum_{i,j} [I_1^c(i,j) - I_2^c(i,j)]^2 \tag{2}$$

B. SSIM

SSIM (Structural Similarity) [12] is an effort to reduce the drawback related to absolute difference in PNSR. SSIM is a perception-based criterion that considers the perceived change in structural information. SSIM between images I_1 and I_2 is based on SSIM between a window X in I_1 and a window Y in I_2. X and Y have the same size and are at the same location.

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \tag{3}$$

μ_X, μ_Y is the average of window X, Y. σ_X, σ_Y is the variance of window X, Y. σ_{XY} is the covariance of X and Y. c_1 and c_2 are small constants to avoid a null denominator.

Then, $\mathrm{SSIM}(I_1, I_2)$ is the average of $\mathrm{SSIM}(X, Y)$ over all locations in the image. In this paper, we used a window of size 9×9. c_1 and c_2 are chosen as default values in [12].

C. False color

Demosaicing algorithms may create false colors, which often occur at edges or objects' boundary due to interpolation error. It has been of great interest to measure, by an objective criterion, a difference in color that is consistent with human perception. In order to obtain this goal, many color spaces other than RGB have been proposed to try to represent color in a similar way as in the human visual system. Some of the most popular color spaces are YCrCb, HSV, Lab, Luv. One important characteristic of these spaces is the separation between one channel of luminance and two channels of chrominance.

Table 1. Results for different criteria and demosaicing algorithms

	PSNR	**SSIM**	**False color**	**Zipper effect**
Bilinear	29.5053	0.9003	1.8230	0.4083
Laroche	34.3766	0.9585	1.3345	0.2953
Hamilton	37.0910	0.9757	1.0142	0.2230
Alleysson	35.5444	0.9748	1.1603	0.2332
Kiku	39.1951	0.9845	0.8460	0.1494
Ho-Phuoc	38.6929	0.9827	0.9306	0.1693

In the present paper, we will use the Lab color space to quantify difference between an original image and its demosaiced one. Moreover, we follow the method proposed by Zhang [13], in which color images are spatially filtered to take into account the different spatial sensitivity of luminance and chrominance. Color difference is then described by a distance, e.g. Euclidean distance. The source code for this false color criterion can be downloaded from [14].

D. Zipper effect

The zipper effect is characterized by the appearance of "on-off" pattern, which arises when the green plan is not correctly interpolated. The zipper effect at a pixel is determined as an increase or a decrease of its contrast with respect to its most similar neighboring pixel, when passing from the original image to the demosaiced one [15, 16]. Given a reconstructed color image, the zipper effect, represented by the percentage of pixels affected by this effect, is computed as follows:

- For each pixel p in the original color image, identify the pixel p* that is in the 8-pixel neighborhood and has the minimum color distance to the pixel p. This distance, $d_1(p)$, is computed in the Lab color space.
- Compute the color distance, $d_2(p)$, between these two pixels but in the reconstructed color image.
- The pixel p is considered affected by the zipper effect if $|d_1(p) - d_2(p)| > \delta$.
- The zipper effect between I_1 and I_2 is

$$Z(I_1, I_2) = \frac{\#\{p : |d_1(p) - d_2(p)| > \delta\}}{MN} \tag{4}$$

In this paper, we chose $\delta = 2.5$ as in [15].

In summary, these four criteria reflect various aspects of discrepancy between an original image and its demosaiced one. PSNR is the simplest criterion: it just evaluates pixel-by-pixel difference. SSIM considers image quality at a higher level in taking into account structural similarity or dissimilarity between the two images. False color criterion tries to examine color quality as in the human visual system. Finally, zipper effect considers the "on-off" pattern that appears very often in demosaicing algorithms. We hope that using a wide range of criteria will be able to evaluate image quality appropriately.

IV. EXPERIMENT

We will evaluate the performance of the demosaicing algorithms described in section II using four criteria presented in section III. Our purpose is to determine whether the performance of an algorithm is consistent over different criteria.

The well-known Kodak base of 24 lossless color images is used for testing. Each image is of size 1028×768. For each demosaicing algorithm A, we generated 24 demosaiced images (Fig.4, Fig. 5). Then, for each criterion C, we compute the average value over the 24 images to obtain the result for the couple algorithm A and criterion C. This process is applied to all demosaicing algorithms and criteria; the result is summarized in table 1. It is important to notice that for PSNR and SSIM, the higher these criteria, the better a demosaicing method; meanwhile for false color and zipper effect, the smaller these criteria, the better a demosaicing method.

Table 2. Correlation between criteria

Correlation coefficient	**PSNR**	**SSIM**	**False color**	**Zipper effect**
PSNR		0.9582	-0.9958	-0.9921
SSIM			-0.9710	-0.9555
False color				0.9910
Zipper effect				

From table 1, Kiku's method always provides the best result for each criterion. Bilinear interpolation performs the least. Ho-Phuoc's method continues to show its satisfied results despite its low computational complexity: this algorithm has better results than other methods, except Kiku's method, which is a state-of-the-art algorithm and requires the solving of an optimization problem. These results confirm observations in [6], where only PNSR is used.

For a better representation of the relation between criteria, we group the performance of all algorithms for each criterion (Fig. 3). The heights of the bars are relative: they are normalized by the highest value of each criterion. From Fig. 3 we can observe that the tendency of the algorithms is consistent for different criteria. For instance, if an algorithm provides good performance in PSNR it is also good in SSIM, false color or zipper effect.

However, if we look at these results in detail, we can observe subtle difference between criteria. For example, with Hamilton's method and Alleysson's method their difference in SSIM is less evident than in false color. In other words, the false color criterion may allow detecting difference in demosaicing algorithms' quality more effectively. Indeed, by comparing Fig. 4e and Fig. 4f, we can see more artifacts in Alleysson's method than in Hamilton's method. This difference is however not clearly reflected in SSIM; by contrast, it is well disclosed in the false color criterion. This observation seems reasonable when we know that SSIM concerns more about structure in intensity, while the false color criterion is sensitive to color error.

Relation between criteria

The above analysis broadly shows the same tendency of demosaicing performance over different criteria. Now we try to quantify the relation between these criteria. The correlation between two criteria is measured by the correlation coefficient for two variables X and Y:

$$\rho_{X,Y} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - E^2[X]}\ \sqrt{E[Y^2] - E^2[Y]}} \tag{5}$$

X or Y is a vector representing the results of all algorithms for one criterion.

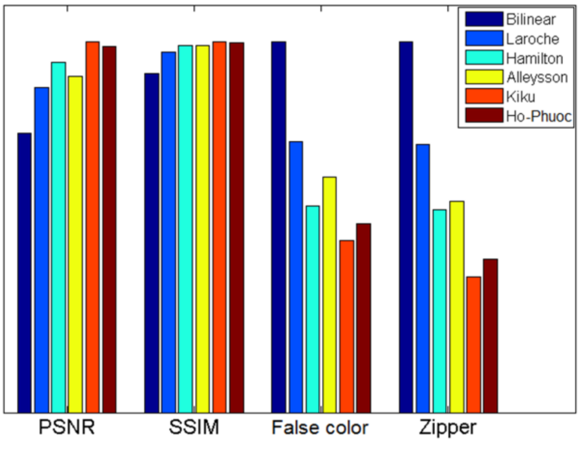


Fig. 3. Criteria displayed in relative values

The closer to one the absolute value of $\rho_{X,Y}$, the closer the relation between X and Y. When $|\rho_{X,Y}| = 1$, X and Y are linearly related (i.e. they can be represented by a line.) A negative value of $\rho_{X,Y}$ indicates that the variations of X and Y are in opposite directions.

We computed the correlation coefficient for every couple of criteria and reported these results in table 2. The very high value of $|\rho_{X,Y}|$ reveals the strong correlation between these criteria: it confirms the consistent tendency of demosaicing performance observed previously. Particularly, the correlation between PSNR, false color, and zipper effect is very strong (greater than 0.99). This observation is interesting as it shows that PSNR, which so far has been considered simple and purely mathematical, is closely related to false color and zipper effect, which are of more perceptual evaluation. Besides, we observe once again slightly less strong correlation between SSIM and the other criteria.

V. CONCLUSION

This paper presented a thorough evaluation of various demosaicing algorithms over different criteria. The results from the experiment showed highly similar tendency of well-known criteria for the demosaicing problem. We observed an interesting result in which PSNR, in spite of its simplicity, provides evaluation very close to those by false color and zipper effect. According to our best knowledge, the present paper is the first to study the correlation between well-known quantitative criteria. It is worth conducting further studies to confirm the above observation. For example, other image databases can be used with these criteria. In this case, images or short video sequences, which are very often used in super resolution [17, 18], may be considered.

Yet we keep in mind that false color and zipper effect can reveal and quantify color artifacts perceptually, which may not be reflected in SSIM or PSNR. While a universal metric for color image evaluation is still an open question, it would be necessary to consider combining various criteria to effectively evaluate image quality in general and demosaicing performance in particular.

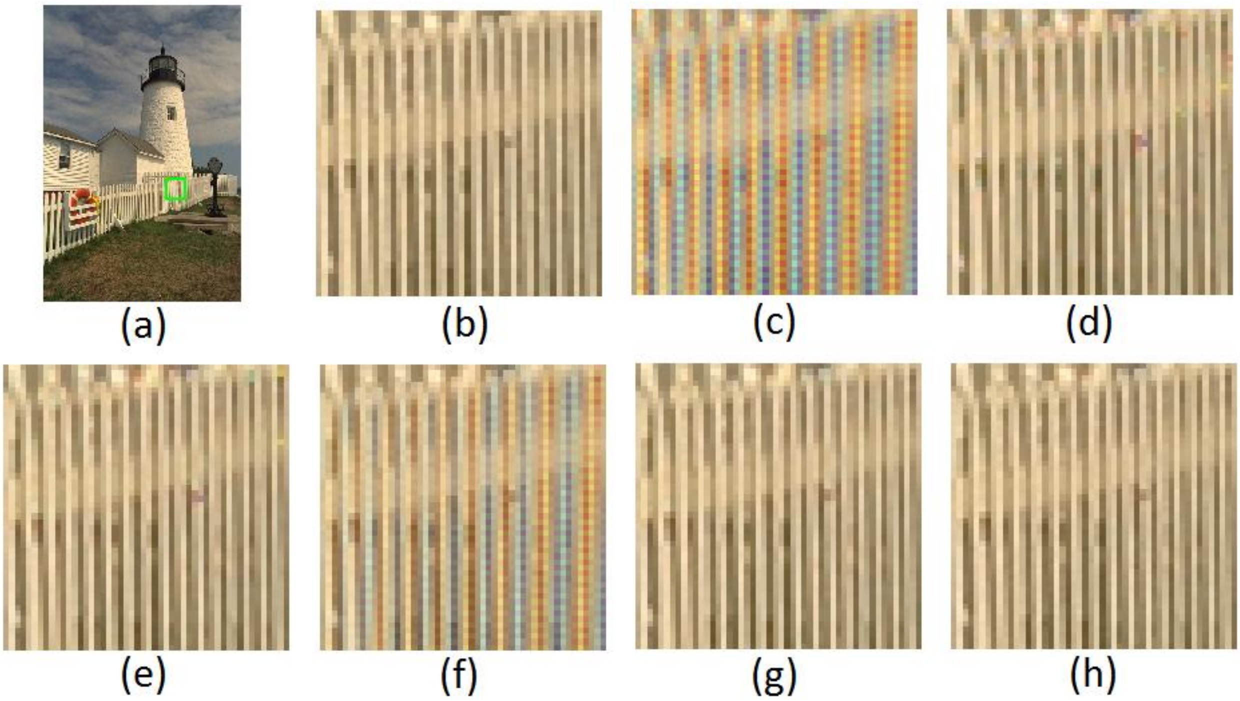


Fig. 4. Zipper effect (and also false color): (a) original image Kodim19, (b) zoom of the green square in the original image, (c) bilinear interpolation, (d) Laroche's method, (e) Hamilton'd method, (f) Alleysson's method, (g) Kiku's method, (h) Ho-Phuoc's method

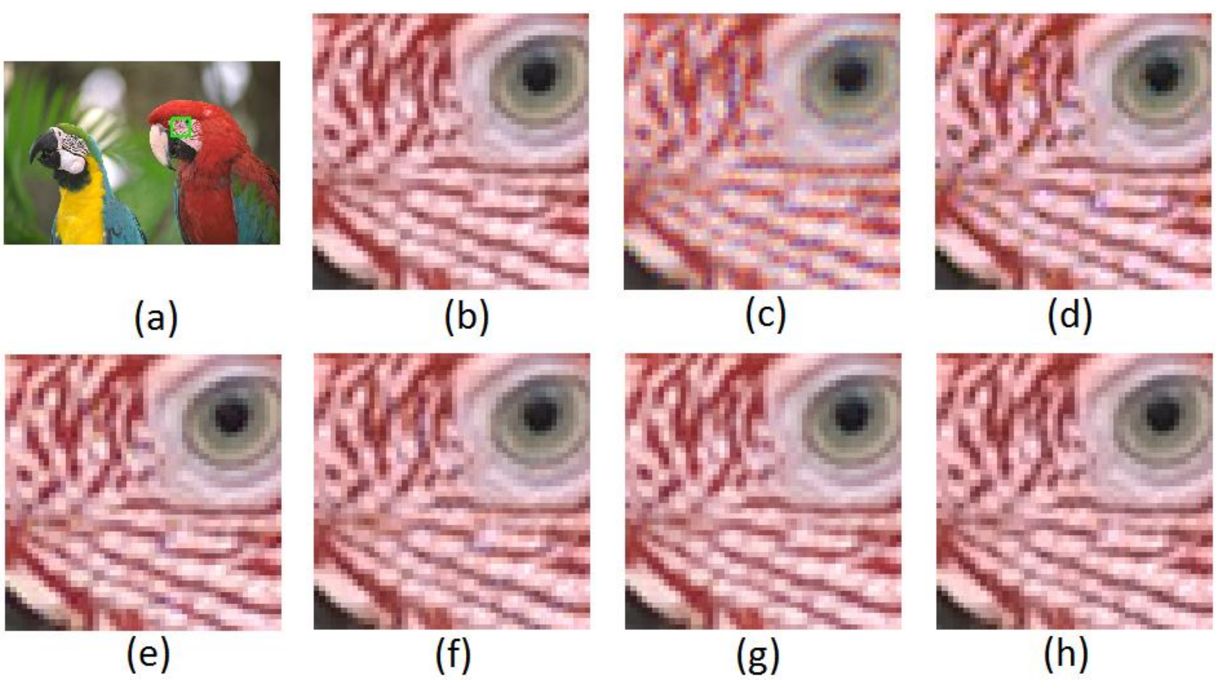


Fig. 5. False color: (a) original image Kodim23, (b) zoom of the green square in the original image, (c) bilinear interpolation, (d) Laroche's method, (e) Hamilton's method, (f) Alleysson's method, (g) Kiku's method, (h) Ho-Phuoc's method.

REFERENCES

[1] B. E. Bayer. "Color imaging array". U.S. Patent 3971065, 1976.

[2] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau. "Demosaicking: color filter array interpolation". IEEE Signal Processing Magazine, 22(1):44-54, 2005.

[3] A. A. Moghadam, M. Aghagolzadeh, M. Kumar, and H. Radha. "Compressive Framework for Demosaicing of Natural Images". IEEE Trans. Image Process., 22(6): 2356 - 2371, 2013.

[4] S. P. Jaiswal, O. C. Au, V. Jakhetiya, Y. Yuan, and H. Yang. "Exploitation of Inter-color correlation for Color Image Demosaicing". Proc. of IEEE Int. Conf. on Image Processing (ICIP), 2014.

[5] X. Huang and O. Cossairt. "Dictionary Learning based Color Demosaicing for Plenoptic Cameras" in Proc. of CVPR, pp. 449-454, 2014.

[6] T. Ho-Phuoc and D. N. Truong-Cong. "Compact and Efficient Algorithm for Color Demosaicing". Proc. of ATC, pp. 671-676, 2015.

[7] C. A. Laroche and M. A. Prescott. "Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients". U. S. Patent 5 373 322, 1994.

[8] J. F. Hamilton Jr. and J. E. Adams. "Adaptive color plane interpolation in single color electronic camera". U. S. Patent 5 629 734, 1997.

[9] D. Alleysson, S. Susstrunk, and J. Herault. "Linear demosaicing inspired by the human visual system". IEEE Trans. Image Process., 14(4):439-449, 2005.

[10] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi. "Minimized-Laplacian Residual Interpolation for Color Image Demosaicking". Proc. SPIE 9023, Digital Photography X, 90230L, 2014.

[11] http://www.ok.ctrl.titech.ac.jp/res/DM/RI.html

[12] 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, 13(4): 600-612, 2004.

[13] X. Zhang and B. A. Wandell. "A spatial extension of CIELAB for digital color image reproduction" in Soc. Inform. Display Symp. Tech. Dig., vol. 27, pp. 731-734, 1996.

[14] http://white.stanford.edu/~brian/scielab/scielab.html.

[15] A. Buades, B. Coll, J. M. Morel, and C. Sbert. "Self-similarity driven color demosaicking". IEEE Transactions on Image Processing, 18(7): 1192-1202, 2009.

[16] W. Lu and Y. Tan. "Color filter array demosaicking: new method and performance measures". IEEE Transactions on Image Processing, 12(10): 1194-1210, 2003.

[17] T. Ho-Phuoc, A. Dupret, and L. Alacoque. "Super resolution method adapted to spatial contrast". ICIP, 976-980, 2013.

[18] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang. "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution". CVPR, 2017.

PHƯƠNG PHÁP ĐÁNH GIÁ GIẢI THUẬT DEMOSAICING

Bùi Đức Thọ, Hồ Phước Tiến, Nguyễn Tấn Khôi

TÓM TẮT: *Demosaicing đóng vai trò quan trọng đối với việc tạo ra ảnh màu trong các máy ảnh số. Nói chung, các giải thuật demosaicing đều không hoàn hảo và làm ảnh hưởng đến chất lượng của ảnh màu thu được. Do đó, tiêu chí đánh giá khách quan về chất lượng ảnh sau demosaicing có ý nghĩa quan trọng, góp phần xây dựng giải thuật demosaicing hiệu quả. Khác với những nghiên cứu trước đây, bài báo này sẽ khảo sát đồng thời nhiều tiêu chí khác nhau, cũng như nhiều giải thuật demosacing, nhằm xem xét một cách đầy đủ về chất lượng của ảnh sau demosaicing. Kết quả cho thấy có sự tương đồng trong xu hướng thay đổi giữa các tiêu chí khác nhau đối với các giải thuật demosaicing được xem xét. Bên cạnh đó, bài báo còn cho thấy nên kết hợp nhiều tiêu chí khách quan để có thể đánh giá chất lượng ảnh một cách hiệu quả.*