Perceptual Evaluation of Demosaicing Techniques

Tomasz Sergej*
Supervised by: Radoslaw Mantiuk†

West Pomeranian University of Technology in Szczecin
Poland

Abstract

Most of the modern digital cameras capture colour images based on the CFA (Colour Filter Array) pattern. Each point of the sensor matrix acquires one of the R,G, or B values. The demosaicing algorithms convert mosaic consisted of these points to a standard digital image representation with three RGB values for each colour channel. The quality of the demosaicing algorithms is measured as a level of deformation in an image after demosaicing compared to the original appearance of the scene. In particular, unfavourable are the artefacts that are well seen by a man, like an object blurring or colour halos. In this paper we conduct a series of the perceptual experiments to measure the degradation of the image quality after demosaicing. We evaluate six demosaicing techniques that differ in a concept of used approximations. Perceptuallly visible artefacts are searched for a native image size and also for its magnification. The results of the experiments can be considered as the reference to another quality assessment algorithms.

Keywords: image difference metrics, image quality, demosaicing, perceptual experiments, image processing

1 Introduction

Modern digital cameras use CCD (charge-coupled device) or CMOS (complementary metal-oxide semiconductor) sensors covered by the CFA (Colour Filter Array) filters to register colours. The intensity of light is measured by only one sensor, CFA filter separates the light sensitivity for each colour separately. CFA is an array of three alternating colour filters that pass only one colour (Red, Green or Blue) to the sensor. Green colour filters occupy half of the array, because the (HVS) Human Vision System is the most sensible to this colour. Each pixel has only one colour measured. The demosaicing algorithms are designed to recover the missing values.

Demosaicing algorithm are approximating algorithms. Each algorithm like this cause the degradation of the image quality. It is hard to assess this degradation using the objective methods, because the difference between values of pixels does not reflect changes in the perceptual quality. Even large differences indicated by the MSE (Mean Square Error) metric may not be noticeable to the human eye. On the other hand, small differences indicated by the MSE could be noticed as a significant difference in the appearance of the image.

In this work we conduct experiments which evaluate the image quality degradation after demosaicing process. Observer is asked to indicate which of the two image is of better quality. The images are generated by various demosaicing algorithms. The measure of quality is the incidence of the artefacts in the image. The results of the experiment taken on 18 people were subjected to statistical analysis. Analysis shows statistical significance of the observed trends and makes final results.

We evaluate both the images displayed in their native resolution, and their twice- and four times enlargements. This way, we test how the image magnification influences the visibility of the artefacts. This operation is a typical image processing performed before displaying or printing digital images and, in our opinion, should be considered during the quality evaluation.

In Sect. 2 the artefacts occurring after demosaicing are presented and discussed. In Sect. 3 we discuss the basic concepts of the subjective image quality assessments. In Sect. 4 we present details of the conducted experiments. Their results are analysed in Sect. 4.4. The paper ends with conclusions and providing directions for further work in Sect. 5.

2 Demosaicing

Demosaicing algorithms

Demosaicing algorithms used during the test are of different degree of complexity and algorithmic rules. The simplest algorithm - *BI* (Bilinear Interpolation) is based on bilinear interpolation of missing colour channels. This method is fast but introduces the artefacts in high contrast and with colour changes areas (e.g at the edges of objects, see Fig. 1).

A better approach is to interpolate values along the edges of objects (edge-directed interpolation). The green channel of Bayer mosaic contains the most important information about the progress of the edge. There is the

*tomaszs89@gmail.com
†rmantiuk@wi.zut.edu.pl

highest sampling rate of the scene for this channel. To detect the edge the difference / gradient between green colour components in horizontal and vertical direction are calculated. The smaller difference determines the course of the edge, and hence the direction of the interpolation [6]. Advanced methods use higher pixel environment (e.g 5x5) and include information from the red and blue channels in the calculations. This approach reduce the aliasing of the final image [5]. Along the edges interpolation algorithm is used in this paper, which is called *GBI*, and was proposed in [8].

The correct selection of the direction of interpolation is the determining factor of democaicing algorithm quality. In [7] an algorithm was proposed, in which the vertical or horizontal direction is chosen on the basis of the local similarity between pixel values (local homogeneity). This similarity determines the difference of luminance and chrominance in the immediate vicinity of the pixel. It is calculated in the CIE L*a*b color space. The similarity coefficient is determined for two directions. The selected is the one with the greater value of similarity coefficient. This technique is called *AHD* (Adaptive Homogeneity-Directed) and is considered to be one of the best demosaicing algorithms. *AHD* is used in a popular program to convert images from RAW format - *dcraw* [3].

We also included demosaicing algorithms: *Gunturk*, *Lu* and *Li*. Gunturk algorithm is designed on the basis of observation of the correlations between the red, green, and blue channels. This demosaicing algorithm removes aliasing in these channels using an alternating-projections scheme. More in [4]. Lu method consists of two successive steps: an interpolation step that estimates missing colour values, and a post-processing step that suppresses noticeable demosaicing artefacts by adaptive median filtering. More in [11]. Li algorithm presents an iterative approach. The quality of the resulting image depends on the number of the iterations. More in [9].

During the experiment we used original implementations of *AHD*, *Gunturk*, *Lu* and *Li* algorithms. GBI algorithm was used in his original implementation built in Matlab. The BI algorithm was written independently.

Demosaicing artefacts

The goal of demosaicing techniques is to fill in the missing information in colour channels. In other words, these algorithms define the missing R,G,B pixel values. The perfect demosaicing algorithm can estimate the same that appear in the real scene. Due to the lack of values one needs to use algorithms based on the interpolation. Interpolation process causes artefacts in the resulting images. The most exposed to artefacts are contrasted and colourful places in the image. Demosaicing algorithms are designed to minimise those artefacts. The most disturbing are those artefacts that are easily perceived by people, like artefacts caused by the wrong direction of interpolation that are recognisable as the zippering effect (see Fig. 1a).The disruption of the ratio between the values of the colour channels results in the false colour effect (see Fig. 1b). These two types of artefacts occur mainly along strong edges in the image. Accumulated false colour effect provides to Moire effect (see Fig. 1c). The incorrect technique of averaging values may cause image blurring (see Fig. 1d).

3 Subjective quality metrics

Subjective methods are designed to measure the quality of digital images [10]. These methods are referred as subjective because they are based on the opinion of people (observers). These opinions are gathered during the perceptual experiments. The observer determines the rank of the compared images through a series of decisions. Often they also defines a quantitative measure of quality using Likert scale.

There are four most commonly used methods of experimental evaluation of image quality. The evaluation with a reference image (*double stimulus*) and without reference image (*single stimulus*) represent a categorical ranking. In the *forced choice* method, the observer is forced to make a choice between the two images. In the *similarity judgement* method the observer is not only judging which of the two images is better, but also determines the difference in quality between them using the Likert scale

In this work, we decided to use the forced choice technique (see Fig. 2). This method is granted as the most effective [10]. During the experiment, pairs of images (test image and reference image) are displayed in random order. The observer is asked to select an image of the better quality. He/she is always forced to choose one of the images, even when they does not notice any differences in quality between them. There is no time limit for making a decision. The method is straightforward, so it is expected to be more precise than other techniques. However, it requires a great number of repetitions to compare every possible combination of images ($(0.5(n^2 - n)$ for n input images).

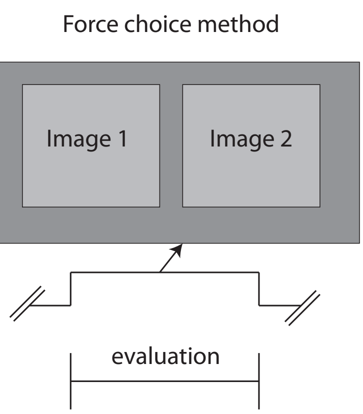

Figure 2: Force choice technique (image reproduced by [10]).

4 Experimental Evaluation

The main goal of the experiment was to compare the perceptual quality of images generated by the set of demosaicing algorithms. The images are compared among themselves and with the corresponding reference images.

4.1 Stimuli

During the experiment we used a set of four scenes selected from the Kodak Image Suite (see Fig. 3). This dataset is commonly used to evaluate demosaicing algorithms. It is characterised by a wide range of presented scenes, colours, and textures.

The preparation of the stimuli was based on converting the original images from the Kodak set to the Bayer mosaic. The conversion was performed by the elimination of the corresponding colour channels. After that the RGB images were reconstructed using every chosen demosaicing algorithm. Additionally, these RGB images were twice and four times enlarged choosing the most interesting areas (see Fig.3). We used the bicubic interpolation technique from Adobe Photoshop CS5 to achieve the high quality of this magnification.

Summing up, we generated three images for every scene for every demosaicing technique and additionally three reference images presenting original scene (one of a native resolution and two magnified). The size of all the images was 512 x 768 pixels.

4.2 Apparatus and experimental procedure

The experiment was performed in a darkened room. Images were displayed on 24" Eizo ColorEdge CG245W monitor with a native resolution of 1920 x 1200 pixels. The display was equipped with a hardware colour calibration module and was calibrated before each experimental session.

The software implemented for this experiment uses the forced choice method (see Fig. 2). It was implemented in Matlab with Psychtoolbox package [1].

During the experiment, the observer was sitting in front of the display at a distance of 65 cm. His head was stabilised by the chin rest. Two images were displayed on the monitor (see Figure 4. The observers task was to point using the mouse cursor, the image of the better quality. The grey background was displayed for 2 seconds between every comparison.

4.3 Participants

The experiment was conducted for 18 individual observers (age between 21 and 25 years old, 16 males and 2 females). All participants had normal or corrected to normal vision and correct colour vision. The participants were aware that the image quality is evaluated, but they were naive about the purpose of the experiment.

Figure 4: The experimental setup.

To shorten the time of a single experiment session, the input images were divided into 4 sets. The evaluation of each set lasted less than 20 minutes and every set was evaluated by at least 4 observers.

4.4 Results

The results of the experiment are presented in Fig. 5 . The horizontal bars show the number of cases in which the observers decided that a given demosaicing technique generated better images.

It's worth emphasising that excluding the BI algorithm, there was no statistically important difference between the algorithms when the images of a native size were compared (see Fig. 5,*averagex1*). In practice, only four times magnification reveals the target ranking (see Fig. 5,*averagex4*).

The AHD algorithm proved to be the best solution. Statistically, its rating is the same as the reference scene. The Li, Lu, and Gunturk techniques can be classified to the second group with slightly worse results. They are statistically better than the GBI algorithm. As it was expected, BI is the worst technique.

5 Conclusions and future work

In this work we conducted the perceptual study of visibility of the demosaicing artefacts. The quality of images after demosaicing using six different techniques was evaluated. Interestingly, the observers did not notice any differences among distorted images and the reference scenes presented in their native resolutions. To reveal the differences among the algorithms, the images had to be enlarged around their most significant areas. The AHD algorithm proved to be the best demosaicing techniques with the quality close to the reference.

In future work we plan to evaluate the visibility of artefacts in the form of the localised distortion maps [2]. We plan to run experiments where the observers use a brush-painting interface to directly mark image regions with the

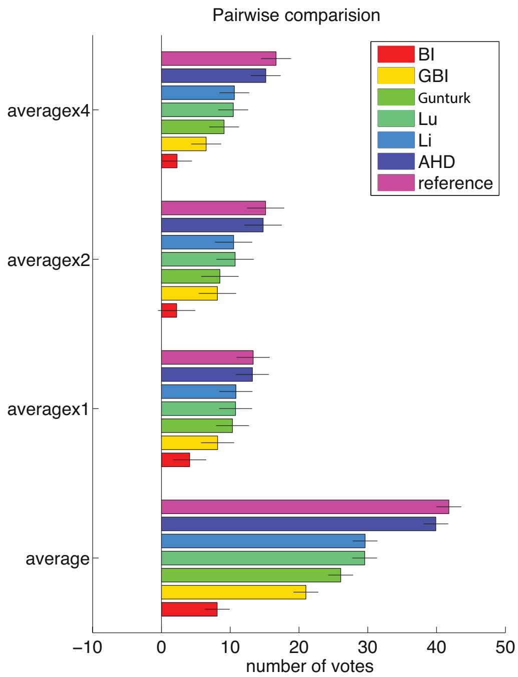

Figure 5: The results of the perceptual experiment. The thin black error bars visualise pair-wise statistical testing. If the two thin bars from two different conditions overlap at any point, the difference between them is too small to be statistically significant.

noticeable distortions in the presence of a high-quality reference image.

References

[1] David H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436(4), 1997.

[2] Martin Čadík, Robert Herzog, Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. New measurements reveal weaknesses of image quality metrics in evaluating graphics artifacts. In *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, volume 31, pages 1–10. ACM, 2012.

[3] Dave Coffin. dcraw: camera RAW file format parser.

[4] Bahadir K Gunturk, Yucel Altunbasak, and Russell M Mersereau. Color Plane Interpolation Using Alternating Projections. *IEEE Trans. Image Processing*, 11(9):997–1013, 2002.

[5] J.F. Hamilton Jr and J.E. Adams Jr. Adaptive color plane interpolation in single sensor color electronic camera, May 1997.

[6] R.H. Hibbard. Apparatus and method for adaptively interpolating a full color image utilizing luminance gradients, January 1995.

[7] Keigo Hirakawa and Thomas W. Parks. Adaptive Homogeneity-Directed Demosaicing Algorithm. *IEEE Trans. Image Processing*, 14:360–369, 2005.

[8] M.A. Laroche and C. A. Prescott. Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients, 1994.

[9] Xin Li. Demosaicing by successive approximation. pages 1–2, 2005.

[10] Rafal Mantiuk, Anna M. Tomaszewska, and Radoslaw Mantiuk. Comparison of four subjective methods for image quality assessment. *Comput. Graph. Forum*, 31(8):2478–2491, 2012.

[11] Yap-Peng Tan Wenmiao Lu. Color filter array demosaiciking: New method and performance measures. pages 1–2, 2003.

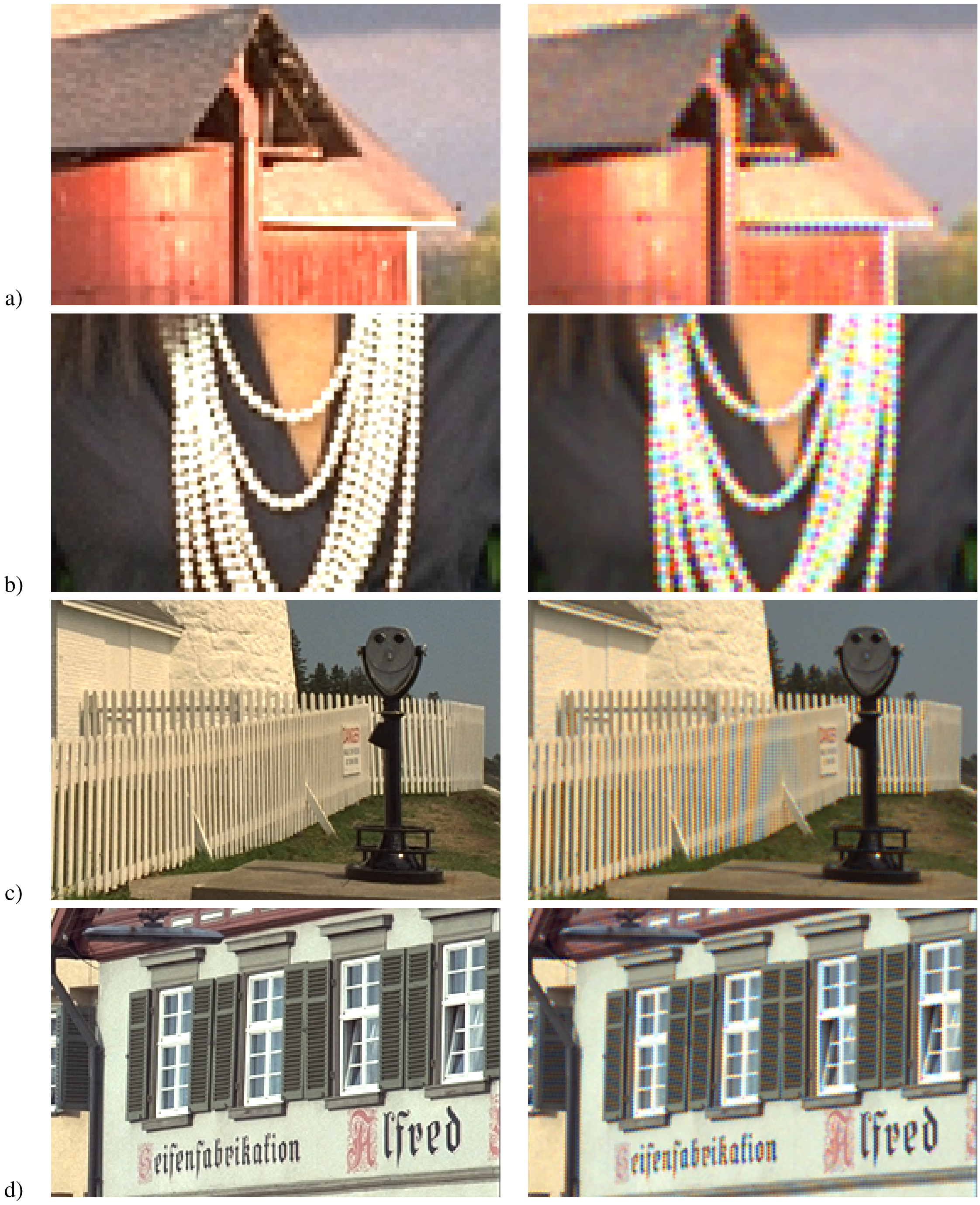

Figure 1: The artefacts generated by the demosaicing algorithms. The rows from top to bottom: a) zipper effect, b) false colour effect, c) Moire effect, d) image blur. Left column: the reference image, right column: the image with the artefact generated by the bilinear interpolation demosaicing algorithm. To better depict the artefacts, the size of the images was upscaled using the NNI (Nearest Neighbour Interpolation) method.

a)

b)

b)

d)

Figure 3: Images and their magnifications used in the experiments.