

CORRELATION-BASED JOINT ACQUISITION AND DEMOSAICING OF VISIBLE AND NEAR-INFRARED IMAGES

Zahra Sadeghipoor¹, Yue M. Lu², and Sabine Süsstrunk¹

¹School of Computer and Communication Sciences

École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

²Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138, USA

¹{zahra.sadeghipoor, sabine.sustrunk}@epfl.ch, ²yuelu@seas.harvard.edu

ABSTRACT

Joint processing of visible (RGB) and near-infrared (NIR) images has recently found some appealing applications, which make joint capturing a pair of visible and NIR images an important problem. In this paper, we propose a new method to design color filter arrays (CFA) and demosaicing matrices for acquiring NIR and visible images using a single sensor. The proposed method modifies the optimum CFA algorithm proposed in [1] by taking advantage of the NIR/visible correlation in the design process. Simulation results show that by applying the proposed method, the quality of demosaiced NIR and visible images is increased by about 1 dB in peak signal-to-noise ratio over the results of the optimum CFA algorithm. It is also shown that better visual quality can be obtained by using the proposed algorithm.

Index Terms— Joint demosaicing, near-infrared (NIR), color filter array (CFA), correlation.

1. INTRODUCTION

Sensors of most digital cameras are inherently sensitive to the near-infrared (NIR) part of the spectrum (700-1100 nm). However, an interference filter called “hot mirror” is usually used in digital cameras to block the NIR information, which is considered noise when capturing visible images. Thus, to capture NIR images with current cameras, one needs to remove the hot mirror and mount a visible-blocking filter in front of the lens [2].

The extra information offered by NIR images has recently been exploited along with visible (RGB) images in some applications such as illuminant detection [3], dark flash photography [4], video conferencing [5], realistic skin smoothing [6], and object segmentation [7].

To take advantage of both NIR and visible information, it is necessary to simultaneously capture NIR and color images for each scene. Currently, two main methods are used to jointly capture the NIR and RGB images. The first approach uses two cameras with a beam splitter [6, 8]. This setup is expensive and inconvenient to use. In the second method, a camera with NIR and visible filters is used to sequentially capture two images [2]. Capturing a pair of NIR/visible images using this method is time consuming and leads to registration issues due to camera movement and scene changes between the two exposures. Contrary to the mentioned methods, our goal is to design a system that captures NIR and color images at the same time



Fig. 1. (a) The luminance and (b) NIR images of a common scene.

using a single sensor. Measuring 4 channels (red, green, blue and NIR) with one sensor, only one color value can be recorded in each spatial position. Thus, a color filter array (CFA) is placed in front of the sensor to subsample the scene during acquisition. To reconstruct the full resolution four-channel image, a post processing step called “demosaicing” needs to be applied on the image. The quality of the reconstructed image is mainly affected by the CFA pattern and the demosaicing algorithm.

To jointly capture the NIR and visible images using one sensor, Lu *et al.* proposed an algorithm to design the CFA and the demosaicing matrices [1]. For the sake of clarity, we call this algorithm hereafter the “optimum CFA algorithm”. As mentioned above, the subsampling strategy (CFA pattern) affects the quality of demosaiced images. Thus, the authors of [1] suggested to design the CFA by solving an optimization problem, which minimizes the error of reconstructing the full resolution visible and NIR images.

In this paper, we propose a regularized optimization problem for NIR/RGB joint demosaicing by modifying the optimum CFA algorithm. Regularization problems have successfully been used in different imaging applications by taking advantage of RGB images’ features (see [9] and references). However, since we address the problem of RGB/NIR *joint* capturing, the regularized optimization problem proposed in this paper takes into account the prior knowledge about the correlation of NIR and visible images.

Despite all the differences between NIR and visible images, NIR and the luminance channel of the RGB image share a considerable amount of information. However, this information is not necessarily perceivable in the spatial domain. For example, in daylight conditions “sky” usually appears black in the NIR image while it is much brighter in the visible image (see Fig. 1 for an example). Since the pixel values of the NIR image and visible luminance channel are generally different, their lowpassed representations are not highly correlated. On the other hand, the general shape of objects and edges

This work was supported by the Swiss National Science Foundation under grant number 200021-124796/1 and the Xerox Foundation.

of the real scene are preserved. Thus, the luminance and NIR images are likely to be more correlated in the high frequency bands rather than in the low frequency band.

To exploit the shared information between NIR and visible images, we computed the NIR/luminance correlation in different frequency bands using shifted versions of the Gaussian filter. Results of our experiment verify that the NIR/visible correlation is highly frequency dependent and that these images are more correlated in highpass frequencies compared to other parts of the frequency space. Therefore, we propose a new design criterion, which minimizes the norm of differences between the highpassed representations of the NIR demosaiced image and the luminance channel of the visible counterpart. The proposed optimization problem also minimizes the norm of estimation error at the same time. Simulation results prove the effectiveness of the proposed algorithm. The quality of NIR and visible images demosaiced by the proposed algorithm is improved by about 1 dB in peak signal-to-noise ratio (PSNR) as compared to the demosaicing results of the optimum CFA algorithm.

2. THE OPTIMUM CFA ALGORITHM

The optimum CFA algorithm is an approach for designing the CFA pattern to jointly capture the NIR and RGB images using a single sensor [1]. It has been assumed that the hot-mirror of the camera is removed and NIR sensitivities of color filters in the CFA are similar. Thus, each pixel of the captured image receives the same amount of NIR radiation. As a result, each pixel of the subsampled image can be written as:

$$y = P_r r + P_g g + P_b b + NIR, \quad (1)$$

where r, g, b and NIR are the intensities of the four channels in the original image and CFA coefficients (P_r, P_g and P_b) are in the interval of $[0, 1]$. Assume that \mathbf{x} is a vector formed by stacking the columns of a block with size $\sqrt{n} \times \sqrt{n}$ in the four channel original image, i.e.:

$$\mathbf{x}^{4n \times 1} = [\mathbf{r}^{n \times 1}, \mathbf{g}^{n \times 1}, \mathbf{b}^{n \times 1}, \mathbf{NIR}^{n \times 1}]^T \quad (2)$$

Let \mathbf{y} be the vector form of the corresponding block in the subsampled image. We can write

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (3)$$

where \mathbf{A} contains CFA coefficients (P_r, P_g, P_b , and 1 for the NIR values). The goal of the demosaicing process is to estimate \mathbf{x} knowing \mathbf{y} . In the case of linear demosaicing, a block in the estimated four channel image ($\hat{\mathbf{x}}$) can be written as:

$$\hat{\mathbf{x}} = \mathbf{D}\mathbf{y} \quad (4)$$

where \mathbf{D} is the demosaicing matrix.

The main novelty of the optimum CFA algorithm is to design mosaicing and demosaicing matrices (\mathbf{A} and \mathbf{D} , respectively) by minimizing the error of estimating the four channel image, which requires solving the following problem:

$$\{\mathbf{A}^*, \mathbf{D}^*\} = \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \mathbb{E} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \quad (5)$$

In the above problem, “E” represents the mathematical expectation. In practice, the expectation can be approximated by averaging the error norm $\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$ over a large ensemble of test images.

Combining (3), (4) and (5) leads to:

$$\{\mathbf{A}^*, \mathbf{D}^*\} = \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \mathbb{E} \|\mathbf{x} - \mathbf{D}\mathbf{A}\mathbf{x}\|_2^2 \quad (6)$$

As shown in [1], solving the above optimization is equivalent to solving the following problem:

$$\{\mathbf{A}^*, \mathbf{D}^*\} = \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \|\mathbf{S}\mathbf{P} - \mathbf{D}\mathbf{A}\mathbf{P}\|_F^2 \quad (7)$$

where \mathbf{P} is the square root of the image correlation matrix (\mathbf{C}), i.e. $\mathbf{C} = \mathbf{P}\mathbf{P}$, and \mathbf{S} is a simple selection matrix (see [1] for more details).

On solving (7), \mathbf{A}^* and \mathbf{D}^* can be used to capture the NIR/visible pair and demosaic the subsampled image.

3. THE PROPOSED ALGORITHM

In this section, we propose a new algorithm to design CFA and demosaicing matrices. As shown in (5), the design process of the optimum CFA method uses the general cost function of most estimation problems, which is the difference of the estimated signal and the original one. One possibility to further improve this algorithm is to explore the prior knowledge about NIR and visible images. In this paper, we use information about the NIR/visible correlation to modify the cost function of problem (5). In the following subsections, first the experimental setup used to quantify the NIR/visible correlation is explained. Then, we present the proposed optimization problem, which takes into account the correlation of NIR and RGB images.

3.1. Correlation of NIR and visible images

To compute the NIR/luminance correlation, we applied N frequency selective filters on NIR and luminance images. Changing the center frequencies of these filters from 0 to 0.5, it is possible to calculate the correlation in the whole frequency space. In our experiment, Gaussian filters are employed to extract image information in different frequency bands, because the shape and filtering characteristics of the Gaussian filter can be effectively controlled by just two parameters (filter bandwidth and center frequencies). A Gaussian filter with bandwidth of σ and center frequencies of f_{cx} and f_{cy} in the frequency domain can be written as:

$$g(f_x, f_y) = \sum_{m=-1,1} \sum_{n=-1,1} \exp - \frac{(f_x - m f_{cx})^2 + (f_y - n f_{cy})^2}{2\pi\sigma^2} \quad (8)$$

Assume that the result of filtering the NIR image by the k^{th} filter is NIR_{g_k} and the corresponding luminance image is L_{g_k} , i.e.:

$$NIR_{g_k} = g_k(NIR), \quad L_{g_k} = g_k(L) \quad (9)$$

where g_k is the k^{th} Gaussian filter and NIR and L are NIR and luminance images, respectively. The NIR/luminance correlation value in the k^{th} frequency band is the Frobenius norm of the $2D$ correlation image of NIR_{g_k} and L_{g_k} , normalized by the Frobenius norm of two images:

$$c_k = \frac{\|corr_{2D}(NIR_{g_k}, L_{g_k})\|_F}{\|NIR_{g_k}\|_F \times \|L_{g_k}\|_F} \quad (10)$$

Note that the correlation should be computed in the spatial domain to take into account the local neighboring relations of images.

The center frequencies of the Gaussian filters in horizontal and vertical directions are equally spaced in the interval of $[0, 0.5]$ to sweep the whole frequency space. In each direction, six center frequencies are selected. Thus, 36 filters are used to compute the NIR/visible correlation. To avoid overlap between consecutive filters, the filter bandwidth in this experiment is 0.05 (the distance of

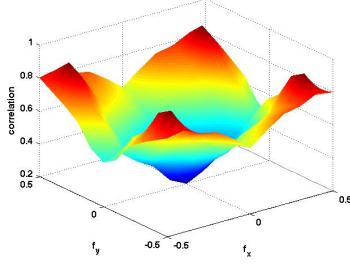


Fig. 2. The correlation of NIR and luminance images in the frequency space.

two successive center frequencies is 0.1). Figure 2 illustrates the correlation values of NIR and luminance images in our dataset for all frequencies. Note that for our dataset, the maximum correlation is obtained when the NIR and luminance images are filtered with the Gaussian filter with $f_{cx} = f_{cy} = 0.4$, where f_{cx} and f_{cy} are center frequencies in horizontal and vertical directions, respectively. This result is used below as prior knowledge about visible and NIR images in the design process of the CFA and demosaicing matrices.

3.2. The proposed optimization problem

We thus propose to modify the optimization problem of the optimum CFA algorithm such that it takes into account the NIR/visible correlation. Here, we exploit the correlation of NIR and spatial (luminance) information by modifying (5) in the following way:

$$\{\mathbf{A}^*, \mathbf{D}^*\} = \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \mathbb{E} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \lambda \|f(\mathbf{NIR}) - f(\hat{\mathbf{L}})\|^2 \quad (11)$$

where $f(\cdot)$ is a frequency selective filter, \mathbf{NIR} is the estimated NIR image, $\hat{\mathbf{L}}$ is the luminance of estimated visible image, and λ is the regularization parameter. Since $\hat{\mathbf{x}} = [\hat{\mathbf{r}}, \hat{\mathbf{g}}, \hat{\mathbf{b}}, \mathbf{NIR}]^T$, it can easily be proved that:

$$\begin{aligned} \hat{\mathbf{r}} &= \mathbf{S}_r^{n \times 4n} \hat{\mathbf{x}} & \hat{\mathbf{g}} &= \mathbf{S}_g^{n \times 4n} \hat{\mathbf{x}} \\ \hat{\mathbf{b}} &= \mathbf{S}_b^{n \times 4n} \hat{\mathbf{x}} & \mathbf{NIR} &= \mathbf{S}_{NIR}^{n \times 4n} \hat{\mathbf{x}} \end{aligned} \quad (12)$$

where:

$$\begin{aligned} \mathbf{S}_r &= [\mathbf{I}^{n \times n}, \mathbf{\Phi}^{n \times 3n}] & \mathbf{S}_g &= [\mathbf{\Phi}^{n \times n}, \mathbf{I}^{n \times n}, \mathbf{\Phi}^{n \times 2n}] \\ \mathbf{S}_b &= [\mathbf{\Phi}^{n \times 2n}, \mathbf{I}^{n \times n}, \mathbf{\Phi}^{n \times n}] & \mathbf{S}_{NIR} &= [\mathbf{\Phi}^{n \times 3n}, \mathbf{I}^{n \times n}]. \end{aligned} \quad (13)$$

Here, $\mathbf{I}^{n \times n}$ is the identity matrix of size $n \times n$ and $\mathbf{\Phi}^{m \times k}$ is the all zero matrix of size $m \times k$. Combining (3), (4) and (12), problem (11) can be re-written as:

$$\begin{aligned} \{\mathbf{A}^*, \mathbf{D}^*\} &= \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \mathbb{E} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \\ &+ \lambda \|f(\mathbf{S}_{NIR} \mathbf{D} \mathbf{A} \mathbf{x}) - f(\frac{\mathbf{S}_r + \mathbf{S}_g + \mathbf{S}_b}{3} \mathbf{D} \mathbf{A} \mathbf{x})\|_2^2 \end{aligned} \quad (14)$$

Using a linear filter, the filtering operator in the above problem can be written in a matrix form, i.e. $f(\mathbf{z}) = \mathbf{F} \mathbf{z}$. Therefore,

$$\begin{aligned} \{\mathbf{A}^*, \mathbf{D}^*\} &= \operatorname{argmin}_{\mathbf{A}, \mathbf{D}} \mathbb{E} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \\ &+ \lambda \|\mathbf{F} \mathbf{S}_{NIR} \mathbf{D} \mathbf{A} \mathbf{x} - \mathbf{F} \frac{(\mathbf{S}_r + \mathbf{S}_g + \mathbf{S}_b)}{3} \mathbf{D} \mathbf{A} \mathbf{x}\|_2^2. \end{aligned} \quad (15)$$

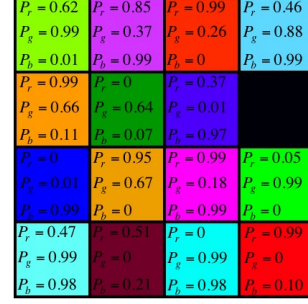


Fig. 3. The 4×4 CFA designed by the proposed algorithm. There is one NIR-only filter in the CFA (black pixel in the second row).

To solve (15), steepest descent and interior point algorithms can be used to change \mathbf{D} and \mathbf{A} , respectively, in the iterative search procedure. Since (15) is non-convex, repeating the search procedure with different initial values can increase the chance of reaching the global minimum.

Important parameters of the proposed algorithm, which have a considerable effect on the performance, are center frequencies of the frequency selective filter (\mathbf{F}). These parameters can be adjusted based on the correlation results presented in the previous subsection. The center frequencies should be chosen such that the filter \mathbf{F} captures the highpass frequencies ($f_x = f_y = 0.4$) of each image, which are the most correlated information between NIR and visible images for our dataset.

The regularization parameter (λ) controls the importance of NIR/visible correlation compared to the error of estimation. This parameter was empirically chosen to be 0.01 in our simulation.

4. EXPERIMENTAL RESULTS

In this section, we present experimental results to show the performance quality of the proposed algorithm and compare it with the optimum CFA algorithm. We used a dataset of 50 NIR/visible images. The images were captured using a two-shot approach, and a registration algorithm was applied on each pair of NIR/visible images to compensate for the camera movements between two shots. All images were downsampled from about 2000×3000 pixels to 512×768 pixels to decrease the effect of remaining registration errors. We used 25 images as training set to compute the correlation matrix and the corresponding CFA and demosaicing matrices. The remaining images were used to test the effectiveness of the obtained CFA and demosaicing matrices.

We used a 2D Gaussian filter with bandwidth of $\sigma = 0.05$ and center frequencies of $f_x = 0.4$ and $f_y = 0.4$ as \mathbf{F} . The 4×4 CFA obtained by the proposed algorithm is shown in Fig. 3. The color of each pixel in this figure is a linear combination of 3 color channels (R, G , and B) and the numbers on each pixel indicate the coefficients of each color in the linear combination. The color coefficients for the black pixel in Fig. 3 are zero after optimization. Since we assume that the hot-mirror is removed, the NIR coefficient in all pixels is 1.

Table 1 summarizes the average CPSNR and PSNR results of the demosaiced images for the visible (RGB) and NIR channels, respectively, using the proposed algorithm and the optimum CFA method.

Algorithm	CPSNR (RGB)	PSNR (NIR)
The optimum CFA algorithm	33.48	34.10
The proposed algorithm	34.30	35.34

Table 1. Average CPSNR and PSNR results for the optimum CFA and the proposed algorithms.

The CPSNR and PSNR values are computed as:

$$\begin{aligned} \text{CPSNR} &= \frac{1}{M} \sum_{j=1}^M 10 \log \frac{255^2}{\frac{1}{3} \sum_{i=1}^3 (\|\hat{\mathbf{x}}_{ij} - \mathbf{x}_{ij}\|_2^2)} \\ \text{PSNR} &= \frac{1}{M} \sum_{j=1}^M 10 \log \frac{255^2}{\|\hat{\mathbf{NIR}}_j - \mathbf{NIR}_j\|_2^2}, \end{aligned} \quad (16)$$

where \mathbf{x}_{ij} is the i^{th} color channel of the j^{th} original image, $\hat{\mathbf{x}}_{ij}$ is the corresponding channel in the j^{th} demosaiced image and M is the number of test images (in our experiment $M = 25$).

The simulation results show that using the proposed algorithm increases the quality of NIR and color demosaiced images by about 1 dB. It should be mentioned that using the optimum CFA algorithm to design the visible-only CFA results in average CPSNR of 38.47 dB for the current dataset.

To illustrate the visual quality of images demosaiced by the proposed algorithm, a pair of NIR/RGB images are shown in Fig. 4. By using the proposed algorithm, less color artifacts are visible in the demosaiced image (tree in the left side of the image) and high frequency parts of the image are retrieved slightly better compared to the optimum CFA algorithm (the right tree in the image). The dataset and more image results can be viewed at http://ivrg.epfl.ch/supplementary_material/SLS_ICIP2011/index.html.

5. CONCLUSION

In this paper, we address the problem of simultaneously capturing RGB and NIR images, which is of significant importance considering the various applications of NIR/visible joint processing. The optimum CFA algorithm has been proposed to design CFA and demosaicing matrices for RGB/NIR joint capturing [1]. We propose a new design algorithm by modifying the optimum CFA algorithm. We propose a new optimization problem that maximizes the correlation of NIR images and the luminance channel of visible images in highpass frequencies. It is justified by our experimental results showing that NIR and luminance images are highly correlated in this frequency band. Simulation results show that our algorithm improves the accuracy of estimating color and NIR images by about 1 dB.

For future work, we will exploit the correlation of NIR and color channels to improve the current algorithm. We will also study the effect of realistic camera sensitivity functions on the performance of the joint demosaicing algorithm.

6. REFERENCES

[1] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk, “Designing color filter arrays for the joint capture of visible and near-infrared images,” in *Proc. of IEEE International Conference on Image Processing*, 2009, pp. 3797–3800.



Fig. 4. The first row shows the original visible and NIR images, the second row represents images demosaiced by the optimum CFA algorithm [1], and the third row illustrates the results of the proposed algorithm.

- [2] C. Fredembach and S. Süsstrunk, “Colouring the near-infrared,” in *Proc. of IS&T/SID 16th Color Imaging Conference*, 2008, pp. 176–182.
- [3] C. Fredembach and S. Süsstrunk, “Illuminant estimation and detection using near-infrared,” in *Proc. of IS&T/SPIE EI: Digital Photography V*, 2009.
- [4] D. Krishnan and R. Fergus, “Dark flash photography,” *ACM Transactions on Graphics, SIGGRAPH 2009 Conference Proceedings*, 2009.
- [5] P. Gunawardane, T. Malzbender, R. Samadani, A. McReynolds, D. Gelb, and J. Davis, “Invisible light : Using infrared for video conference relighting,” in *Proc. of IEEE International Conference on Image Processing*, 2010, pp. 4005–4008.
- [6] S. Süsstrunk, C. Fredembach, and D. Tamburrino, “Automatic skin enhancement with visible and near-infrared image fusion,” in *Proc. of the 18th International Conference ACM MULTIMEDIA 2010*, 2010.
- [7] N. Salamati and S. Süsstrunk, “Material-based object segmentation using NIR information,” in *Proc. of IS&T/SID 18th Color Imaging Conference*, 2010.
- [8] X. Zhang, T. Sim, and X. Miao, “Enhancing photographs with near infra-red images,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [9] D. Menon and G. Calvagno, “Regularization approaches to demosaicking,” *IEEE Trans. on Image Processing*, vol. 18, no. 10, pp. 2209–2220, 2009.