INCOHERENT COLOR FRAMES FOR COMPRESSIVE DEMOSAICING

Abdolreza A. Moghadam[#1], *Mohammad Aghagolzadeh*[#2], *Mrityunjay Kumar*[*3]*, and Hayder Radha*[#4]

 # Dept. of Electrical & Computer Engineering, Michigan State University, * *Eastman Kodak Company*
[1] abdolhos@msu.edu, [2] aghagol1@msu.edu , [4] radha@msu.edu, [3] mrityunjay.kumar@kodak.com

ABSTRACT

In this paper, we explore the notion of using frames to project sensed colors within their inherently 3D space onto a larger number of color basis vectors. In particular, we develop a new frame design, Incoherent Color Frames (ICF), which can include an arbitrary number of incoherent color vectors. An ICF frame possesses key desired properties including the ability to sparsify colors in 3D and to decorrelate color channels utilizing a spatial-frequency selective strategy. We present a low complexity algorithm for constructing ICF frames targeted for the problem of image demosaicing. Our simulation results show that when incorporating the proposed ICF within a Compressive Demosaicing (CD) framework [8], significant visual improvements can be achieved when compared with traditional and Compressed Sesnsing-based demosaicing solutions.

Index Terms— Compressed sensing, Demosaicing.

1. INTRODUCTION

The majority of consumer cameras utilize a Color Filter Array (CFA) that admits, for each image pixel, a single color (e.g., R, G, or B) onto the image sensor. Hence, the captured image is a mosaic of different single-color pixels. This framework made it necessary to develop *demosaicing* algorithms that recover the original three color channels for each image pixel. Due to its impact on millions of consumer cameras, the area of demosaicing has received a great deal of attention over the past decade (e.g., [1]).

Few novel attempts have been made to solve the problem of demosaicing based on solutions developed under the area of Compressed Sensing (CS) [9, 8, 12]. Despite their relative success, recent CS based demosaicing approaches follow the traditional view of 3D representation of the color channels. More recently, a *Compressive Demosaicing* (CD) framework [8] was developed, and which is based on (a) utilizing panchromatic CFAs; and (b) projecting traditional three color channels onto an Equiangular Tight Frame (ETF). The notion of projecting the traditional 3D color channels into an ETF opened the door for a new paradigm for solving the demosaicing problem. In a nutshell, applying an ETF within the 3D color space enables further sparsification, and hence enhances the probability of a CS solver to find the original three-color signal. Nevertheless, and despite being maximally

incoherent, an ETF based approach for demosaicing suffers from several shortcomings and faces key challenges. One key obstacle for using ETF is that only a single non-trivial ETF frame is known to exist in 3D. This single non-trivial ETF in 3D has only six basis vectors; and hence the level of sparsity that can be achieved is rather limited. Furthermore, prior CS-based demosaicing formulations [8, 9, 11] are not sufficiently flexible and do not consider key characteristics of color images (e.g. positive values in RGB color planes) and related attributes of the human vision system (e.g. low sensitivity of human eyes to high frequency chroma components). Thus, the deployed CS-solver might diverge from the correct solution and lead to demosaiced images with artifacts.

Consequently, there is a clear need for developing a more general framework than ETF for representing color images in their inherently 3D space. Such framework should exploit the notion of projecting the three traditional color channels onto an overcomplete sparsifying color dictionary using an arbitrary number of basis vectors while considering key features of the color space and its relationship with the spatial frequency domain. It is crucial to note that a viable CS-based demosaicing solution needs to consider the sparisification of the color image in both (a) the spatial (pixel) domain through an inter-pixel decorrelation transform and (b) the 3D space of the color channels through inter-color channel (or simply "interchannel") decorrelation. This sparsification process (in both domains) can be achieved in a separable manner (e.g., by employing a spatial-frequency independent color transform) or by a non-separable framework. In this paper, we introduce a new frame design that we refer to as Incoherent Color Frames (ICF). ICF can be realized using a separable or non-separable sparsifying framework. However, as we show in this paper, a non-separable strategy (i.e., by employing an ICF frame that is a function of the spatial-frequency) can outperform its separable counterpart significantly. Consequently, the primary focus of this paper is on a non-sparable ICF framework.

It is important to note that employing well known "optimization" methods for constructing tight frames or for achieving minimum coherence projection matrices [10] are not suitable (not even applicable) in the context of the proposed ICF frames, since a set of ICF constraints (outlined later) lead to a highly non-convex problem and any alternating projection algorithm may diverge after a few iterations. Hence, we

present a heuristic algorithm for constructing sub-optimum ICF frames that are tailored for demosaicing. The algorithm attempts to capture the key properties of an ICF (e.g., the necessity of having a luminance component and considering attributes of the human vision system) to achieve good sparsifying results. Our simulation results show that the proposed ICF design provides visibly improved quality for demosaiced images when compared to leading approaches. The remainder of the paper is structured as follows: In Section 2, we introduce the notations used in this paper and briefly describe a CD framework that employs a generic ICF. In Section 3, we present our algorithm for constructing sub-optimal ICF color frames for CD. The simulation results are presented in Section 4 and Section 5 concludes this paper.

2. COMPRESSIVE DEMOSAICING WITH ICF

Throughout this paper, we use the vectorized forms of images (formed by stacking the columns of respective matrices). Assume that for the l-th pixel, the CFA effectively multiplies the red, green and blue intensities by a factor of α_l, β_l and γ_l respectively. Thus, the photosensor for the l-th pixel senses: $y_l = \alpha_l R_l + \beta_l G_l + \gamma_l B_l$, where R, G and B are red, green and blue color planes of the image in the vectorized form. Extending this formulation for the whole image gives:

$$y = \phi [R^T\ G^T\ B^T]^T \tag{1}$$

where $y, R, G, B \in \mathbb{R}^N$ and N is the total number of pixels. Furthermore, ϕ can be defined by means of matrices $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ as $\phi = \begin{bmatrix} \bar{\alpha} & \bar{\beta} & \bar{\gamma} \end{bmatrix}$. Here, $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ are diagonal matrices where $\bar{\alpha}_{i,i} = \alpha_i$, $\bar{\beta}_{i,i} = \beta_i$, $\bar{\gamma}_{i,i} = \gamma_i$ and zero elsewhere. Now the problem of demosaicing can be stated as follows: given samples y and the CFA matrix ϕ in (1), find color planes R, G and B. Note that (1) is an under-determined system of linear equations and one can resort to a CS approach if the RGB vector can be mapped into a sufficiently sparse representation ξ through a sparsifying transform Ψ and if the CFA matrix ϕ leads to a projection matrix $P = \phi\Psi$ with a low measure of coherency. In that case, the sparse representation of the image (block) would be estimated by solving:

$$\hat{\xi} = \arg\min \|\bar{\xi}\|_1 : y = P\bar{\xi} \tag{2}$$

After recovering $\hat{\xi}$, the vectorized forms of RGB color planes of the target image would be reconstructed by $\Psi\hat{\xi}$.

The success of any CS-based demosaicing algorithm or equivalently the quality of the demosaiced image tightly depends on how well the transform Ψ captures intrinsic correlations within color images. Broadly speaking, there are two approaches for designing the sparsifying transform Ψ for natural color images: (a) methods (e.g. [9]) which are inherently oblivious to the distinction between the spatial and color domains; and (b) approaches that are mindful of the distinction between the two types of correlations (spatial and color-channel) [8, 11, 12]. A prime example of the first approach is the one introduced in [9] where trained dictionaries were employed to sparcify color images. Despite of its superior results, such approach relies on image-dependent transforms and hence can suffer from high computational complexities due to the training process. In this paper, we focus on the second approach, which is mindful of the distinction between spatial and color channels' domains, yet it employs fixed non-separable frames that are based on using spatial-frequency dependent ICF's. This strategy lowers the computational complexity and simplifies the demosaicing process while providing high-quality visual results.

Consider signal $x = (x_1, x_2, \ldots, x_l)$ where transform $\Psi^{(i)}$ de-correlates $x_i \in \mathbb{R}^{d_i}$. Then, one can use the *separable* transform $\Psi^{(1)} \otimes \ldots \otimes \Psi^{(l)}$ to de-correlate the vectorized form of x, where $\otimes$ denotes the Kronecker product. Prior works [8, 12, 11] in CS-based demosaicing utilized inter-channel and inter-pixel correlations in this separable form. More specifically, in our work [8, 12], the sparsifying transform Ψ is formed by: $\Psi = \Theta \otimes \psi$, where ψ and Θ capture spatial and color correlations, respectively. This separable formulation, which is similar to some traditional demosaicing methods, attempts to find spatial transform (e.g., DCT or Fourier) coefficients of the image in fixed color spaces (e.g., YUV). In [8], we utilized Equiangular Tight Frame (ETF) along with the luminance (Y) axis in a separable form for Θ and 2D-DCT for the spatial transform ψ. At first glance, this choice seems to be an optimal one since an ETF is the most incoherent redundant frame for a given number of atoms. Although the ETF based solution may work well for image regions with limited colors, a demosaiced image can still have visible color shifts in other regions. The failure of the separable ETF based color frame is due in part to the inflexibility of the separable formulation. Another formulation for $\Psi = \Theta \otimes \psi$ is: $\Psi = \begin{bmatrix} \Theta \otimes \psi_{:,1} & \ldots & \Theta \otimes \psi_{:,N} \end{bmatrix}$ where $\psi_{:,i}$ is the i-th column of ψ. In this formulation, Θ can be considered a frame to decorrelate the triplet $(\hat{R}_i, \hat{G}_i, \hat{B}_i)$, where $R = \psi\hat{R}$, $G = \psi\hat{G}$ and $B = \psi\hat{B}$. Unfortunately, designing a Θ which can sparsify $(\hat{R}_i, \hat{G}_i, \hat{B}_i)$ for all frequencies i, is cumbersome and practically infeasible. This is due the fact that if an image patch has more than one color (which is most likely the case), then the optimum Θ changes for different frequencies as well. Thus, the CS-solver might diverge from the correct solution and hence introduces some artifacts. In this paper, we solve this problem by introducing the concept of utilizing *non-separable* Incoherent Color Frames (ICF). More specifically, we divide the frequency indices $[N] := \{1, 2, \ldots, N\}$ into K groups S_i, $i \in [K]$, where $\bigcup S_i = [N]$. Then for the i-th frequency group, an ICF $\Theta_i \in \mathbb{R}^{3 \times q_i}$ that has q_i color atoms would be utilized. Thus, given the frequency group $\{S_i\}_i$, spatial dictionary $\psi^{(\cdot)}$ and the color coordinates $\{\Theta_i\}_i$, the final sparsifying dictionary is formed by:

$$\Psi = \begin{bmatrix} \Theta_1 \otimes \psi^{(\cdot)}_{:,S_1} & \ldots & \Theta_K \otimes \psi^{(\cdot)}_{:,S_K} \end{bmatrix} \tag{3}$$

Note that in an extreme case, where all Θ_i are the same, then the non-separable formulation (3) turns into the separable for-

mat. Another extreme case might be the case where we design a distinct color frame for each frequency, i.e. $K = N$. In this paper, we consider the case when frequencies are partitioned into three groups ($K = 3$): (a) DC frequency (b) low frequency atoms and (c) high frequency atoms. In the next section, we develop some guidelines for designing the corresponding three ICFs ($\boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Theta}_3$).

3. INCOHERENT COLOR FRAMES

Our proposed design for matrices $\boldsymbol{\Theta} := \{\boldsymbol{\Theta}_i\}_i$ considers the following:

The Luminance Vector for DC: $\boldsymbol{\Theta}_1$ has to include the luminance vector $Y = [\frac{1}{\sqrt{3}} \ \frac{1}{\sqrt{3}} \ \frac{1}{\sqrt{3}}]^T$. This can be justified easily by recalling that the luminance of any color image usually contains the most energy of that image. Inclusion of the Y vector in ICFs clearly contributes to improved sparsity.

Over-complete ICFs for DC and low frequencies: Note that the role of matrices $\{\boldsymbol{\Theta}_i\}$ is two-fold: (a) to make the solution vector (more) sparse and (b) improve the mutual incoherency conditions [5] of the projection matrix $\boldsymbol{P} = \phi\boldsymbol{\Psi}$. Considering that a CD approach would operate on blocks of images and each image block has only a few colors, then $\boldsymbol{\Theta}_1$ should sparsify the average color of any possible image patch. Clearly for a given $\boldsymbol{\Theta}_1$ where $q_1 = 3$, this task is impossible (the same argument holds true for low frequency atoms but not for high frequency ranges). Consequently, a basic requirement for DC and low frequencies is $q_i > 3$ for $i = 1, 2$. Further, these frames have to be incoherent to keep the mutual coherency as low as possible.

The Proper Number of Color Bases: Note that there is a limitation on the number of atoms in color coordinates $\{\boldsymbol{\Theta}_i\}$. First, as we add more and more atoms to $\boldsymbol{\Theta}_i$ (increase q_i), the length of the solution vector ξ increases as well. Since the complexity of most of the CS decoder algorithms are nonlinear as a function of the length of the solution vector, thus the time required for demosaicing increases at a much higher rate as we add more atoms. Second, adding too many colors to an ICF could lead to a degraded color quality or to a saturated quality in the best case (where no improvements on visual quality/PSNR can be achieved), since the sparsity ratio of the solution would fall out of the working region of CS-decoder due to the phase transition phenomenon [7].

In our simulations, $\boldsymbol{\Theta}_1$ (for DC) is a color frame composed of seven incoherent vectors (including the luma axis) in the positive Orthant of $\mathbb{R}^3$. Meanwhile, $\boldsymbol{\Theta}_2$ (for low frequency atoms) contains twelve incoherent vectors in $\mathbb{R}^3$ and for high frequencies, $\boldsymbol{\Theta}_3$ only consists of the luma vector Y. The aforementioned selected color atoms and their numbers in these color transforms have been chosen based on our experience, and hence, they are seemingly (and admittedly) heuristic. However, we can justify such selection of color transforms for different frequency ranges by the following arguments: (a) The DC coefficients (sum of pixel values) in differ-

---

Input: $C, u, l, v, \mathbb{H}, \boldsymbol{\Theta}_i$
Output: $\boldsymbol{\Theta}_i$
for $i=1$ to C **do**
    Generate $W \in \mathbb{H}$;
    if $\max\{W^T\boldsymbol{\Theta}_i\} \in [l, u]$ **then**
        $\boldsymbol{\Theta}_i \leftarrow [\boldsymbol{\Theta}_i \ W]$;
    end
end

Algorithm 1: Finding an Incoherent Color Frame.

---

ent color planes (and hence $(\hat{R}_1, \hat{G}_1, \hat{B}_1)$) are always positive. Hence, the color frame for the DC coefficient has to span only positive Orthant of $\mathbb{R}^3$ and our choice of color frames enables us to span a wide range of colors sparsely. (b) For low frequencies, spatial coefficients $(\hat{R}_i, \hat{G}_i, \hat{B}_i)$, $i > 1$ might take any value (including negative numbers). To sparsely represent all possible triples and meanwhile keep the size of the respective color frame as small as possible, we have to use ETFs with a sufficient number of atoms. However, as stated before in $\mathbb{R}^3$ there exists only one non-trivial ETF with six atoms. Hence, we deploy an incoherent frame (which is the best approximation to an ETF) with twelve atoms. (c) For high frequencies, we have considered the well-known fact that the human visual system is much less sensitive to high frequency chrominance changes. Hence for simplicity of design and by choosing $\boldsymbol{\Theta}_3 = Y$ as the only color coordinate for high frequency atoms, we enforce the demosaiced image to have the same high frequency coefficients for luma and chroma. This also prevents the CS-decoder to choose high frequency chroma atoms that might lead to color artifacts.

Our approach to finding such ICF is presented in Algorithm 1. Bounds l and u dictate the maximum number of attainable atoms. Number of iterations is chosen large enough (e.g. one hundred) to increase the chance of finding all possible atoms. Note that, we set $\max\{W^T\boldsymbol{\Theta}_i\}$ to be higher than l, since otherwise we may end up with a few vectors that occupy a big space in $\mathbb{R}^3$ and prevent adding any further vector to $\boldsymbol{\Theta}_i$. The restriction of $\max\{W^T\boldsymbol{\Theta}_i\} < u$ guarantees that atoms are distinct and meaningful (incoherent).

4. SIMULATION RESULTS

In this section, we present our simulation results for demosaicing of natural images. Due to space limitations, we can only present results for a cropped region of the well-known "Lighthouse" test image. This particular image region has been the choice for virtually all leading demosaicing methods including CS-based ones; and hence we used it for consistency. As clearly demonstrated in Fig.1(b)-1(d), leading traditional demosaicing algorithms such as [4, 2, 3], introduce visible color artifacts. Even though, the spatio-spectral method [3] produces a color image with a very high PSNR Fig.1(d), the reconstructed image still suffers from visible color artifacts. Note that the CS based approach of [9] (Fig.1(e)),
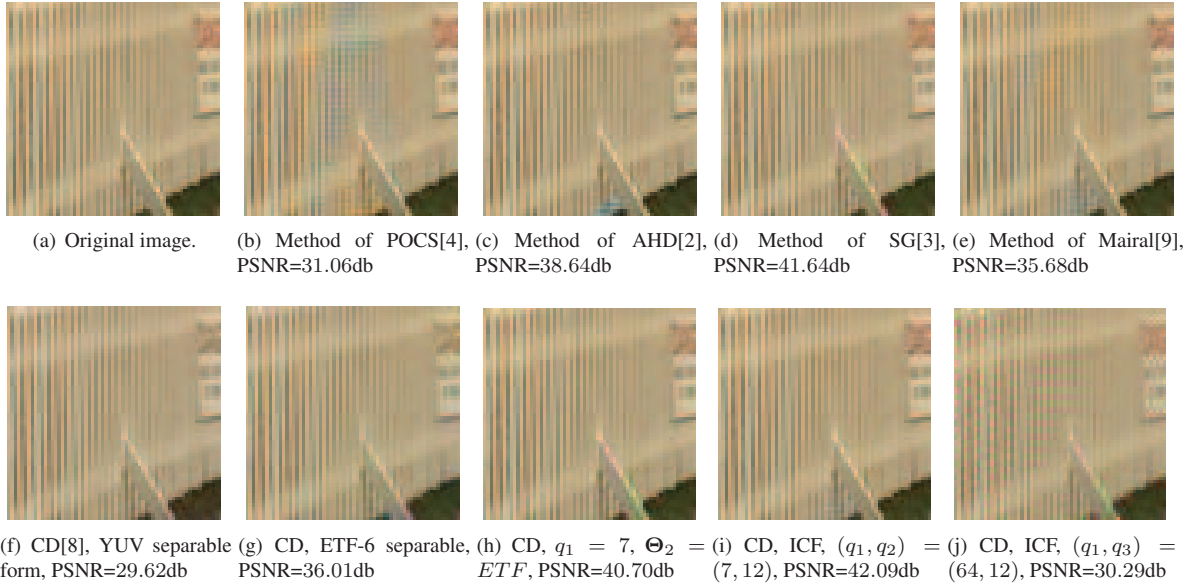
(a) Original image. (b) Method of POCS[4], PSNR=31.06db (c) Method of AHD[2], PSNR=38.64db (d) Method of SG[3], PSNR=41.64db (e) Method of Mairal[9], PSNR=35.68db

(f) CD[8], YUV separable form, PSNR=29.62db (g) CD, ETF-6 separable, PSNR=36.01db (h) CD, $q_1 = 7$, $\Theta_2 = ETF$, PSNR=40.70db (i) CD, ICF, $(q_1, q_2) = (7, 12)$, PSNR=42.09db (j) CD, ICF, $(q_1, q_3) = (64, 12)$, PSNR=30.29db

Fig. 1. A 64×64 cropped region from the "lighthouse". The reported PSNRs are calculated only for the cropped region.

requires some training for deriving an "optimal" sparsifying dictionary (tailored for the specific CFA used) during the demosaicing process, and naturally this increases the complexity of their algorithm. In contrast, the proposed ICF-based CD method is a generic, non-adaptive framework for demosaicing and works on all types of CFAs. Here, we have used a second generation CFA [3] which might not be the optimal one for our CD framework. To investigate the effect of ICFs on the quality of demosaiced images, we also run our algorithm to generate ICFs with different number of atoms to utilize in our CD solver. For instance, Fig.1(f) and Fig.1(g) show the cases where, respectively, color coordinates YUV and ETF-6 are utilized in a separable form for designing the sparsifying dictionary while in Fig.1(h)-Fig.1(j) non-separable ICFs are utilized. More specifically, for Fig.1(h)-Fig.1(j), Θ_1 is an ICF with $q_1 = 7$ atoms where $(u_1, l_1) = (.83, .73)$. For Fig.1(h), an ETF is utilized for Θ_2 while Fig.1(i) illustrates the case where Θ_2 consists of 12 atoms with parameters $(u_2, l_2) = (.82, .72)$. Finally, Fig.1(j) shows an extreme case where $q_1 = 64$ and $(u_1, l_1) = (.99, .89)$. As clearly illustrated in these figures, increasing the number of color atoms in utilized ICFs (q_i) clearly improves the visual quality (and PSNR values) of the demosaiced images. However, after a certain number of color basis vectors, adding more atoms to Θ_is does not necessarily improve the quality and eventually the quality of the demosaiced images may actually degrade in the extreme cases where q_is are relatively large as in Fig. 1(j).

5. CONCLUSION

In this paper, we investigate the design of new non-separable color frames for demosaicing applications. Our results show that utilizing incoherent color frames in a non-separable form for different frequency ranges in the frequency spatial domain, leads to demosaiced image with high visual quality.

6. REFERENCES

[1] B.K.Gunturk, J.Glotzbach, Y.Altunbasak, R.W.Schafer, R.M.Mersereau, "Demosaicking: color filter array interpolation," IEEE SP Magazine, Jan 2005.

[2] K.Hirakawa, T.W.Parks, "Adaptive homogeneity-directed demosaicing algorithm," IEEE Trans. on IP, 14(3):360-369, Mar 2005.

[3] K.Hirakawa, P.J.Wolfe, "Spatio-spectral color filter array design for optimal image recovery," IEEE Trans. IP, 17(10):18761890, Oct. 2008.

[4] B.K.Gunturk, Y.Altunbasak, R.M.Mersereau, "Color plane interpolation using alternating projections," IEEE Trans. IP, 11(9):997-1013, 2002.

[5] D.Donoho, "Compressed sensing". IEEE Trans. on IT, 52(4):1289-1306, Apr 2006.

[6] S.Chen, D.Donoho, M.Saunders. "Atomic decomposition by basis pursuit". SIAM J.Sci.Comp. 20:3361, 1999.

[7] D.Donoho, A.Maleki, A.Montanari, "Message-passing algorithms for compressed sensing", PNAS, 2009.

[8] A.A.Moghadam, M.Aghagolzadeh, H.Radha, M.Kumar, "Compressive Demosaicing", IEEE MMSP 2010.

[9] J.Mairal, M.Elad, G.Sapiro, "Sparse representation for color image restoration," IEEE Trans. on IP, 17(1):5369, 2008.

[10] M.Elad, "Optimized projections for compressed sensing," IEEE Trans. SP., 5(12):56955702, 2007.

[11] P. Nagesh and B. Li, "Compressive imaging of color images," Proc. IEEE ICASSP, 2009.

[12] M.Aghagolzadeh, A.A.Moghadam, H.Radha, M.Kumar "Bayer and panchromatic color filter array demosaicing by sparse recovery", SPIE, Digital Photography, 2011.