Super-Resolution using Image Sequence in Raw Sensor Domain

Mejdi Trimeche and Markku Vehvilainen

Multimedia Technologies Laboratory
Nokia Research Center
Tampere, Finland
mejdi.trimeche@nokia.com

Abstract— **Although the performance of imaging sensors is constantly improving, there are still several physical and practical constraints that limit the final image quality. In this paper, we present a framework for producing a high-resolution color image directly from a sequence of images captured by a CMOS sensor that is overlaid with a color filter array. The algorithm attempts to utilize the additional temporal resolution in order to improve the demosaicing of the color data and filter the noisy and blurred image data. The method is based on iterative super-resolution that performs separately the filtering of the individual color image planes. We present experimental results using synthetic image sequence as well with real data from CMOS sensors.**

I. INTRODUCTION

Lately, the industry trend has been focused to reducing the pixel size in order to improve the spatial resolution. However, because CMOS sensor performance is limited by low quantum efficiency and by dark current non-uniformity [1], this approach leads to reduced sensitivity of the individual pixels and amplification of the noise levels. In fact, regardless of the sensor manufacturing technology, there is a fundamental trade-off between the spatial sampling (number of pixels), pixel size, and the temporal sampling.

Super-resolution (SR) [2], [3], [4] is considered to be one of the most promising techniques that can help overcome the limitations due to optics and sensor resolution. The technique consists in combining a set of low-resolution (LR) images portraying slightly different views of the same scene in order to reconstruct a high-resolution (HR) image of that scene. The idea is to increase the information content in the final image by exploiting the additional spatio-temporal information that is available in each of the LR images.

In most consumer products, the cameras consist of a single imaging sensor that uses Color Filter Arrays (CFA) to sample different spectral components. The most common sampling pattern is called the Bayer-matrix (Fig. 2), which consists of color filter elements arranged such that green component is 50% of the total number of pixels, whereas red and blue components are each 25% of the total amount of pixels. At the location of each pixel, the missing colors must be interpolated from neighboring samples. This color plane interpolation is known as demosaicing, and it is one of the important tasks in the image reconstruction (formation) pipeline.

There has been significant work related to demosaicing of raw color planes [6], [8], [7]. In [7] some promising results were shown using an alternating-projections scheme across the color channels in order to exploit inter-channel correlation. However, this method uses uniquely the spatial correlation of the pixel, without utilizing adjacent frames.

In this paper, we consider the application of super-resolution directly on a sequence of raw images. The idea is to enhance the interpolation of the different color planes by utilizing the data from neighboring image frames. Note that almost all super-resolution methods to date have been designed to increase the resolution of a single monochromatic channel (luminance component). It was not until recently that use of multi-frame processing was considered in the problem of demosaicing raw image data [9], [10], [11]. In [9], it is argued that although a two-pass algorithm (demosaicing followed by super-resolution) improves the overall resolution, this approach results in blurring effects and artifacts similar to those observed in demosaiced images. One benefit from applying super-resolution as a pre-processing step (before cascading into the following operations) is that it ensures that the linear modeling of the problem holds best, at least conceptually. In earlier work [12], we have demonstrated that by applying deblurring directly on the raw color components, we were capable of producing superior results, especially after implementing simple mechanisms to avoid color mismatch.

In the following section II, we define the image formation model that we considered, in section III, we present the solution using super-resolution of color components and we highlight some implementation issues. In section V we show experimental results of the proposed algorithm with synthetic image set, as well as with real sensor data. In the last section, we present some conclusions.

II. IMAGE FORMATION MODEL

Fig. 1 depicts the block diagram of the proposed super-resolution image reconstruction chain. In order to produce the final image, the processing pipeline includes several cascaded operations such as automatic white balance (AWB), gamma correction, contrast enhancement, color filter array interpolation (CFAI), denoising, sharpening. Most of the implementations for these functions are non-linear. Hence, the choice of applying super-resolution as a pre-processing

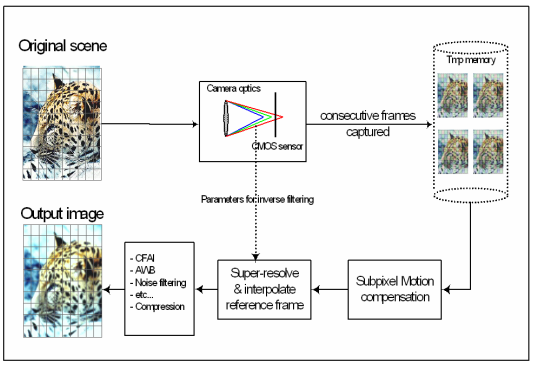


Fig. 1. Image formation (reconstruction) model using proposed integrated super-resolution approach

step (before cascading into the following operations) ensures that the linear modeling of the problem holds best.

In the assumed model, the incoming light is blurred by the camera optics, and the image data is measured by a sensor through the Bayer sampling pattern. The optical blurring and the noise sensitivity of each color channel can be different. We assume linear sensor response as well as linear space-invariant blur for each color channel. We consider N observed raw LR images and assume that these images are obtained as different views of the same scene. Further, we assume that each subsampled color component image is an independent realization of the imaged scene, i. e., for the i^{th} LR image g_i, we process separately 4 channels $g_{i(c)}$, where the index $c = \{1,2,3,4\}$ denotes respectively the data of the *Green1*, *Red*, *Blue*, and *Green2* color channels, as measured according to the Bayer sampling pattern (Fig. 2). Using the assumptions above, the image formation model can be written as:

$$\begin{aligned} g_{i(1)}(x,y) &= S\downarrow(h_i(u,v) * f_G(\xi_i(x,y))) + \eta_{(1)}(x,y) \\ g_{i(2)}(x,y) &= S\downarrow(h_i(u,v) * f_R(\xi_i(x,y))) + \eta_{(2)}(x,y) \\ g_{i(3)}(x,y) &= S\downarrow(h_i(u,v) * f_B(\xi_i(x,y))) + \eta_{(3)}(x,y) \\ g_{i(4)}(x,y) &= S\downarrow(h_i(u,v) * f_G(\xi_i(x,y))) + \eta_{(4)}(x,y) \end{aligned} \quad (1)$$

$f = (f_R, f_G, f_B)$ is the HR reference image corresponding to the imaged scene in RGB domain. h_i denotes the point spread function, or the psf due to optical blurring, $*$ denotes the convolution operator, and $S\downarrow$ the down-sampling operator. Note that in equation (1) each color component is subsampled at a different offset due to the specific pattern of the Bayer sampling matrix. ξ_i is the mapping function corresponding to the geometric warping due to the scene displacement in each of the LR images relative to the HR image f. $\eta_{(c)}$ is an additive noise term that is associated separately with each color channel.

After discretization, the model can be expressed in matrix form as follows:

$$\begin{aligned} \overline{g}_{i(1)} &= \mathbf{A}_{i(1)}\overline{f}_G + \overline{\eta}_{(1)} \\ \overline{g}_{i(2)} &= \mathbf{A}_{i(2)}\overline{f}_R + \overline{\eta}_{(2)} \\ \overline{g}_{i(3)} &= \mathbf{A}_{i(3)}\overline{f}_B + \overline{\eta}_{(3)} \\ \overline{g}_{i(4)} &= \mathbf{A}_{i(4)}\overline{f}_G + \overline{\eta}_{(4)} \end{aligned} \quad (2)$$

The matrix $\mathbf{A}_{i(c)}$ combines successively, the geometric transformation ξ_i, the convolution operator with the blurring parameters of h_i, and the down-sampling operator $S\downarrow$ over the Bayer grid. For notation convenience, we integrate the RGB correspondence in matrices $\mathbf{A}_{i(c)}$, and we express the image model using the following equation:

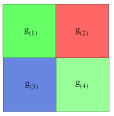


Fig. 2. Bayer matrix sampling pattern

$$\overline{g}_{i(c)} = \mathbf{A}_{i(c)}\overline{f} + \overline{\eta}_{(c)} \quad (3)$$

where $\overline{g}_{i(c)}$, $\overline{f}$ and $\overline{\eta}_{(c)}$ are lexicographically ordered.

In equation (3), each of the LR images $\overline{g}_{i(c)}$ is quarter size of the sensed raw image. The separation of the raw image data into 4 individual subsampled images introduces significant aliasing, since for each color plane we are skipping the pixels from the next color component without low-pass filtering. In principle, super-resolution algorithms exploit the aliasing of the image data in order to reconstruct high frequency information. In the following section, we present the solution for interpolating each color component and combining the information from adjacent frames.

III. Super-Resolution from Raw Data

In this section, we describe the algorithm that performs simultaneously the demosaicing of the color components while fusing the data from the LR frames. The HR image is in RGB domain, while the individual LR images are in sub-sampled raw monochrome domain. The super-resolution reconstruction problem can now be described as estimating for each color channel the best HR image $\overline{f} = (\overline{f}_R, \overline{f}_G, \overline{f}_B)$, which when appropriately projected and down-sampled by the model in (2) will generate the closest estimates of the LR images $\overline{g}_{i(c)}$, $c = \{1 \cdots 4\}$.

A. Cost Function

For each observation $\overline{g}_{i(c)}$, we associate the following cost function:

$$\varepsilon_{i(c)} = ||\widehat{g}_i - \overline{g}_i||^2 = ||\mathbf{A}_{i(c)}\overline{f} - \overline{g}_{i(c)}||^2, \quad (4)$$

where $\widehat{g}_{i(c)}$ is the simulated LR image through the forward imaging model. If we assume that all LR images $(i = 1 \cdots N)$ contribute equally to the cost function, then the overall cost function is the following:

$$\varepsilon_{(c)} = \sum_{i=1}^{N} \varepsilon_{i(c)} = \sum_{i=1}^{N} ||\mathbf{A}_{i(c)}\overline{f} - \overline{g}_{i(c)}||^2 \quad (5)$$

Further, if Gaussian noise is considered in the imaging model $(\overline{\eta}_{(c)})$, then minimizing the cost function in (5) (Least squares solution) is equivalent to the maximum likelihood solution.

B. Iterative Super-Resolution

In order to minimize the error functional in (5), we use the method of iterative gradient descent. This technique seeks to converge $\varepsilon_{(c)}$ towards a local minimum by following the trajectory of the negative gradient. i.e., at iteration n, the high-resolution image is updated as follows:

$$\overline{f}^{n+1} = \overline{f}^{n} + \mu^{n}\overline{r}^{n}_{(c)}. \tag{6}$$

where μ^n is the step-size, and $\overline{r}^n_{(c)}$ is the residual gradient due to the LR color images (c).
The residual gradient $\overline{r}^n_{(c)}$ is computed as follows:

$$\overline{r}^{n}_{(c)} = \sum_{i=1}^{N} \mathbf{W}_{i(c)} \left(\overline{g}_{i(c)} - \mathbf{A}_{i(c)}\overline{f}^{n} \right). \tag{7}$$

The matrix $\mathbf{W}_{i(c)}$ corresponds to $\mathbf{A}^{(-1)}_{i(c)}$, i.e., the inverse process of the image formation. In practice, $\mathbf{W}_{i(c)}$ combines successively the up-sampling and the inverse geometric warp ξ_i^{-1} such that we map the i^{th} LR image grid onto the HR grid.

In the update equation (6), we use the same step size μ^n for all color components. We calculate the step size for a single channel $(c = 1)$ that achieves the steepest descent [5] for that channel:

$$\mu^{n} = \frac{1}{N}\sum_{i=1}^{N} \frac{||\overline{g}_{i(1)} - \mathbf{A}_{i(1)}\overline{f}^{n}_{G}||^2}{||\mathbf{A}_{i(1)}\overline{r}^{n}_{i(1)}||^2}. \tag{8}$$

IV. Implementation

A. Motion Estimation

One critical aspect to achieve efficient implementations of image super-resolution is the need for accurate sub-pixel registration of the input images. The problem of estimating sub-pixel motion from raw data has been investigated in detail in [10]. In our implementation, we used the subsampled components to estimate global projective motion parameters. The motion was estimated for each of the color channel separately, and the resulting motion parameters were finally refined by a simple averaging operation.

Because the overall performance of super-resolution algorithms is particularly degraded in the presence of persistent outliers, we have included a simple mechanism in the motion estimation process that asserts the confidence of the obtained estimates; i.e., if the MSE between the reference frame and the motion-compensated LR image is larger than a given threshold, then we mark that frame so that we skip it throughout the entire reconstruction process.

B. Initialization of Iterative Super-Resolution

It is well known that the iterative Least Squares solution in equation (6) is prone to divergence, especially when the number of input images is limited. If this happens, annoying artifacts start to appear when over-iterated. This is due to the absence of a proper regularization term. To avoid this, we use a smooth initial estimate of the HR image and we limit the number of iterations, especially when we know that the input sequence is noisy. The initial HR estimate is obtained by demosaicing the reference frame (which is used in the motion estimation process) by applying simple bilinear CFAI, and then interpolated to the desired zoom factor, also using bilinear interpolation of all color components.

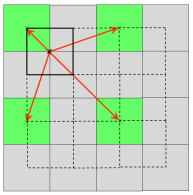


Fig. 3. Pixel projection from interpolated RGB domain (dashed lines) onto a single raw color component (green 1). Note the uneven spacing that is used in the forward and backward projections.

C. Projection Functions

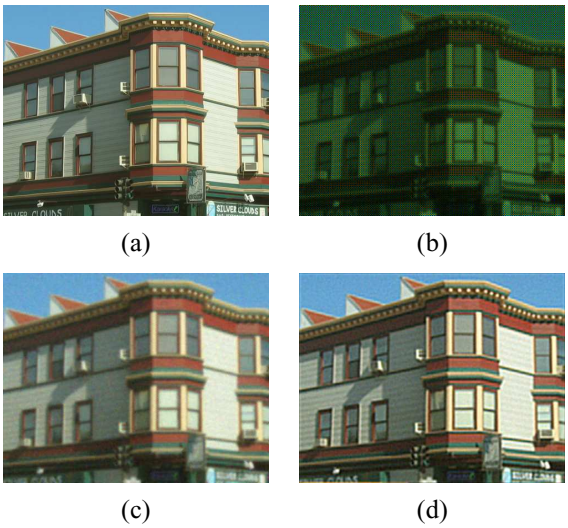
When implementing the image synthesis and the inverse process defined respectively by $\mathbf{A}_{i(c)}$ and $\mathbf{W}_{i(c)}$, we used a process similar to that described in [4]. In the synthesis process, or the forward-projection process, we warp the HR image as point samples and convolve with a continuous form of the point-spread function (psf), finally we downsample at the required positions on the Bayer pattern. We assumed the psf can be approximated with Gaussian function, so that we can easily integrate the blurring as a single parameter in the convolution process. This operation is rather delicate to implement, Fig. 3 shows an example on the assumed positioning of the HR image grid (in dashed line) with respect to the LR image grid. The corresponding half pixel shifts need to be integrated in the motion parameters of each LR image. The inverse or the back-projection process $(\mathbf{W}_{i(c)})$ is handled similarly, instead of downsampling, we interpolate to the required Bayer pattern positions. The "region of influence" of the backprojected pixels is determined by the interpolation Gaussian filter. The smaller the variance of this filter, the sharper the result HR image, however this also means that we need more LR samples (N) to avoid amplification of the noise and the pixelized effect of the solution. A smoother interpolation filter will make a compromise between the number of input images, the noise level and the sharpness of the result.

D. Processing the Green Channel

Another problem that we need to take in consideration is the fact that for each LR image, we have 2 sub-images corresponding to the green spectral component, $\overline{g}_{i(1)}$ and $\overline{g}_{i(4)}$, which we need to correspond to a single channel in the HR image $(\overline{f}_G)$. This can be handeled in many ways, for example, by averaging the backprojected components, i.e., the residual gradients $\overline{r}^n_{(1)}$ and $\overline{r}^n_{(4)}$ corresponding to the green spectral component.

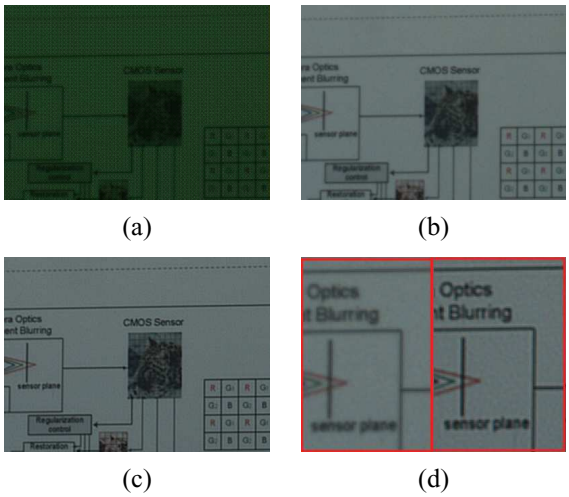
V. Experimental Results

In this section, we present experiments on synthetic and on real sensor data. First, we tested the algorithm on a sequence of synthetic test images. The images, 6 in total, were



(a) (b) (c) (d)

Fig. 4. (a) Original HR image. (b) Example LR image obtained according to model in equation (1), Gaussian psf ($\sigma_{psf} = 1.5$) and additive Gaussian noise ($\sigma^2 = 20$) (c) Image obtained using bilinear CFAI interpolation $PSNR(Y) = 23.87$ (d) Image obtained by applying proposed algorithm, 2 iterations, $PSNR(Y) = 24.68$.



(a) (b) (c) (d)

Fig. 5. (a) Example of raw data captured (4 images) with Micron test camera board (MI SOC1310) (b) image obtained using bilinear CFAI interpolation of reference image (c) Image obtained by applying proposed algorithm, 3 iterations, $zoomfactor = 1$. (d) Close-up comparison between zoomed portions of the images shown in (b) and (c).

generated from a single HR image according to the imaging model described in equation (1). The original HR image was randomly warped using an 8 parameter projective model. We used a continuous Gaussian psf ($\sigma_{psf} = 1.5$) as the blurring operator, and we down-sampled the images by 2 to obtain the 6 LR images. We added to all input images Gaussian noise ($\sigma^2 = 20$). The results of the proposed algorithm after 2 iterations are shown in Fig. 4. Both *PSNR* values and visual inspection confirm the performance of the proposed approach, it is also possible to see from the results that the algorithm is capable of recovering some additional spatial details that were not possible to see from a single interpolated image. The visibility is also improved, since the contrast is enhanced, whereas the colored artifacts due to noise are less visible. In Fig. 5, we have applied the proposed algorithm on a set of 4 images captured by a CMOS camera board (Micron SOC1310). The images were taken, on purpose, slightly out-of-focus to simulate fixed optics system. In Fig. 5(d), we show a zoomed portion to compare the results obtained by applying simple bilinear interpolation against the images obtained using proposed algorithm. Although the parameters that were used in our algorithm were used blindly without exact knowledge of the forward imaging model; the obtained results were satisfactory, i.e. the details were sharper, noise was decreased, and the contrast was much better.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we presented a super-resolution algorithm that takes a sequence of raw color images, and produces a demosaiced color image in RGB domain. Experimental results have confirmed that this approach for image processing is promising, and is capable of producing superior results.

VII. ACKNOWLEDGMENTS

The first author would like to thank Prof. Azzedine Baghdadi for extending his invitation to participate in this special session.

REFERENCES

[1] A. Blanksby and M. Loinaz, "Performance Analysis of a Color CMOS Photogate Image Sensor", *IEEE Transactions on Electron Devices, vol. 47, no. 1, pp. 55-64, Jan. 2000.*

[2] *Park S., Park M., Kang M.: Super-Resolution Image Reconstruction: A technical Overview. IEEE Signal Processing Magazine, Vol. 20, pp. 21-36, 2003.*

[3] *R.C. Hardie, K. Bernard and E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Under-sampled Images", IEEE Transactions on Image Processing, vol. 6, pp. 1621-1632, 1997.*

[4] *D. Capel, A. Zisserman, "Super-resolution Enhancement of Text Image Sequences", Proceedings of the International Conference on Pattern Recognition (ICPR'00), Vol. 1, pp. 1600-1605, 2000.*

[5] *Bertero M., Boccaci P.: Inverse Problems in Imaging, Chap. 6, Institute of Physics Publishing, Bristol, 1998.*

[6] *J. E. Adams Jr., "Design of practical color filter array interpolation algorithms for digital cameras", Proc. SPIE vol. 3028, pp. 117125, Feb. 1997.*

[7] *B. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections",* IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 997-1013, Sep. 2002.*

[8] *O. Kalevo and H. Rantanen, "Sharpening methods for images captured through Bayer matrix",* Sensors, Cameras, and Applications for Digital Photography V, IS&T-SPIE's Electronic Imaging Science and Technology conference*, Jan. 2003.*

[9] *T. Gotoh and M. Okutomi, "Direct Super-Resolution and Registration Using Raw CFA Images", Proc.IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR04), vol. 2, 600-607, 2004.*

[10] *M. Shimizu, T. Yano and M. Okutomi, "Super-resolution under image deformation" Proceedings of the IEEE Conference on Pattern Recognition, Vol. 3, pp. 586-589 , August 2004.*

[11] *S. Farsiu, M. Elad and P. Milanfar, "Multi-Frame Demosaicing and Super-Resolution from Under-Sampled Color Images", Proceedings of the SPIE Conference on Computational Imaging, vol. 5299, pp. 18-22, Jan. 2004.*

[12] *M. Trimeche, D. Paliy, M. Vehvilainen and V. Katkovnic, "Multichannel image deblurring of raw color components" Proceedings of the SPIE Conference on Computational Imaging, vol. 5674, pp. 169-179, Jan. 2005.*