JOINT MULTI-FRAME DEMOSAICING AND SUPER-RESOLUTION WITH ARTIFICIAL NEURAL NETWORKS

Theodor Heinze, Martin von Löwis and Andreas Polze

Operating Systems and Middleware Group, Hasso-Plattner-Institute, Potsdam, Germany

ABSTRACT

This paper introduces an artificial neural networks (ANN) based framework for joint demosaicing of color field array (CFA) raw image sequences. We propose an algorithm that offers superior resolution, signal to noise ratio and dynamic range when compared to single-frame demosaicing. A rich set of both synthetic and real world experimental results illustrates its capabilities.

1. INTRODUCTION

Working in the field of telemedicine for infrastructurally underdeveloped regions [19], our current research considers improving image quality of simple imaging devices to allow for in-field telemedical assistance and machine learning decision support. CFA sensors, especially the ones used in mobile devices like smartphones are limited towards spatial resolution, noise ratio and dynamic range. Furthermore, standard raw processors apply a high degree of noise reduction, which spoils rendition of low contrast detail like human skin. Single-frame CFA interpolation, also called demosaicing, uses only spatial correlation of neighbor pixels, not exploiting information from adjacent frames. A comprehensive review is provided from [6].

A set of neighbor frames with very similar view on the same scene contains information that can be exploited in joint demosaicing. Image characteristics like resolution and noise ratio can be improved. This technique is known as multi-frame super-resolution and is described in [2][8][9][16]. Super-resolution has also been proven to work in the medical context [5]. [11] demonstrates the application of super-resolution to medical imaging.

ANN is a technique from machine learning which is widely used both for classification and approximation problems [1]. ANN needs a part of its parameters to be hard-coded (e.g. network topology, error function) while the other part (connection weights) is learned from examples. Training from examples can be very rewarding in terms of performance. A review on image processing with neural networks can be found in [13]. Approaches to single-frame demosaicing with ANN are described in [10][17][18].

The main contribution of this paper is a new CFA raw multi-frame demosaicing ANN based algorithm which is self-adaptive to the number of frames, noise ratio and exposure fluctuation. It can be trained to render images with a higher pixel and detail resolution than the source frames. We show that it produces high quality output even if compared to state-of-the-art industry raw processors.

2. CONCEPT OF SUPER-RESOLUTION

The goal of the proposed algorithm is to use information of multiple sequential frames to obtain one single frame of higher quality. One important preposition is that the source frames are adjacent observations of a mostly static scene. One of the main ideas of super-resolution is that due to camera movement, for every differentiable visible point of the real world, every captured image frame provides a unique description with slightly differing CFA color surrounding. This fact allows for resolution gain by jointly demosaicing multiple frames into one.

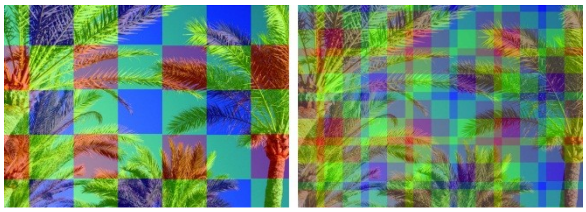

Fig.1. Information gain of an image sequence.

Figure 1 shows how a CFA sensor sees the world on the left. On the right, we overlapped 3 frames, which are slighty shifted to simulate camera movement. Now it is obvious, how much more exploitable demosaicing information an image sequence has to offer. It is also obvious, that the extra spatial color information from the irregularly positioned source frames is very hard to mine with common techniques.

3. DEMOSAICING WITH NEURAL NETWORKS

The joint demosaicing process consists of several major parts. At first, before actual demosaicing, we need to estimate the shift between the source frames, so that we can make exact projections of the source frames onto the target plane [14]. For shift estimation, also called image registration any sub-pixel precision algorithm will do [7][4]. Second, source frame data needs to be processed into input vectors for the ANN. Target images are processed pixel-wise, so for every target pixel image registration information is used to determine related neighbor pixels from every source layer. Third, for every target pixel, the ANN processes the input vector into an output vector with three floating point values between 0 and 1 which contain color information for the R, G and B channels of the target pixel. Last (optional) step is mapping the ANN output to the 0-255 range so that we have 24-bit color information for every pixel. The easiest way to do so is to just multiply ANN output values with 255.

4. ANN TRAINING METHODOLOGY

The core of our approach is training of an ANN to interpolate single pixels from multiple CFA surroundings. This is only practicable with artificial examples from simulated CFA raw data capturing, which has to be performed using an imaging model as precise as possible. The trained ANN can then be used to render images from real raw image files.

When capturing a CFA image from the real world, we basically get only a distorted reflection of reality. To take that into account, we simulate CFA raw data capturing by applying the following transformations to the source image:

1. Downsampling + Shifting.
2. Blur.
3. CFA filtering.
4. Noise addition.

Simulating capturing a series of raw data files means that step 1 (downsampling) needs also to be added a random shift to model camera movement between the shots. The image formation model we used for algorithm design and experiments is largely the same as described in [2] and [9]. Experimental results will show that our modelling assumptions hold very well for real world CFA raw data.

Demosaicing is done pixel-wise, so the goal is training of an ANN to interpolate single pixels from multiple CFA surroundings. For every target pixel, related source pixels from e.g. 5x5 neighborhood mask are collected from every source frame and the ones with the least distance to the target pixel are arranged in the input vector for the ANN. Input vector also includes data like number of source frames and scaling factor of the target image, so that the predicted target image does not need necessarily be of the same size as the source frames. Output values are color values taken directly from the (optionally downscaled) training image. The ANN employed here is a feed-forward backpropagation network with sigmoidal activation function and a sum-of-squares error function [1]. It is trained to predict target pixel color (all three channels simultaneously) from input values which represent neighbor source pixels. Figure 2 is a rough sketch of how input and output values of training samples are generated.

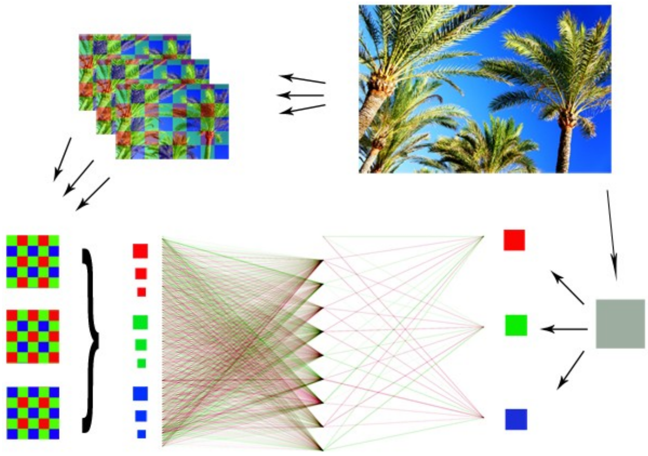

Fig.2. ANN training from artificially generated samples.

To provide the ANN with the capability of accurately guessing target pixels from that input vector, a supervised training must be performed. Training means that for a given set of examples a configuration of ANN weights is determined which produces a minimal error on that training set. Therefore, it is important to provide a wide variety of training samples, including as many different cases as possible. For our case this means that training samples must vary from bright to dark, frequency of detail and noise, saturation, exposure, number of source frames and target scaling factor.

ANN training and testing methodology is firmly aligned with [20]. For the training set, we extracted 160.000 sample pixels from 400 digital images. For testing, we used 32.000 sample pixels from another 80 images. The number of hidden neurons strongly depends on the number of training samples [1]. We used 50 hidden neurons in one hidden layer. Training samples varied from 3-10 input frames and 1-2 upscaling factor. Squared error percentages [20] were 0.251 for training and 0.294 for testing. Therefore, we expect a good generalization performance.

5. EXPERIMENTAL SETUP AND RESULTS

First, we experiment with joint demosaicing of synthetically generated CFA raw files. From one single high-resolution image we derive a series of randomly shifted, downsampled, blurred, CFA filtered images with Gaussian noise (“raw image series”). Figure 3 shows the color image *seals* synthetically decomposed into a series of 10 CFA raw images with random sub-pixel accuracy shift and then rendered bilinearly from one frame (bottom) and by the proposed algorithm with scaling factor 2 (top).

Fig. 3. *seals* example. Proposed algorithm (top) compared to bilinear demosaicing (bottom). Left side pair shows bilinear rendering at its original size and ANN result downscaled, while mid and right pairs bilinear result is upscaled while ANN result stays original size.

Fig. 4. *car* example. Proposed algorithm (top) compared to ACR (bottom).

Figures 4, 5 and 6 are comparisons with real raw data acquired from a Sony® NEX-5 digital camera (NEX) with a 14.2 megapixels 23.4x15,6 mm sized CFA sensor. To simulate rendering from a small and noisy sensor in the *hand* example, NEX was set to very high sensitivity (ISO 12800). Figure 4 shows ANN rendering from 4 raw frames (top) compared to Adobe® Camera Raw® (ACR) default output (bottom). Resolution (Moire, left; EU stars, middle) and noise ratio (right) are slightly better. Figure 5 shows how the proposed algorithm is capable of accumulating dynamic range from source frames with varying exposure. Tonal detail extraction can be steered by tone mapping techniques. Figure 6 shows the rendition of human skin (top) compared to the default output of the NEX (bottom). Here, multi-frame demosaicing from 8 source frames exhibits much better reproduction of low-contrast detail.

Fig. 5. Three source frames with different exposure (left), target frame after tone mapping (right).

Fig. 6. *hand* example. Proposed algorithm (top) compared to demosaicing from NEX (bottom).

6. CONCLUSION AND FUTURE WORK

We have shown experimentally that

1. Joint multi-frame demosaicing with ANN is capable of high-quality results considering resolution, noise ratio and dynamic range.

2. Image processing steps like demosaicing, super-resolution and noise reduction are special cases of neural network processing and thus can be integrated into one single processing step.

With exposure-controlled frames as described in [3] it is possible to achieve super-resolution, low-noise and high dynamic range video. It is also possible to use our algorithm for video upscaling. We are working on a telemedical video-streaming-setup to perform some in-field tests of our framework. One of the problems may be computational performance. Our current prototype, implemented in Java and not optimized at all, renders approx. 50.000 target pixels per second (on a 2.5 Ghz desktop computer), far too slow for real-time video data processing.

7. REFERENCES

[1] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New-York, 2006.

[2] M. Trimeche, "Color Demosaicing Using Multi-Frame Super-Resolution," *European Conference on Signal Processing 2008*, (Eusipco), 2008.

[3] S.-J. Youm, W.-H. Cho, K.-S. Hong, "High Dynamic Range Video through Fusion of Exposure-Controlled Frames", *Proceedings of the IAPR Conference on Machine Vision Applications*, pp. 546-549, 2005.

[4] G. Ward, "Fast, Robust Image Registration for Compositing High Dynamic Range Photography from Handheld Exposures", *Journal of Graphic Tools*, 8(2):17-30, 2003.

[5] H. Greenspan, "Super-Resolution in Medical Imaging," *The Computer Journal* (2009) 52(1): pp. 43-63, 2008.

[6] X. Li, B. K. Gunturk, and L. Zhang, "Image Demosaicking: A Systematic Survey," *Proc. of SPIE Electronic Imaging Conference*, vol. 6822, 15 pages, 2008.

[7] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing vol.21*, 11 (2003), p. 977-1000 (2003).

[8] P. Vandewalle, K. Krichane, D. Alleysson & S. Süsstrunk, "Joint Demosaicing and Super-Resolution Imaging from a Set of Unregistered Aliased Images," *Proc. IS&T/SPIE Electronic Imaging: Digital Photography III*, Vol. 6502, 2007.

[9] S. Farsiu, M. Elad, & P. Milanfar, "Multi-Frame Demosaicing and Super-Resolution of Color Images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141-159, 2006.

[10] O. Kapah and H.Z. Hel-Or, "Demosaicing using artificial neural networks," *Applications of Artificial Neural Networks in Image Processing V (SPIE Vol 3962)*, San Jose, CA; 2000; pp. 112-120.

[11] M.D. Robinson, S.J. Chiu, J.Y. Lo, C.A. Toth, J.A. Izatt and S. Farsiu, "Novel Applications of Super-Resolution in Medical Imaging," *Super-Resolution Imaging, Peyman Milanfar (Editor).* (CRC Press, pp. 383-412, 2010).

[12] H. Takeda, S.Farsiu and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Transactions on Image Processing*, 2007, Vol. 16, pp. 349-366.

[13] M. Egmont-Petersen, D. de Ridder and H. Handels, "Image Processing with Neural Networks – a Review," *Pattern Recognition*, Vol 35, Issue 10, 2002. pp. 2279-2301.

[14] M. Gevrekci, B. K. Gunturk, and Y. Altunbasak, "Restauration of Bayer-sampled Image Sequences," *Oxford University Press, Computer Journal*, vol. 52, no. 1, pp. 1-14, January 2009.

[15] Ho-Gun Ha, In-Su Jang, Kyung-Woo Ko and Yeong-Ho Ha, "Robust Subpixel Shift Estimation Using Iterative Phase Correlation of a Local Region," *Color Imaging XIV Proceedings*, vol. 7241.

[16] S.C. Park, M.K. Park, & M.G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, Vol. 20, No. 3. (May 2003), pp. 21-36.

[17] Y. Huang and Y. Long, "Super-Resolution Using Neural Networks Based on the Optimal Recovery Theory," *Journal of Computational Electronics*, vol. 5, number 4, pp. 275-281, 2006.

[18] V.H. Patil, D. S. Bormane and V. S. Pawar, "Super Resolution using Neural Networks," *Second Asia International Conference on Modelling & Simulation, 2008*, AICMS 08, pp. 492-496.

[19] T. Heinze, R. Wierschke, A. Schacht, & M. von Lowis, "A Hybrid Artificial Intelligence System for Assistance in Remote Monitoring of Heart Patients," *Hybrid Artificial Intelligent Systems Proceedings (2) 2011*, pp. 413-420, 2011.

[20] L. Prechelt, "PROBEN1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules", *Technical Report 21/94*, Faculty of Informatics, University of Karlsruhe, 1994.