Theodor Heinze
Hasso-Plattner-Institute for Software Systems Engineering
Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany
theodor.heinze@hpi.uni-potsdam.de

ABSTRACT
This paper introduces a robust super-resolution algorithm for joint color multi-frame demosaicing. We show that our algorithm, although fast and simple, exhibits convincing results not only within the modeling assumptions, but also for real raw data series. The ultimate goal is its application to telemedical patient monitoring through mobile devices with limited computing power and low quality imaging devices.

KEY WORDS

1. Introduction
Recent developments in the field of Color Field Array (CFA) imaging sensors reveal a trend towards high frame rates while resolution of sensor / lens seems to converge towards physical limitations. Usage of mobile devices like smartphones, which often offer primitive but high-framerate video functionality, spreads rapidly. Working in the field of telemedicine for infrastructurally underdeveloped regions [12], our current research considers improving image quality of simple imaging devices to allow for in-field telemedical assistance [11] and machine learning decision support. Imaging processors of current mobile devices use excessive noise reduction in the demosaicing process [2] to improve visual appeal of the data coming from tiny, noisy sensors. This leads to smearing of low contrast parts of the image thus awkwardly rendering important details like human skin, which is highly undesirable in the telemedical context.

One approach to improving image quality is multi-frame super-resolution [9]. The idea is to use information of multiple sequential frames to obtain one single frame of higher quality. Such algorithms have been introduced already [1] [7] [10], mostly with maximum a posteriori (MAP) estimation techniques. In [8] an algorithm based on the projections onto convex sets technique is proposed, which uses single-frame demosaicing before applying super-resolution steered by several constraint sets. In [4], the authors experiment with normalized convolution interpolation. The main contribution of this paper is a new multi-frame super-resolution algorithm which is highly adaptive to frame-rate, sensor sensitivity and computation time. We will show experimentally, that our algorithm performs well within modeling assumptions, as well as in real world conditions. Our further research will include optimizations of this algorithm towards rendering of human skin and tests with simple imaging devices.

2. Imaging Model
When capturing a CFA image from the real world, we basically get only a distorted reflection of the reality. To take that into account, the image formation model we used for algorithm design and experiments is largely the same as described in [7] and [1]. Thus, to simulate CFA raw data capturing, we need to apply the following transformations to the source image:

1. Downsampling.
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 compensate for camera movement between the shots.

One of the main ideas of super-resolution is that due to that 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.

Figure 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 slightly shifted to simulate camera movement. Now it is obvious, how much more exploitable demosaicing information an image sequence has to offer. Not only is there an overlapping of different color channels, the color patches are also much finer, which makes it possible to derive a higher resolution image. To make use of that, it is very important to align the image sequence as precisely as possible. The possible image quality gain depends on the number of the frames in the sequence, as well as on shift characteristics. If frames are not shifted at all, then any additional resolution in the joint demosaicing process can not be gained either.

We will refer to the multiple raw CFA frames as “source”, while the goal image we will call the “target”. The target is often, but not necessarily, of higher resolution than the source frames. In fact, its resolution can be chosen freely within our algorithm.

3. Image Registration

The main goal of image registration is to provide transformation information of one series of frames to fit one coordinate system. E.g. one of the frames in the series is picked to be the “pivot” frame. Usually, this is the frame with the least cumulative distance (mean square error sum of all RGB pixels) to the others. Then, for each pivot pixel p(i,j) transformation information (shifting, rotating, warping) towards each other frame is determined. This information is then passed on the demosaicing algorithm.

In our implementation, for speed and simplicity reasons we have only determined the mean frame shift information for x and y axes. For resolution increase, it is highly desirable to estimate shift between two frames not only as an integer value, but as precise as possible. There are many approaches like phase correlation and cross correlation [3] to estimate similarity between two images. Since for our demosaicing algorithm accuracy in shift estimation is crucial, we have implemented the following algorithm to estimate the exact shift between two frames:

```
REPEAT r times
    Pick p random pixels
    Find shift (integer value) based on cross correlation
    Compute mean value (floating point) from all r runs

    The idea is that the r semi-random runs will distribute around the real shift. Of course, estimation accuracy improves with increasing number of runs, while a small r might lead to a very inaccurate estimation. In our experiments we usually set r=100.
```

4. Joint Demosaicing Algorithm

After image registration, joint demosaicing can be performed. The idea is to use the exact distance information between the target pixel and the corresponding source pixels projected onto the target plane to calculate a robust and reasonably precise estimate of the real value. Take a look at the following pseudocode:

```
FOR each pixel (i,j) of target image t
    FOR each color channel {r,g,b}
        make a list of n neighbor source pixels from all source frames (value, distance)
        target value = weighted sum of the neighbor pixel values (closer distance means more weight)
```
For every target pixel, for each color channel a list with some nearest source pixels is made. Shift information is used to determine the nearest source pixels of a particular color for every source frame.

Figure 2. The nearest three red source pixels for target(i=5, j=4) located on frame 1 (pivot).

Figure 3. The nearest three red source pixels for target(i=5, j=4) located on frame 2, which is slightly shifted towards the pivot frame.

Figure 4 illustrates the demosaicing problem. The goal is to estimate the value of the target pixel. We assume that light hitting the sensor follows a certain 2d-distribution and source pixel values correspond to the density of that distribution. Now the goal is to estimate that distribution for the target pixel area. For distribution estimation, there are plenty of known techniques, including machine learning. Since our algorithm needs to be fast, we went for a much simpler approach.

It is easy to configure our algorithm towards speed or noise reduction. A low value for \( n \) would mean fast calculation, while a high number of contributing source pixels would mean better noise reduction. The formula is as follows:

\[
\text{color value} = \frac{\sum_{i=0}^{n_f} (s_i \cdot w(d_i))}{\sum_{i=0}^{n_f} w(d_i)}
\]

with
- \( n \) number of source pixels per frame.
- \( f \) number of frames.
- \( s \) color value of source pixel.
- \( d \) distance of source pixel to target pixel.
- \( w(d) \) weight with which a source pixel contributes to the target pixel value. Function of distance.
Varying the $w(d)$ term allows to configure the algorithm for more spatial resolution by giving the closest source pixel a dominating impact, while considering many source pixels to a nearly equal part will give heavy noise reduction at the cost of a slightly blurred image. Flexible involving of source pixels makes the algorithm very adaptive towards sensitivity, resolution, frame rate and processing speed issues.

5. Experiment Setup

Our experiments consist of two main parts. First, we experiment with joint demosaicing of synthetically generated raw files. From one single high-resolution image ("real world") we derive a series of randomly shifted, downsampled, blurred, CFA filtered images with Gaussian noise ("raw image series"). The goal for our algorithm is then to estimate the shift between the frames (image registration) and jointly demosaic them.

In the second part, we show that our modelling assumptions hold very well for the real world. Therefore, we show joint demosaicing results we gained from a CFA raw image series shot with a Sony NEX-5 camera and compare them to single frame demosaicing results from a single frame a) JPEG straight from the camera and b) Adobe ® Camera Raw ® at default settings. To simulate a small and noisy sensor, Sony NEX-5 was set to very high sensitivity (ISO 12800 and ISO 3200 resp.) in the examples “Hand” and “Color Theory”.

6. Experimental Results

Figure 7 shows a resolution chart (© Cornell University, Stephen H. Westin) synthetically decomposed into a series of raw images with random sub-pixel accuracy shift and then rendered bilinearly from one frame (bottom) and by multi-frame demosaicing with a pixel resolution increased with factor 2 (5 frames, mid). In this and the following examples, top image shows the overall view of the picture. Frame shift information has been determined by the algorithm described in the image registration section. The mean shift estimation error was 0.197 pixels, which means that the mean error in source/target pixel distance estimation was about 1/5 of a source pixel. There is some difference in resolution visible, our algorithm is also definitely more robust towards Moire and color fringing.

Figure 8 shows a comparison with a frame rendered by Adobe Camera Raw (ACR). First pair exhibits a resolution advantage, especially when considering the rendering of the EU stars, from which ACR skipped one completely. Second pair shows Moire patterns in the ACR image and soft detail roll-off in the multi-frame rendering. Third pair, again, shows difference in treatment of fine detail.

Figure 5. Synthetic Experiment Set-Up
One important goal in the telemedical context is improved rendering of human skin. Here with Figure 6, we show that a high frame-rate might compensate for low image quality. Although the raw files contained extreme amount of image noise, our algorithm proved its robustness by accurately rendering important low contrast skin detail. Figure 9 shows the “Color Theory” chart, with kind permission from Paper Leaf Design (paper-leaf.com). Multi-frame demosaicing (left side) vastly improves resolution and color rendition here.

Since our prototype is implemented in Java and not optimized at all, we did not measure computing time systematically. On our 1-Ghz subnotebook it takes 20 seconds to jointly demosaic 5 source frames of 14.2 megapixels into a target with a resolution of also 14.2 megapixels and “slow” settings ($n=9, 5x5$ mask).
Figure 8. Left side: multi-frame demosaicing (4 frames). Right Side: single frame rendered by Adobe Camera RAW (default settings).

Figure 9. “Color Theory”: JPEG from Sony NEX-5 (ISO 3200) on the right versus multi-frame rendering (10 frames) on the left.
7. Conclusion and Future Work

We have introduced a robust and adaptive random multi-frame joint interpolation process. Experimental results show that it is well capable to extract additional information from image sequences when compared to single frames, even if they are processed by a state-of-the-art raw-convertor. Also, its configurable processing speed and image noise adaptivity alleviate its application to mobile imaging. Our future research will include testing with mobile devices which feature limited imaging and computing capabilities. Also, we will research if the algorithm is appropriate for an application to live-streaming in telemedical context [6].

For further image quality improvement, we are currently experimenting with artificial neural networks, which have already been employed for demosaicing [5]. Our ultimate goal is to apply image sequence processing to telemedical decision support issues.

References


