White paper about improving image quality and noise reduction by temporal averaging

Less noise – nice

The following document describes the usage of the image averaging property which can be found in MATRIX-VISION cameras of the mvBlueFOX3 and mvBlueCOUGAR-X families. It requires camera firmware version 2.8.0 for the mvBlueFOX3-1 family and 2.14.2 or newer for the mvBlueFOX3-2 family. Its purpose is to improve image quality by temporal averaging of images. This is fundamentally different to image processing algorithms e.g. low pass filtering or mean filtering of single images which also lead to loss of details in an image.

What it does

Assume that one adds the content of two images together. While this doubles the image brightness it does not double the noise in the image because the noise is (mostly) statistically distributed. Statistics math says the noise level is only increased by the √2, if we add two images. If we divide the image by a factor of two to achieve the same brightness again, one can say that we have lowered the noise level or increased the signal to noise level by 3dB.

Now we look at what happens if we add more images together. We rely on the fact that EMVA1288 already defines quite nice and handy functions and measurement rules for

- Dynamic range
- Signal to noise ratio
- Dark current or dark noise

Dynamic e.g. is defined by the ratio of the maximum signal, which is the full well N_{well} (or better saturation capacity in EMVA1288 terms) and the lowest signal, which is the dark noise n_{dark} (N_{dark} is the dark current) according to the following formula:

$$DNR_{sensor} \sim \frac{N_{well}}{\sqrt{N_{dark}}} \sim \frac{N_{well}}{n_{dark}}$$

Formula 1: Max. DNR of a sensor

Signal to noise is defined by the ratio of the maximum signal again versus the noise, which is the geometric sum of dark noise and signal noise. Physics laws say that signals in nature are influenced by so called Poisson or shot noise, which is simply proportional to the mean value of the signal itself. In bright scenes the shot noise dominates by far so that SNR is simply defined by the √saturation capacity.

$$SNR \sim \frac{N_{well}}{\sqrt{N_{well} + N_{dark}}} \sim \frac{N_{well}}{\sqrt{N_{well}}}$$

Formula 2: Simplified Max. SNR of a sensor

Adding images is like adding saturation capacity. This means for DNR and SNR:

$$DNR_{avg} \sim \frac{avg * N_{well}}{\sqrt{avg * N_{dark}}} = \frac{avg * N_{well}}{\sqrt{avg * n^2_{dark}}} = \sqrt{avg} * DNR$$

Formula 3: Average DNR of a sensor

$$SNR_{avg} \sim \frac{avg * N_{well}}{\sqrt{avg * N_{well}}} = \sqrt{avg * N_{well}} = \sqrt{avg} * SNR$$

Formula 4: SNR increase via averaging

Both Dynamic and SNR increase by the √avg.

This is 20*(log (√avg)) dB

Example: avg=16: DNR and SNR improve by a factor of 4.

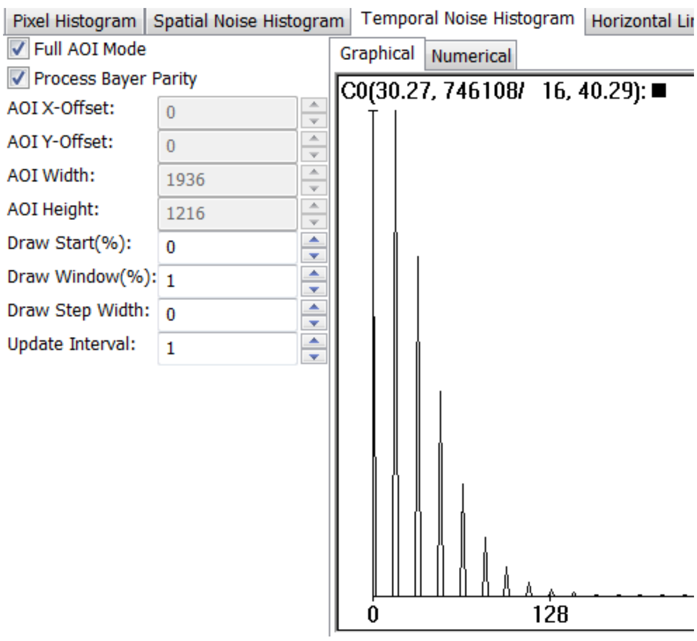
This is 12.04dB or an equivalent of 2 bit more dynamic.

mvBlueFOX3-1xxx series can average up to 16 images in the camera’s internal memory at full bit depth of the sensor. While it calculates and outputs the average image, the sensor images are skipped, so the stream of images which are averaged is discontinuous.

mvBlueFOX3-2xxx series can sum images with an individual limit according to the image size and bit depth (e.g. mvBlueFOX3-2024: 56 images @8bit or 37 images@12 bit) but image intake is continuous, no incoming images are skipped.

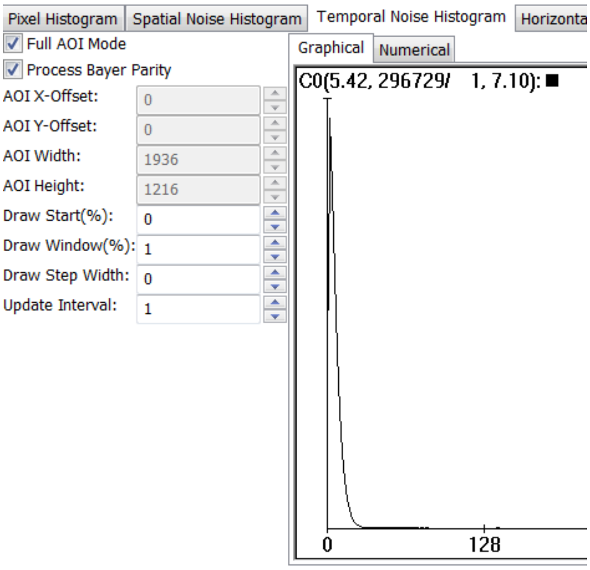
The latter cameras also support a summing output mode. That means that the say 56 images in 8 bit are summed up and output as a 16 bit sum.

The above screenshot shows an image at 42dB gain (via bit-shifting) with no averaging. The standard deviation of the temporal noise (pixel positions from one image to the next) over the whole image is calculated as 40.29 digits (16 bit).



This screenshot shows the same scene after averaging 32 images:

The noise is drastically reduced at no loss in detail. Please note again, that we have 42dB of gain in this image! Consequently, the standard deviation is reduced by a factor of 5.6 =√avg



EMVA1288 measurements of a camera mvBlueFOX3-1013G, using an e2v 76C560 CMOS with a 10 bit ADC sensor show the following EMVA1288 test results:

Average mode	16x	4x	off
Framerate [fps]	3.7	15.1	60.5
SNR,max [dB]	51.9	45.9	39.8
SNR,max [bit]	8.6	7.6	6.6
Dark Noise [DN]	0.7	1.29	2.5
Dynamik Range [dB]	62.7	57.3	51.4
Dynamik Range [bit]	10.4	9.5	8.5

It can be seen that both SNR and DNR are in accordance with the laid-out theory.

We want to answer now the question, what to prefer in the case of a low light application:

Averaging *avg* images at a shorter exposure time *exp* (generating N_{exp} photons) or taking 1 image with exposure time *avg***exp*?

This is answered with the help of the formula for DNR: In darker scenes averaging firstly adds photons while the dark currents are added by the square root.

$$DNR_{avg} \sim \frac{avg * N_{\exp}}{\sqrt{avg * N_{dark}}} = \frac{avg * N_{\exp}}{\sqrt{avg * n^2_{dark}}}$$

Formula 5: DNR increase via averaging

$$DNR_1 \sim \frac{avg * N_{\exp}}{\sqrt{1 * N_{dark}}} = \frac{avg * N_{\exp}}{\sqrt{1 * n^2_{dark}}} = \frac{avg * N_{\exp}}{n_{dark}}$$

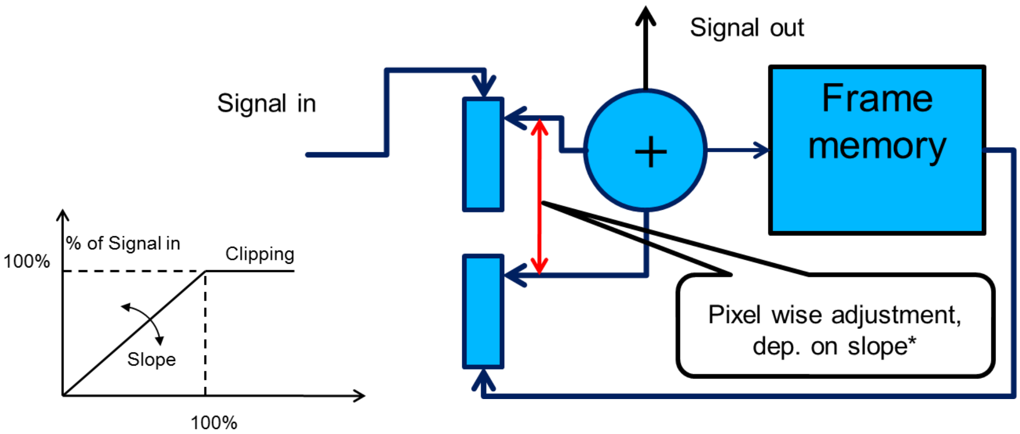
Formula 6: DNR increase via exposure increase

Taking one image at a longer exposure time leads to higher dynamic because the dark noise is to be taken into account only once.

While this averaging works perfectly for stationary images it is obvious that it introduces motion artefacts or blurry images in the case of moving parts in the image.

The last chapter details a principally known enhancement of the averaging, called running average to overcome motion artefacts.

This known principle of an adaptive noise reduction is implemented in the FPGA of several MATRIX VISION CCD cameras of the mvBlueCOUGAR-X type.



Assume that, pixel for pixel, the deviation in grey level of *signal in* versus signal from frame memory is calculated. The incoming signal and the averaged signal are added in an inverse proportional way illustrated by two potentiometers. (The more incoming signal the less averaged signal).

This characteristic is controlled by a slope parameter so that e.g. 10% difference in pixel value leads to say 95% signal in and 5 % of averaged signal. With this parameter one can control noise reduction at the expense of motion artefacts. This approach is very suitable for motion, because it uses the fact that the human eye is less susceptible to noise in moving areas of the image while it is more sensitive to noise in stationary areas. And this is exactly what the camera does: Stationary areas will be denoised because the loop is more closed and less denoised for moving parts because the loop is more open for these pixels.

As a last fact it is worth mentioning that the frame frequency is not lowered by this averaging method and that there is no additional delay because the output is before the memory.

But off course it takes certain frames until the image is denoised.

The following screenshots illustrate that it is possible to reduce noise in stationary image parts while at the same time moving content exhibits only little noise reduction and motion artefacts.

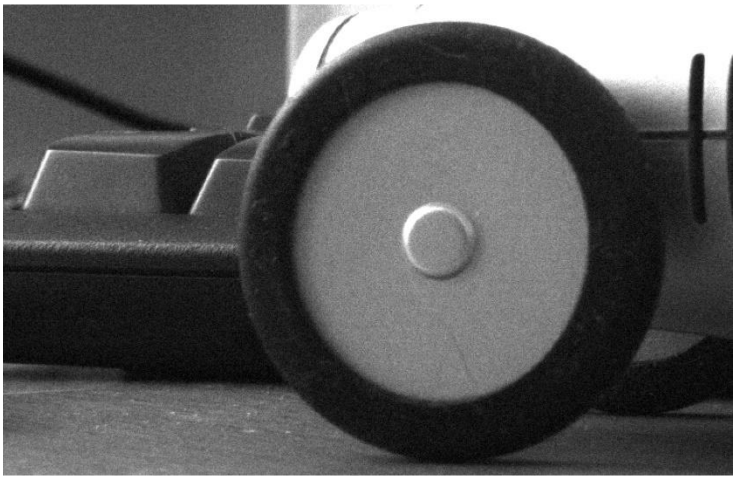
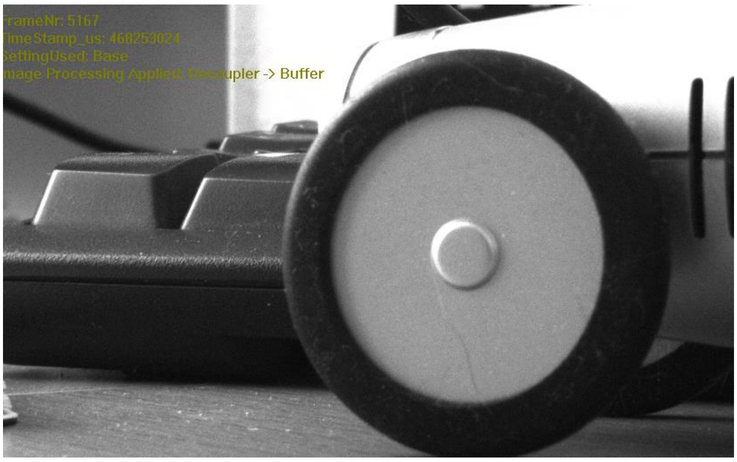
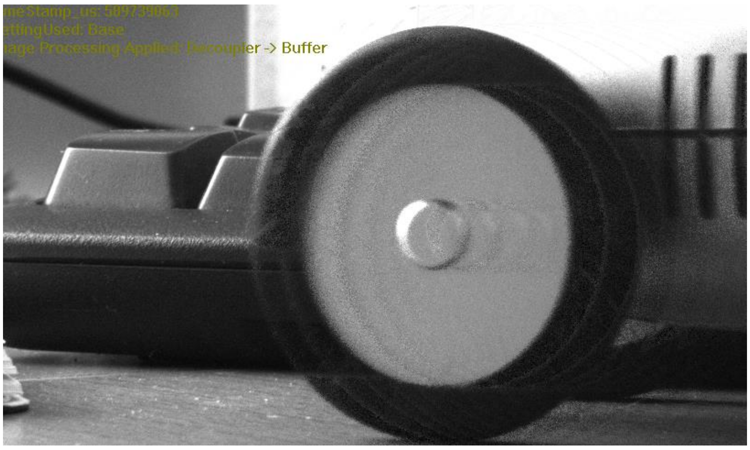


Image with stationary content and high gain; no noise reduction



Same scene denoised (slope 1000)



Scene with wheel moving

Addendum (2021-06-01)

Exposure time vs. averaging / summing

What is the difference in using N-times longer exposure time versus using N-images averaging and use of N-times gain to the images (if the current image is N-times too dark)?

Both methods would result in the same brightness level of the final image.

- First and obvious difference is the frame rate. In averaging mode, we reduce the framerate at least by the number of frames being averaged.

- Second, from image quality perspective, we already learned that temporal noise is reduced by the square root of the number of images being averaged.
  This ends up in a signal-to-noise ratio, which is nearly the same for N-times exposure versus N-times average at N-times gain.
  The main difference is that averaging is also applied to the read-noise of the sensor, which is independent of exposure time. This means, our sensitivity threshold is reduced in either way.

⇨ **Image averaging makes the camera more sensitive, while using longer exposure times will result in larger dynamic range. For low light applications, the latter is mostly less important, so image averaging can be beneficial for such applications.**

Image is N-times too dark:

- Use N-times more exposure time: We get

  SNR_{NxExp}, DNR_{NxExp} and $\mathrm{SensitivityThreshold}_{NxExp}$

- Using N-times gain and N-times image averaging: We get

$$SNR_{NxAvg} \approx SNR_{NxExp},$$
$$DNR_{NxAvg} \approx \frac{1}{\sqrt{N}} DNR_{NxExp},$$
$$SensitivityThreshold_{NxAvg} \approx \frac{1}{\sqrt{N}} SensitivityThreshold_{NxExp}$$

Derivation

SNR

$SNR = \frac{N_{sat}}{\sigma_{sat}} \approx \frac{N_{sat}}{\sqrt{N_{sat}}} = \sqrt{N_{sat}}$, because for Poisson distribution the variance equals the mean.

If we assume, that we only have $\frac{1}{N}$ of the possible saturation level N_{sat} we could use N-times exposure to get N_{sat}:

$$SNR' \approx \sqrt{N'_{sat}}, with\ N'_{sat} = \frac{1}{N} \cdot N_{sat}$$

$$SNR \approx \frac{N \cdot N'_{sat}}{\sqrt{N \cdot N'_{sat}}} = \sqrt{N \cdot N'_{sat}} = \sqrt{N} \cdot SNR'$$

now we add the images (for averaging, the same applies):

$$SNR = \frac{\sum_{n=1}^{N} N'_{sat}}{\sqrt{\sum_{n=1}^{N} N'_{sat}}} = \frac{N \cdot N'_{sat}}{\sqrt{N \cdot N'_{sat}}} = \sqrt{N} \cdot SNR'$$

So, for N-times exposure or N-times averaging/adding, we get same SNR

DNR

$$DNR = \frac{N_{sat}}{\sigma_{dark}}$$

Making the exposure time longer, we get:

$$DNR = \frac{N \cdot N'_{sat}}{\sigma_{dark}} = N \cdot DNR'$$

Adding up the images:

$$DNR = \frac{\sum_{n=1}^{N} N'_{sat}}{\sqrt{\sum_{n=1}^{N} \sigma_{dark}^2}} = \frac{N \cdot N'_{sat}}{\sqrt{N \cdot \sigma_{dark}^2}} = \frac{N \cdot N'_{sat}}{\sqrt{N} \cdot \sigma_{dark}} = \sqrt{N} \cdot \frac{N'_{sat}}{\sigma_{dark}} = \sqrt{N} \cdot DNR'$$

Sensitivity Threshold

$$\mu_{e.min} \approx \sigma_{dark} + \frac{1}{2}$$

With different exposure time, this stays constant, there is no dependency to exposure time.

But if adding up images we get:

$$\sigma_{dark} = \frac{1}{N} \cdot \sqrt{\sum_{n=1}^{N} {\sigma'}^2_{dark}} = \frac{1}{N} \cdot \sqrt{N \cdot {\sigma'}^2_{dark}} = \frac{1}{N} \cdot \sqrt{N} \cdot {\sigma'}_{dark} = \frac{1}{\sqrt{N}} \cdot {\sigma'}_{dark}$$

And therefore:

$$\mu_{e.min} \approx \frac{1}{\sqrt{N}} \cdot {\sigma'}_{dark} + \frac{1}{2}$$

Simulation results:

```
Max Saturation: 10000 e-

Read-Noise: 2 e-

original      : Bright Noise: 27.95 e-, Dark Noise: 2.02 e-, SNR: 93.52 | 39.42 dB, DNR: 1294.49 | 62.24 dB

4 x exposure  : Bright Noise: 58.26 e-, Dark Noise: 2.21 e-, SNR: 180.01 | 45.11 dB, DNR: 4742.76 | 73.52 dB

4 x average   : Bright Noise: 13.96 e-, Dark Noise: 1.01 e-, SNR: 187.28 | 45.45 dB, DNR: 2588.27 | 68.26 dB
```