How to determine camera gain.

First of all, let us determine what is gain and how it relates with ISO. We are all used to the fact that from the times when film was the main light-capturing source that the higher ISO the more sensitive to the light our medium, be it the film itself or camera sensor. However, if in case of the film its ISO value is determined by its chemical composition then for the digital sensors' things are pretty different. ISO change doesn't correspond to the real switch to the more sensitive chip. In our camera we have just one. So, what's the deal?

CCD or CMOS chip is an array of light-sensitive semiconductors that can change their charge value depending on the amount of light that hits them. For our purposes it does not matter whether the charge decreases or increases. The main thing is that the change in charge is linearly related to the number of photons that knock out electrons, changing the charge in each pixel.

Further, the accumulated charge is read, digitized (by the Analog-to-Digital Converter) and written to a file. Unprocessed RAW, or processed jpg. Right before digitization, we can multiply the accumulated charge by a certain factor, and only then digitize it. It would seem, why not multiply it after the ADC. Let's assume that the ADC bit depth is 14 bits (as in most modern cameras). So, we can get 16384 gradations of brightness. If a pixel is capable of accumulating, say, 30,000 electrons, then from the point of view of the ADC, the addition of one electron to a neighboring pixel does not lead to its difference from the first one. In this case, although there is a signal, it turns out to be too weak for digitization. It is especially evident in case of small signal: zero or one electron will be both digitized to zero. Therefore, one needs to multiply the accumulated charge BEFORE passing through the ADC (which, quite inopportunely, also introduces its own small noise into the image). So, the choice of sensitivity is just a change in that multiplicative factor or gain.

Setting the ISO to its minimal value is beneficial since the dynamic range is at its maximum, and this is understandable — we use the entire available depth of the potential well in which we store electrons. However, when trying to brighten the image in post-processing, the structure of the readout noise in the darkest areas will come out in all its glory. By increasing the ISO/gain, we reduce the depth of the well, however, we no longer have to pull up the brightness so much in the editing software, which means that the effect of readout noise decreases[1] with increasing ISO. What should we do? At low ISO, we are waylaid by the readout noise, and at high ISO the dynamic range drastically drops. Obviously, some kind of compromise is needed. This role is frequently played by the native ISO, where the gain is close to unity and each electron gives one unit of the ADC.

How can we find this magic coefficient, and the corresponding magic ISO? The cloudynights website describes the technique, but does not explain how it arises from the physical principles. Here I will try to fix this annoying gap, since I would like not only to blindly follow the recipe, but also to understand what is behind it.

So, let's imagine that we have at our disposal some "black box" — a camera. Which the received light (photons knocking out electrons) multiplies by some unknown coefficient and digitizes the resulting value. We want to find out this coefficient in some way. The so-called "photon noise" will help us with this.

[1] It will be possible to determine the readout noise in absolute units — electrons, by measuring the amount of noise on a frame with a bias current and dividing it by the gain.

A light source with a fixed brightness emits a certain number of photons per second, and on average it will be a constant flux. However, if we take an arbitrary time interval, the number of photons will randomly fluctuate around a certain average value. The magnitude of this fluctuation, according to the Poisson distribution, is equal to the square root of the mean value.

So, we need a light source with a brightness which remains constant during the measurement process and which we can predictably control. Monitor screen with a uniformly filled background is perfectly suited to this task. We will systematically shoot it with a close-placed long-focus lens. Additionally, you can defocus the image so that the pixels of the screen certainly do not show up in our photo.

Suppose we took a couple of photos of our improvised flat field, subtracted the master-bias frame (30-40 frames taken with the lid closed with a minimum shutter speed and averaged would be enough) from them and chose an area free from dust, gradients and other shenanigan. What information do we have now, and what can we do with it?

As I said above, the number of electrons depends on the number of incident (and registered) photons multiplied by the gain, i.e. for two frames we have:

$$N_1 = ke_1 \quad N_2 = ke_2$$

Due to presence of the photon noise average value of e_1, will fluctuate around $\pm\sqrt{e_1}$, and so will e_2. It is convenient to determine this variation by subtracting N_1 from N_2, obtaining:

$$d = N_1 - N_2 = k(e_1 - e_2)$$

Since on average $N_1 = N_2$ our difference gives us standard deviation value with no offsets. Which is by definition, for the error of indirect measurement of two quantities is:

$$\Delta d = \sqrt{\left(\frac{\partial d}{\partial e_1}\right)^2 (\Delta e_1)^2 + \left(\frac{\partial d}{\partial e_2}\right)^2 (\Delta e_2)^2}$$

One can see those partial derivatives equal $\pm k$. And by remembering that the fluctuation is equal to the square root of the mean value we get:

$$\Delta d = \sqrt{(k)^2 e_1 + (k)^2 e_2} = \sqrt{k(ke_1 + ke_2)} = \sqrt{k(N_1 + N_2)}$$

On the other hand — mean of two frames is equal:

$$Av = \frac{N_1 + N_2}{2}$$

It is obvious that we can obtain our desirable k:

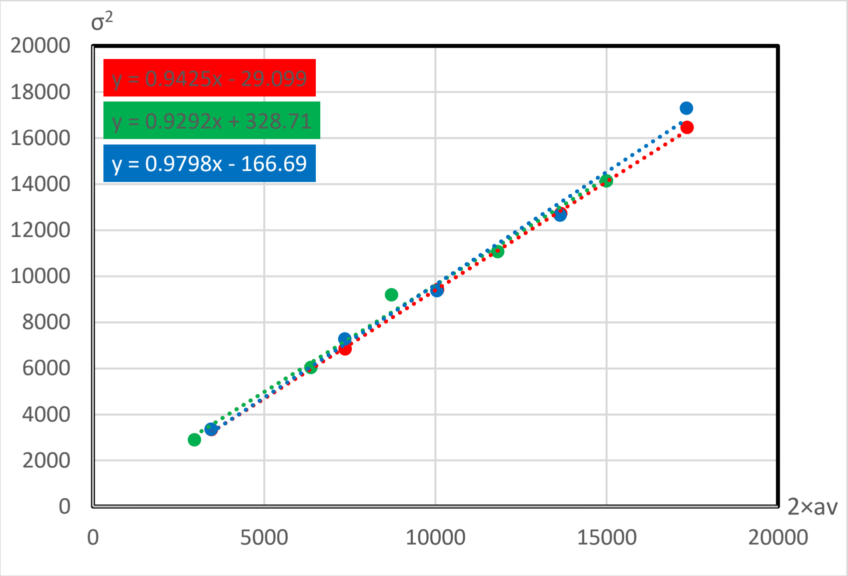
$$\frac{(\Delta d)^2}{Av} = \frac{2k(N_1 + N_2)}{(N_1 + N_2)}$$

Or simplifying:

$$k = \frac{(\Delta d)^2}{2Av}$$

So, we subtract one image from the second, determine the standard deviation and square it, and then divide that result by twice the average.

The more such independent image pairs we get at different monitor brightness, the more accurate our gain estimate will be. And the value equal to $1/k$ will give us the number of electrons per ADC unit. Linear regression can be applied to further increase the accuracy. Set the monitor to its maximum brightness and get a couple of frames with the longest shutter speed, such that there is no overexposure in any of the channels yet (it’s better to navigate not by the camera’s histogram, but by real values, controlling them in your analyzing program of choice). After that, reduce the brightness of the monitor and continue to make pairs of exposures without changing the parameters throughout the session. Plot the resulting averages on the x-axis, and the standard deviations on the y-axis and determine the slope factor: this is how we use the least squares to solve the redundant system of equations to increase accuracy. For a native ISO, you should get a picture like this:



k ~ 1, as we were looking for. Knowing the gain, it is no longer difficult to determine the number of electrons corresponding to saturation of the matrix: we divide the maximum count by the gain. Apparently, the full capacity of a pixel must be sought at the minimum available ISO.

It should be noted that for these measurements it is better not to convert images to color, but to use rggb channels by decomposing the RAW file into the corresponding channels immediately after calibration, so that the debayerization algorithm does not distort the original data with its interpolation mechanisms.

A small afterword. From my point of view, the use of native ISO is justified when shooting high-contrast scenes. When we want to maximize the exposure of dark objects in the frame without overexposing the brightest ones. If there is a limitation in shutter speed, and there are no bright objects on the frame, or they are too small, then, in my opinion, it is more expedient to use higher ISOs. The readout noise continues to decrease with increasing sensitivity, and this, in turn, has a beneficial effect on the transmission of details in the shadows. In modern cameras however after certain ISO value readout noise stops changing that much. And it means that further ISO increase will not help you to preserve faint details but only cuts your dynamic range.

In fact, as always, everything (well, at least the read noise) has already been done before us, and the answer at the end of the textbook can be peeped here.

Anton Chechkin. (4D on astronomy.ru)
2016.11.12
Translated in English 2022.05.18