Module 1.1

OpenCV Basics

Photos and Color

Satya Mallick, Ph.D.
LearnOpenCV.com

Table of Contents

What is a photo?

A photo is simply light reflected by an object and captured by a camera. A grayscale camera records the intensity of reflected light. A color camera records the light reflected at three points of the visible spectrum -- Red (650 nm), Green (510 nm) and Blue (475 nm).

What is color?

Color is a fascinating subject. As kids, we were taught that the Rainbow has all the colors in the world. That is absolutely wrong!

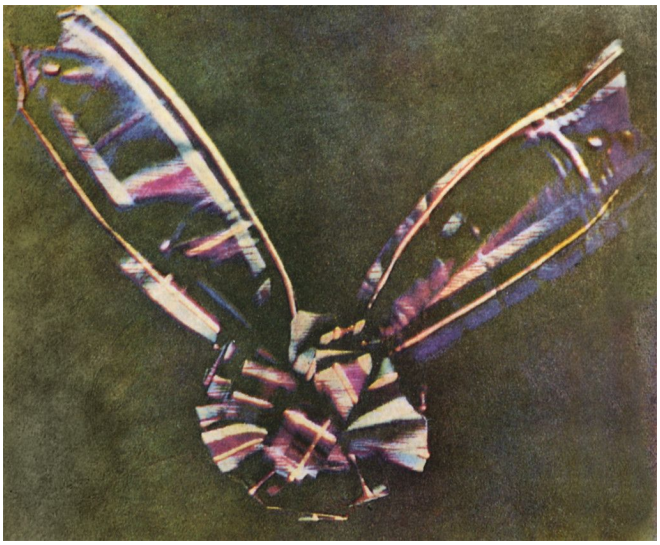
> "A Rainbow does not contain all colors humans can perceive!"

In fact, the color **pink is not present in the rainbow**! It does not correspond to any wavelength of light. Pink is a mixture of Blue and Red.

Color is, therefore a mechanism of the human eye by which it perceives a mixture of lights of different wavelengths.

A brief history of color photography

The idea that you can take three different photos using three primary color filters (Red, Green, Blue) and combine them to obtain a color image was first proposed by **James Clerk Maxwell** in 1855. Six years later, in 1861, English photographer **Thomas Sutton** produced the first color photograph by putting Maxwell's theory into practice. He took three grayscale images of a Ribbon (see image on the right), using three different color filters and then superimposed the images using three projectors equipped with the same color filters. The photographic material available at that time was sensitive to blue light, but



not very sensitive to green light, and almost insensitive to red light. Although revolutionary for its time, the method was not practical.

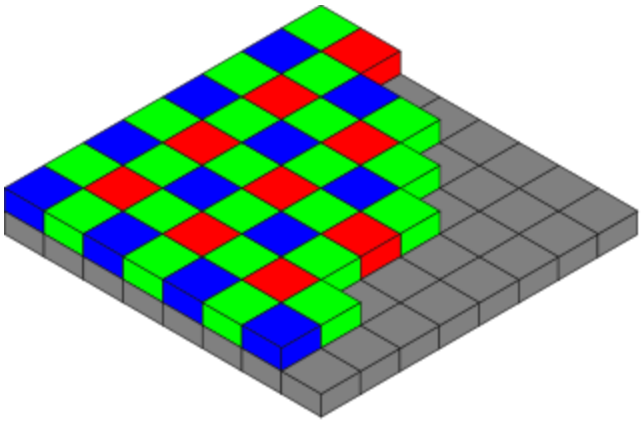
By the early 1900s, the sensitivity of photographic material had improved substantially, and in the first decade of the century, a few different practical cameras were available for color photography. Perhaps the most popular among these cameras, the **Autochrome**, was invented by the **Lumière brothers**.

How is a color recorded by a digital camera

Inside your digital camera is a sensor that is sensitive to visible light (photons). The sensor has many photosites or pixels arranged in a rectangular grid. These photosites produce an electrical charge directly proportional to the amount of light (photons) it receives.

Depending on the scene we are recording, different pixels of the sensor are exposed to different amounts of light and we will see a pattern of electrical charge. The amount of electrical charge is converted to a number between 0 and 255. This corresponds to 8-bits or 256 values.

This rectangular matrix of numbers between 0 and 255 is a digital **grayscale** image.



Color sensors are much more complicated. Every pixel in the sensor records only one color -- Red, Green or Blue. The pattern shown [image credit] on the side is called the **Bayer Pattern**. Notice there are twice as many green pixels compared to blue and red pixels because the human eye is much more sensitive to the green light compared to the red and or the blue light. The two missing channels at every pixel are calculated by interpolating the values from neighboring pixels. This process is called demosaicing**.**

The story does not end there. **Our eyes are not linear sensors.** If you are staring at a dimmed light source and you double the intensity of light, the perceived change in intensity will be much less than 2x. To mimic this behavior, cameras apply a nonlinear transform to the recorded image intensity. This transformation is often approximated by a single parameter function called the **gamma.** Can you get access to the linear image (i.e. the image without gamma correction) the sensor records? Yes, in many SLR cameras you have the option to record the image in the **RAW** format which is nothing but the linear image with the bayer mosaic intact. RAW images also contain more than 8 bits per channels.

How are images stored on disk

As mentioned in the previous section, once the image is read out from the sensor, it is demosaiced and usually converted to an RGB image. Usually, this array of RGB pixels is compressed to JPG or PNG formats before storing on disk. Most image formats, like JPG, have two parts

1. **Image Header:** This part contains image metadata like the width and height of the image, number of channels, the color profile, the number of bits per pixel and so on and so forth.
2. **Data:** This part contains compressed RGB values.

What is OpenCV?

OpenCV is the most popular Computer Vision library in the world with an estimated 14 million downloads. The following qualities make it an excellent library of choice for building commercial Computer Vision applications.

1. **Highly optimized:** OpenCV is written in C/C++ with the goal of building real-time applications. It is highly optimized when compiled with the appropriate options and can utilize multiple cores on your machine. It is also capable of utilizing the heterogenous computing resources (e.g. a GPU) when compiled with OpenCL support.
2. **Open sourced under the BSD license:** It is open source and licensed under the very permissive BSD license. This means you can use it to build commercial applications and do not need to open source your own code. However, there are parts of OpenCV (e.g the opencv_contrib module) that may or may not be under the BSD license.
3. **Language bindings:** It is written in C/C++ with bindings for other languages including Python and Java.
4. **Portability:** It supports Linux, Mac, Windows, iOS and Android operating systems.

How are images represented in memory in OpenCV

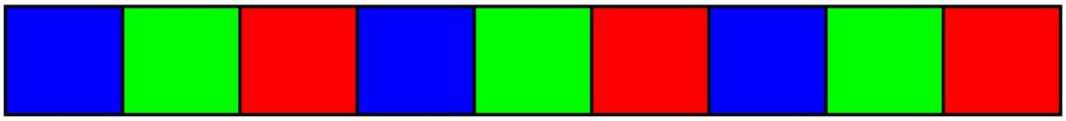
When we read an image using OpenCV, it first decompresses the image, and stores all images in a standard format.

OpenCV Images in C++

In the C++ version, images are instances of the **Mat** class. The **Mat** class has two parts

1. **Image Header:** This part contains image metadata like the width and height of the image, number of channels, the number of bits per pixel and so on and so forth.
2. **Data:** This part is an array of uncompressed **BGR** values in a row major format. Why not RGB? OpenCV stores images in BGR format for historical reasons.

The first three bytes in this data array correspond to the blue, green and red channels in the top left corner of the image. The next 3 bytes correspond to the pixel in the second row and first column of the image. All the pixels in one column of an image are stored as a continuous block in memory. The image below shows 3 pixels.



Note: While pixels in one column of an image are physically stored in one continuous block in the RAM, all the pixel values are not always in a continuous block of memory. We will never have to worry about this detail in this course. However, if you ever implement a new algorithm from scratch and have to efficiently access the image pixel by pixel, you have to keep this detail in mind.

OpenCV Images in Python

In the Python version, an image is a multi-dimensional NumPy array. If you are not familiar with NumPy, we have included a tutorial in Module 1.2.2 of this course.