

# Introduction to Image and Video Processing

**course:** Image and Video Processing

**week:** 1

**Date:** 2013-06-07

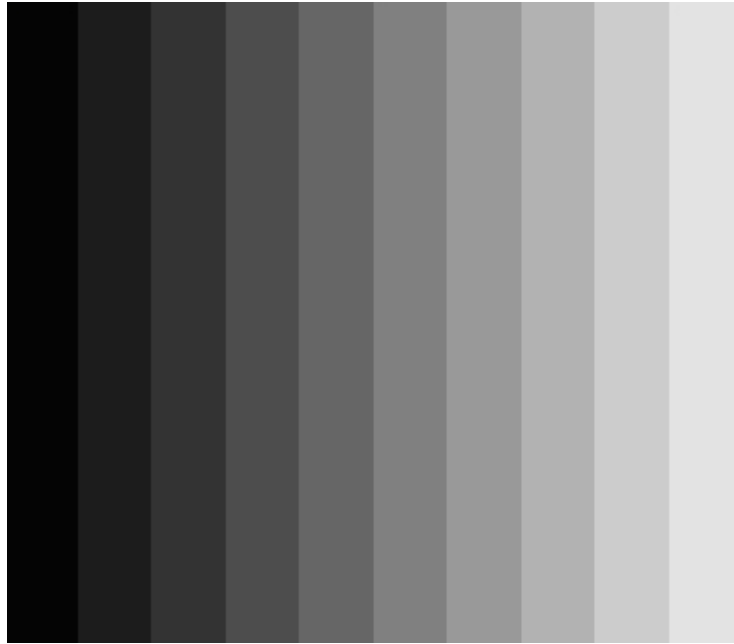
Image and Video processing is used in many fields. Some examples are scientific images such as ones taken on Mars to compress them so that storage and transmission of it is easy for Mars Rovers, movies post-production to manipulate the videos to have certain effects, historical image restoration, identification of similar images and videos, image/video enhancement, patient's organ (such as brain) simulation (such as DBS - Deep Brain Simulation) while carrying out a surgery.

We are mostly familiar with images that we can see with our naked eye but it is a very small part of the spectrum. Images can be captured for the entire electromagnetic spectrum. For e.g., X-Ray imaging used to scan human organs and also inspection of printed circuits, Ultraviolet imaging, light microscopy imaging for both biological and circuit inspection, ultrasound, artificial images created with computers satellite images in different part of spectrums, visible spectrum images. Also, images of different modalities of the same scene can give us wide array of information.

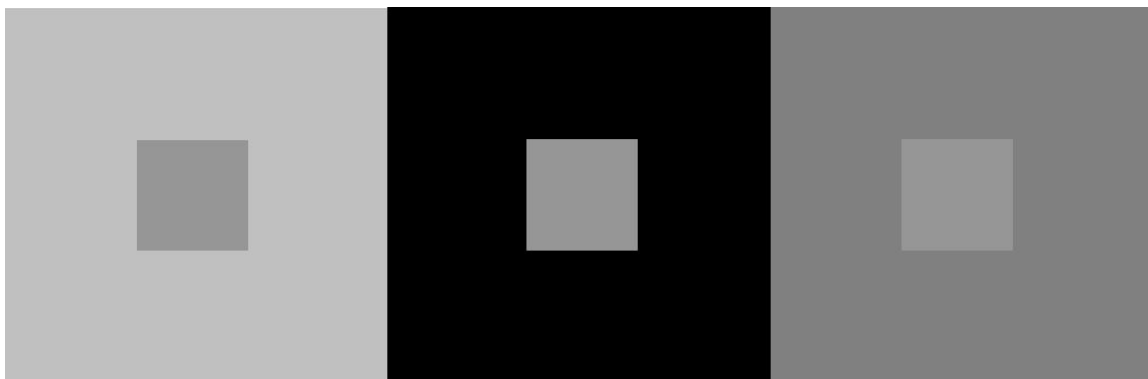
## Human Visual System

Our eyes have sensors in two forms: cones and rods. Cones are very good at seeing the specifics and details of the scene. They are present almost uniformly across the retina. Rods cannot see the specifics but are very good at getting the overall view of the scene. We are able to perceive very large range of intensities but not the extreme ends at a same time. So, we cannot see very dark and very light regions of a room at the same time. When moving from a very dark to very light room (or vice versa), it takes some time for our eyes to adapt to the new lighting environment. In a dark room, we are not able to distinguish between different objects easily. It requires much more amount of change in their intensity to be able to notice the difference. On the other hand, when the background is already very bright, we don't need so much difference between levels to notice a change. This is described a Weber's Law. In other words, we are less sensitive in dark backgrounds than in the light ones.

Mach Band - Mach bands is an optical illusion named after the physicist Ernst Mach. It exaggerates the differences between neighbouring areas of slightly differing shades of gray along the boundaries, thus enhancing edge-detection by the human visual system.



Our perception of a particular color depends upon the background color. If a same grey color (say x) is shown above a black, dark grey and light grey backgrounds, the color x will seem to be lighter on black background, less but still lighter on dark grey background, but darker on the light grey background.



Group 1

Group 2

Group 3

## Image Representation in Computers

The digital cameras we use work in a similar way to our eyes. It has  $A \times B$  number of sensors in a 2D array which can capture the different lights coming towards it. Since, A and B are a not so big and finite number, it is not possible to capture every single ray of light. Values at a certain intervals are only captured and this process is called Sampling. Also, the number of colors that we can represent in our computers is limited so we can't capture all the different levels of intensities of light. So, we convert those not-possible-to-represent values to representable ones (e.g. by rounding) and this process is called Quantization.

Very good quality digital cameras have dense enough sampling and very fine level of quantization so we don't notice that everything is discrete and believe that it is a continuous stream as it is when we see a scene with our eyes. Each of the sensor's captured value is stored in a 2D array and the values are known as pixel (origin - picture element). So, more the number of pixels, the larger will be the picture array and the better the picture will be.

Color images are represented by three colors - Red, Green and Blue (RGB). Camera acquires three images for each of these colors. One thing to note is that most of the cheaper

consumer cameras don't actually capture three images each for R, G and B which we would like to have (and is done by expensive ones). What these cameras do is that they create mosaics interleaving the colors. So, the array would look like -

R	G	R	G	R	G
B	R	B	R	B	R
R	G	R	G	R	G

The values in these arrays are then interpolated with the help of neighbouring values. Usually, a single shade of either of R, G or B is represented with 8 bits. So, a complete RGB color takes up 24 bits or 3 bytes. In a 8 bit color, 0 represents black and 255 represents white.

Since we are now limited to only 256 levels of variations, when an image is too bright we might not be able to represent small variations in brightness which is same as how our eyes work. Our eyes and also the cameras cannot capture most of the information of dark and bright regions at the same time. This is called saturation.

Let us see how Quantization works. Lets say we have an image with 8 bit color values and we want to convert it to an image using only 7 bits color values. We have to divide every value by 2, round it down and then multiply by 2. Lets say the value is  $x$  then,  $\lfloor x/2 \rfloor * 2$  will be how its done. So, every previous color value is now quantized to the intervals of 2. When 8 bit color values are scaled down to 6 then, the same equation can be used except that  $x$  is divided by 4 and not 2. Here are quantization is done at intervals of 4. Same process can be done for different smaller ranges. So, now we know that images are represented as a 2D array of discrete values.

## Simple Image Operations

Every pixel in the image as neighbor pixels. Two types of neighborhoods are considered when doing image processing. One is four neighborhood which includes the pixels in the four directions of a pixel. Another one is eight neighborhood which consist of all the surrounding eight pixels. Following are some simple image operations that can be done-

- One of a very simple and important image operations is addition of two images. For example, when NASA takes an image of a galaxy, it contains a lot of noise. To solve this noise problem, NASA takes hundreds of images without moving the camera and then adds the images and averages them out. This causes the noise to get reduced.
- Another operation is masking/subtracting the images. When images of the same scene are taken under different conditions, then subtraction helps us to see the changes.
- Inversing the color of the image can be used to distinguish the objects in the image which otherwise was not possible in the original image.
- Logical operations like union, intersection, etc.
- Local averaging of the pixels by replacing it with the average of its neighbouring pixels including itself.
- Scaling, Rotation, Translation, Shear, etc.