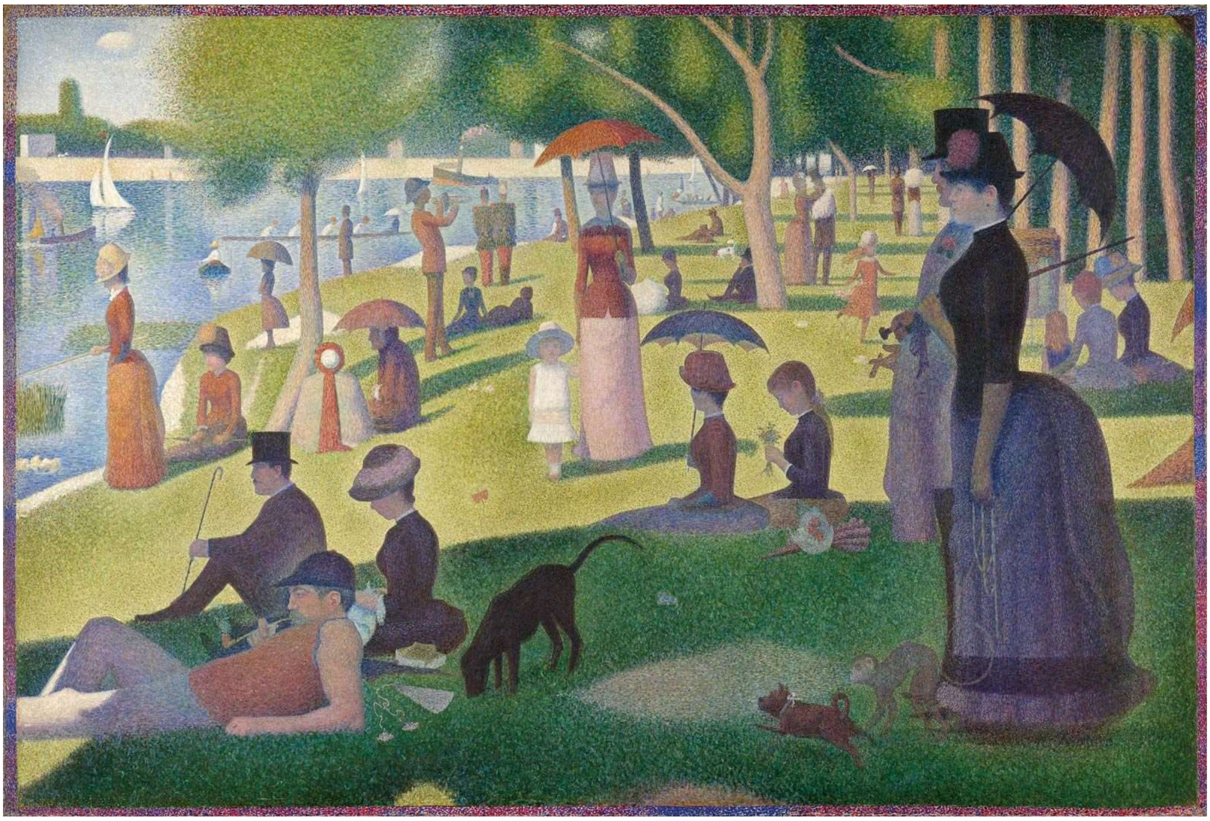
A more real world than the real world

In 1884, when George Surat painted “A Sunday on La Grande Jatte” with dots, he may have never thought that over 100 years later, digital cameras use the same technique to create images. At his time, his work, limited by paint and brush, is quite distant from the reality (maybe that is what he had wanted).

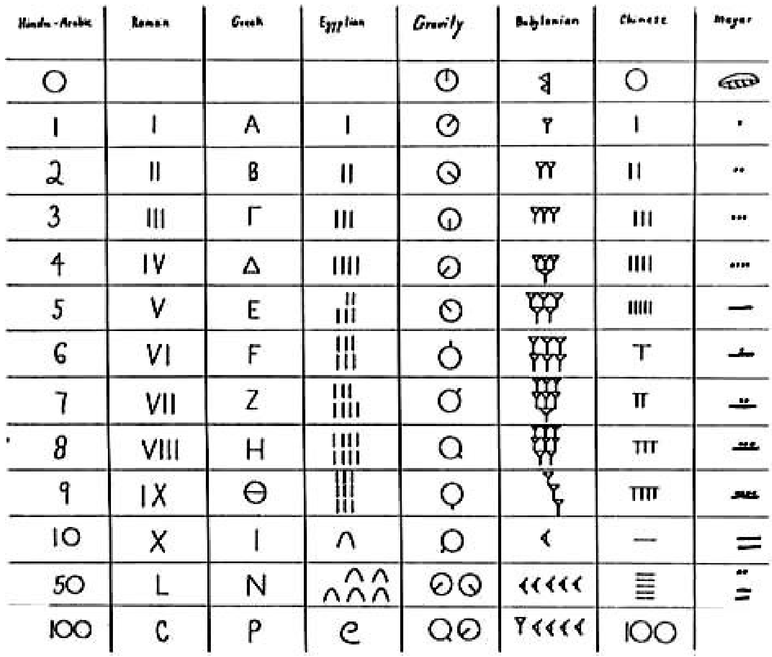


A Sunday on La Grande Jatte by George Surat 1884

Nowadays, digital cameras replicate the idea of Surat but with even more concentrated dots that we now know as pixels. Cameras are so good with the technique that in fact, they are creating a world that is even more real than the reality that human can physically perceive. Let’s take a small peek into this amazing digital world of 0 and 1 where we can break free from the limitation of our bodies and feel a world, arguably more real than our own.

Foundation: a world of 1+1 = 10

Binary can be considered as the start of the digital world just like the numeral system can be considered as the start of mathematics. Number is an abstract idea. To specifically grasp it, a system of symbols is needed, and this is known as the numeral system. Many numeral systems have been created in the past, some of the more famous ones include Roman numeral (a combination between denary and quinary), Egyptian numerals (denary) and Babylon numeral (Combination of denary and sexagesimal).



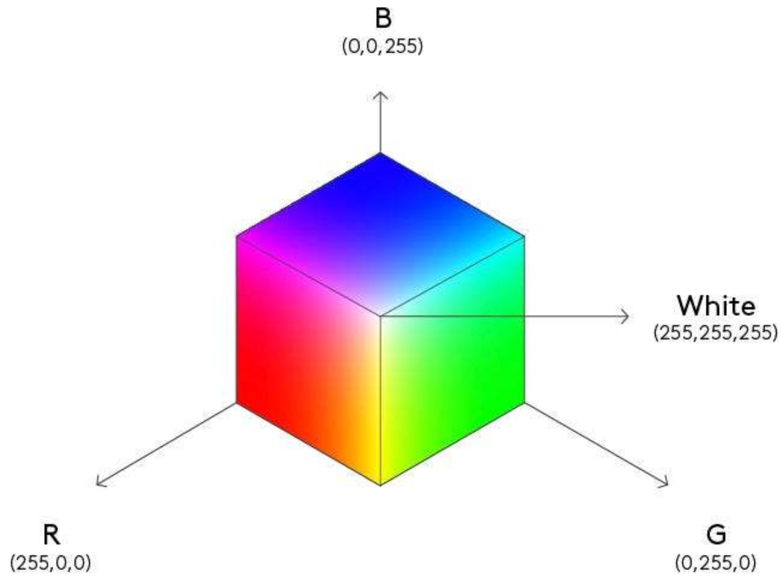
However, all of these are eventually abandoned, and we are left with Denary in Western Arabic numeral because of its simplicity and a nice fit with the ten fingers of the human hand. Yet, we are now seeing the rise of a new numeral system that is quite unusual to human brain at first – Binary.

Why is binary used? Because we can’t find anything with more than two states that are so clearly defined. Maybe we can but it will be much more inaccurate and more prone to ambiguity; so, in order to create a world that is more delicate than the world that we are used to, switches (transistors) with only two states are chosen.

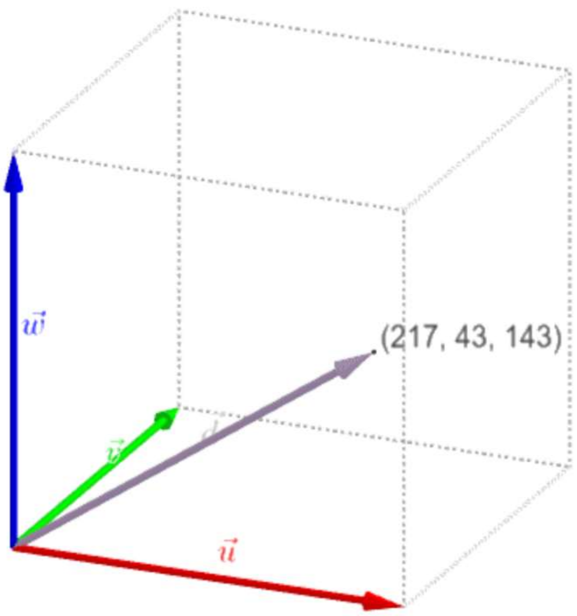
Construction: Colour and Pixel

Now we have the foundation, we can start the construction of the world, but how? Humans perceive the world through their five senses and vision is the most prominent to most of them. So, let’s take vision as an example. We see the world through colours thanks to our eyes, more precisely, thanks to cones which are receptors in the retina in our eyes. There are three types of cones each responsible for a different range of wavelength of light corresponding to the colours red, green, and blue. Hence, our vision is composed of three different colours and any other colours we perceive are a mixture of these three colours, we call them primary colours, just with different shades. You can dot these colours in some blank spaces. If your dots are small and tight enough, human eyes can be cheated to believe that it is a continuous hue. We call these dots pixels. With enough pixels and hue, an image is created.

But how can we model the colour from the real world? Looking at how our eyes work, it is natural to create colour as the mixture of the three primary colours we have in our hands. The first problem in front of us is how can we mix these three lights. This is relatively simple as our vision has its limits. If the source of light is close to each other our eyes mistake it as a single source. The second problem is to mimic shades. This is also simple. As we know that darkness is just the absence of light, we can create shades by just varying the intensity of light from the source. Now we have all we need, and it is time to create a mathematical model. Since we are mixing three colours at different intensities, it is reasonable for us to arrive at the 3D vector model with red, green, and blue each on an axis just like this:



In this vector model, we can arrive at any point on the graph with a position vector (or a set of coordinates) such as:



That's not impressive? Let's put it this way. In theory, this simple model allows us to control the shades of each colour to infinite precision which means that we can create any shades between black and Red (or Green or Blue) leading us to be able to literally create any visible colour! Well, but we live in a practical world with limited resources and infinite is unachievable and true continuity is impossible. In the end, a scale from 0 to 255 is set and a total of 256^3(16581375) colours can be created. That is a lot! Let's put it in comparison with our eyes which can only see around 100 shades for each primary colour and there are only around a million different colours we can clearly perceive which is roughly ten times less than the number of colours in this simple model. There we have it: the RGB colour system.

Optimization: Chroma Subsampling, DCT and Quantization

Our visual world consists of images that are continuously sent to our brains at an average rate of about 60 or 70 images per second. However, as we have said earlier, we have limited resources in our world just like we have limited memories in our brains. We cannot span this information, especially during situations such as uploading and downloading or streaming a video where great information may slow down the process causing a reduction in the frame rates. This is when we notice that the video is not smooth and difference between the real world and the digital one. How can we mitigate this problem?

Since the end of last century, people are working on it. At the moment the most common and, dare I say, one of the most effective methods is compression. Let's take the most common photo compression JPEG (Joint Photographic Expert Group) as an example. For JPG to work, it relies on 2 defects of human.

Chroma Subsampling

The first one is the fact that human can't see colour as well as the grey scale. This is caused by the fact that we have much fewer (cones) photoreceptors in our eyes that are responsible for recognising colour (around 5 million) than rods (around 90 million), which are also photoreceptors responsible for seeing grey scales. But how can we utilise this?

We have to separate the grayscale image (luminance) from the colour image(chrominance). One of the colour systems known as YC_bC_r luckily does that for us. It separates images also into 3 different channels, but unlike RBG, one of the channels is a Luminance channel(Y) instead of 3 colour channels. This conversion can be done using a matrix multiplication:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.256 & 0.504 & 0.098 \\ -0.148 & -0.292 & 0.441 \\ 0.441 & -0.369 & -0.071 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

Or more simply:

$$Y' = 16 + \frac{65.738 \cdot R'_D}{256} + \frac{129.057 \cdot G'_D}{256} + \frac{25.064 \cdot B'_D}{256}$$

$$C_B = 128 - \frac{37.945 \cdot R'_D}{256} - \frac{74.494 \cdot G'_D}{256} + \frac{112.439 \cdot B'_D}{256}$$

$$C_R = 128 + \frac{112.439 \cdot R'_D}{256} - \frac{94.154 \cdot G'_D}{256} - \frac{18.285 \cdot B'_D}{256}$$

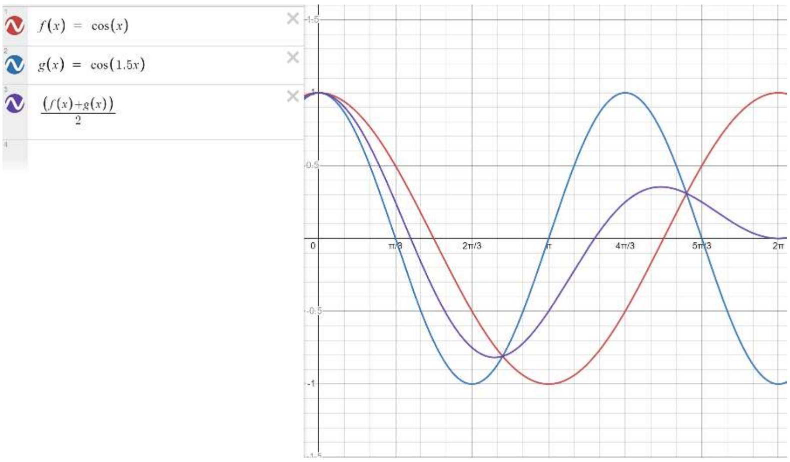
So now we have excluded the luminance channel, we can remove some of the information on the colour by using Chroma subsampling. Normally this is done by separating the image into a lot of 2x2 or 1x2 blocks. For each block, only the front or the average value of C_b and C_r is kept. We are halving/quartering the colour information; surely this will be very noticeable right? Here is an example: (subsampled image on the left)

Not a lot of difference, right? Unless you can zoom in to see individual pixels, these changes are pretty much unnoticeable (Subsampled image on the left). However, they reduce the information for colour by a huge amount.

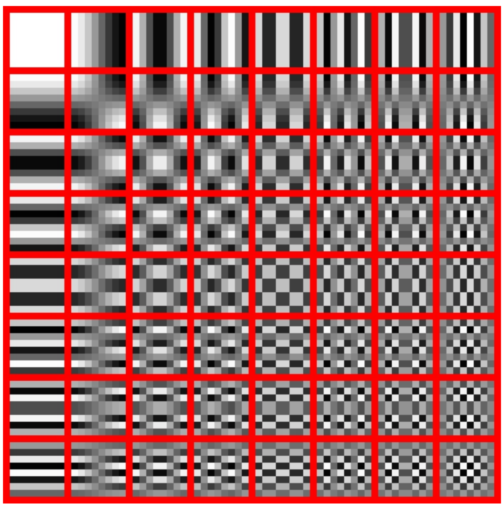
Enlarged (from the yellow box above) side by side comparison

DCT and Quantization

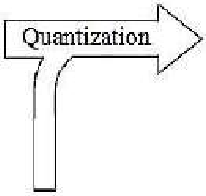
Now we move on to the second defect of human. Our brain is actually not very sensitive to high frequency of shifts in colour or luminance. In general, this means we can get rid of the high frequency shift unnoticeably. This is done by using DCT (Discrete cosine transform) and quantization. In DCT we separate a channel of a picture into thousands of small 8 pixels by 8 pixels block. Here we can replicate these 64 pixels with the weighted sum of cosine waves. What does that mean? As we know, cosine function has the maximum of 1 and minimum of 0. This can be used to mimic the maximum intensity of the channel and the minimum intensity of the channel. If we change the coefficient of the x in the cosine wave, we can get another cosine wave but with different periods.



It was recognised that any wave be reformed by a combination of many other waves. So, if we have summed enough cosine waves, we can obtain the value of a row of pixels. And if we repeat this for the column, we can mimic the whole 8x8 block just with cosine waves. In the actual compression, a table of cosine waves is used. Each red box represents a 2D cosine wave in discrete form.



It is checked how much of the 8x8 block is composed of each of the cosine waves. The “how much” is represented by a coefficient. This coefficient is then divided by a table of values known as the quantization table (a matrix) before being rounded. Remember that we said earlier that human are not very sensitive to high frequency shifts in colour? We want to get rid of all the high frequency patterns at the right bottom corner of the Cosine table as their frequency is so high that it is highly unlikely to be recognised. We can see that a high value is designated in the right bottom corner of our quantization table. This means they will be zeroed when we round.



1243	18	5	2	-3	-1	-6	2
4	1	1	0	3	2	1	2
-19	-5	2	2	-1	2	4	-1
-7	1	-1	0	3	2	-2	2
-6	1	-3	-1	1	1	4	1
14	-1	5	0	-2	1	4	1
24	3	3	0	1	2	12	2
-15	0	-4	-2	0	-1	-6	0

DCT Coefficient

78	2	1	0	0	0	0	0
0	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Output

16	11	10	16	124	140	151	161
12	12	14	19	126	158	160	155
14	13	16	24	140	157	169	156
14	17	22	29	151	187	180	162
18	22	37	56	168	109	103	177
24	35	55	64	181	104	113	192
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	199

Quantization table

The data will be stored as the values in the last table also known as quantization coefficients. Since there are a huge number of zeros due to quantization, the data can be very efficiently stored. When the image is revisited, the image can be built quickly with the reverse process – using the quantization table to regain the DCT coefficient and rebuilding the image using the cosine wave functions. This process not only greatly reduces the image size, but also preserves the resolution while making the changes. Don’t you believe me? Look:

Not a lot of difference is seen, right? But the left image is compressed and more than ten times smaller than the right image. This right one is a raw image I have taken of a gibbous moon that takes me ages even to open.

Conclusion

With these wonderful powers, pictures are able to be presented with astonishing fidelity that is even beyond the resolution of our eyes. Now we come back to our initial thoughts. This digital world has already and literally exceeded our comprehension. This is just a small peek into the visual part of this growing world. There is much more to this world than just vision. We, in our short lives, will not have enough time to do all we want and travel to wherever we want. But the digital world gives us a chance to experience all maybe even in a higher level of detail than those who have physically experienced it in real life. Indeed, this is a more real world than the world we live in.