



Resolution and compression – myths and reality

Information volume of image in relation to its digital saving. Changes of information volume depending on chosen resolution, compression and calibration.

Stanislav Psohlavec, AiP, 2001-24-03

Content :

Influence of [resolution](#) theoretically and [in reality, an example with a microfilm, conclusion](#).
Effect of [compression](#), connection of [resolution and compression](#), [recapitulation](#).

Each objective to provide digital image starts with the decision about the quality of the digital image. This important decision is surprisingly very often based on subjective opinion influenced by unverified third party information or widely accepted rumors.

It is this decision, about how the data should be created and stored, which is above all connected to the cost of the projects and the satisfaction of the users.

Intentionally we will not consider how the particular format affect the stored data. We are assuming that the results of chosen digitizing process satisfy the users needs. We will focus on the effects which the chosen resolution and compression has on data and on the volume of the information the data contain. There are certain statements I consider being truthful:

Satisfaction of the user comes from the available information volume of digital image.

Basic technical features such as resolution and compression determine the size of the data of the digital image.

The size of the image can be greater than the size corresponding to the information volume.

In this discourse I want to show that there are certain impacts to be known when determining effective use of resolution and compression.

Information volume

Practical user is interested in available information and if he gets it, he is satisfied. He does not care about the original size of the data, as long as he has received all the information he needs. Is it possible to write half mesh between the size of the data and the amount of the information that it contains? How to evaluate amount and quality of the recorded information? Line frequency tests are used for photographic film to determinate the limits in recording of the details. Line frequency (number of lines per millimeter) which a film material is certainly capable of recording together with the size of the film frame sets the limits of recording details for photographic film material. The same principles work in recording of electrical signals representing sound. This problem is very well theoretically resolved in the field of signal analyses (Furier transformation, spectral characteristics of signal). The results of these approaches have found their practical use firstly in setting of the technical norms for digital recording of sound, and afterward in the whole development of compression for saving the sound in digital form. Nobody doubts practicability of the compression in this case.

Similar trends in development have also reached the field of image processing.

We can find image data of sophisticated compression and optimized resolution when watching television or using Internet. In TV for example, color information is restrained and the viewer's perception is based more on information in the area of brightness. In recent technologies of processing and saving the images the same principles as the ones already used

for audio and video technologies are applied. The key principle is keeping the kind of the signals carrying the **useful information** and intentional suppression of the information which is not perceivable. In spite of these facts, when digitizing images formal rules are often used when applying resolution and compression. In this case I mean formally set rules for resolution and still living nightmare of a loss compression.

Let's look at this problem from a different perspective. Do the basic technical parameters such as resolution and compression define the volume of recorded information?

I think that we are reaching terminological problem here. As a matter of course, the file containing an information about digitized homogeneous white area carries data about its each point. From this point of the view the answer to above mention question is yes. But in this case I mean the usable information, for example the number and quality of details, which are available. Even though this is very subjective approach, I believe that this definition will be sufficient for further explanation. Looking at the image capturing white area from this point of view, we aren't even able to see if the image is focused. In this case usable information volume is smaller than what corresponds to the size of the image.

What defines the information volume when digitizing?

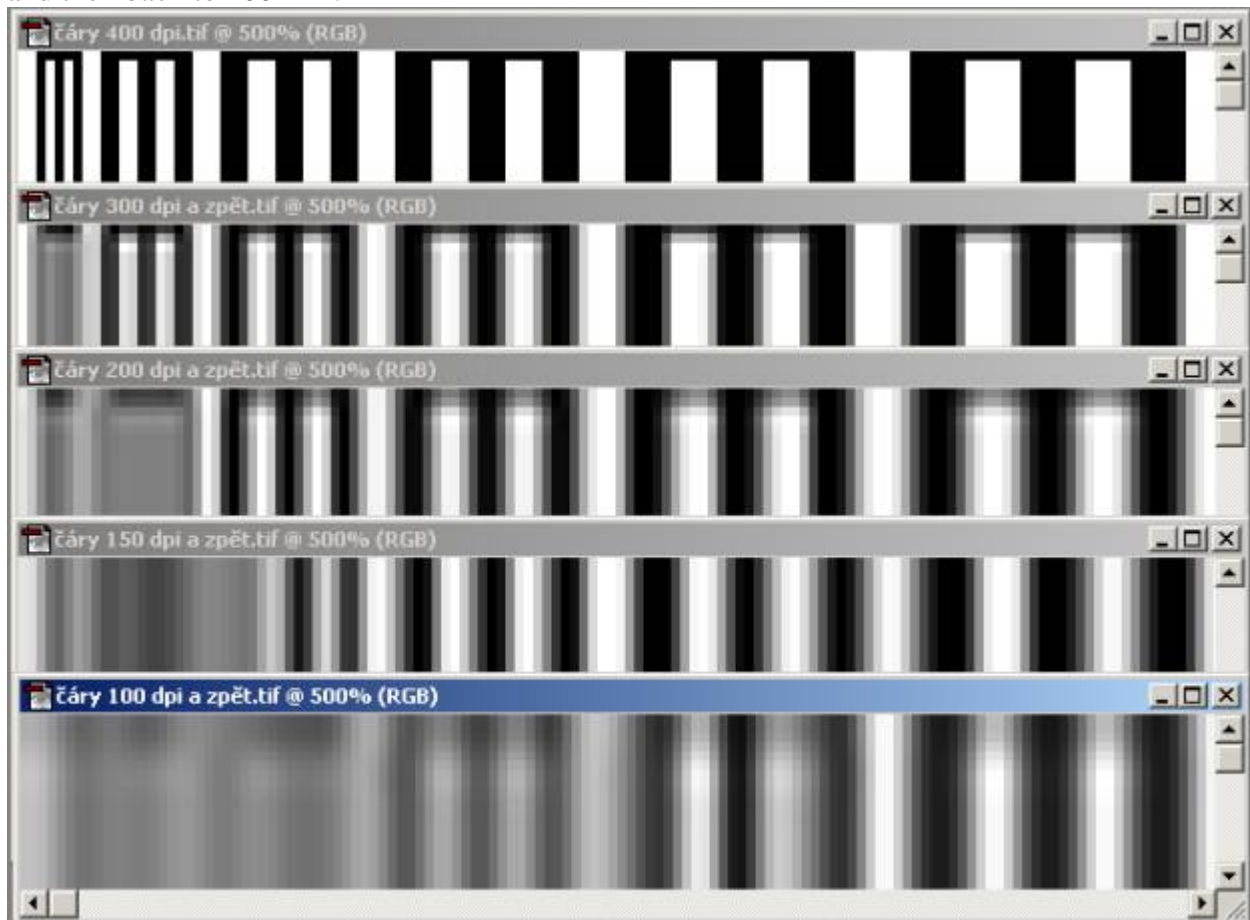
Nature of the digitized object itself – existence of available details

Technical features of the digitizing device

Saving method

Effect of resolution on quality of an ideal image

Firstly let me show you one example with specially prepared samples. The first image is an original in resolution 400 DPI, the other images are its conversions to 300,200,150,100 DPI and then back to 400 DPI.



Note: All of the images presented in this document are just a demonstration. If you want to check the presented facts on your own, repeat the mentioned steps on your image in your image editor.

In the first image, the resolution is fully used. In the other images the information volume was lowered, but the size of the image stayed the same. You can see that each lowering of information volume leads to an evident loss of information – unfocused image. In the same time this action does not necessary lead to change in data size and resolution. This example can look quite trivial. Try to get back to it after reading the whole article.

Effect of resolution on quality of real image

The following images are details of scanned business card in 200, 400 and 800 DPI.

A: 200DPI

B: 400DPI

C: 800 DPI



I think that there is no discussion about the evident difference in information volume. The differences in picture quality is clearly shown by the pixels. The fact that picture C is 4 times larger than picture B is intelligible.

Lets look at these three images.

These images are all in 800 DPI, the file size of all the images is therefore the same.

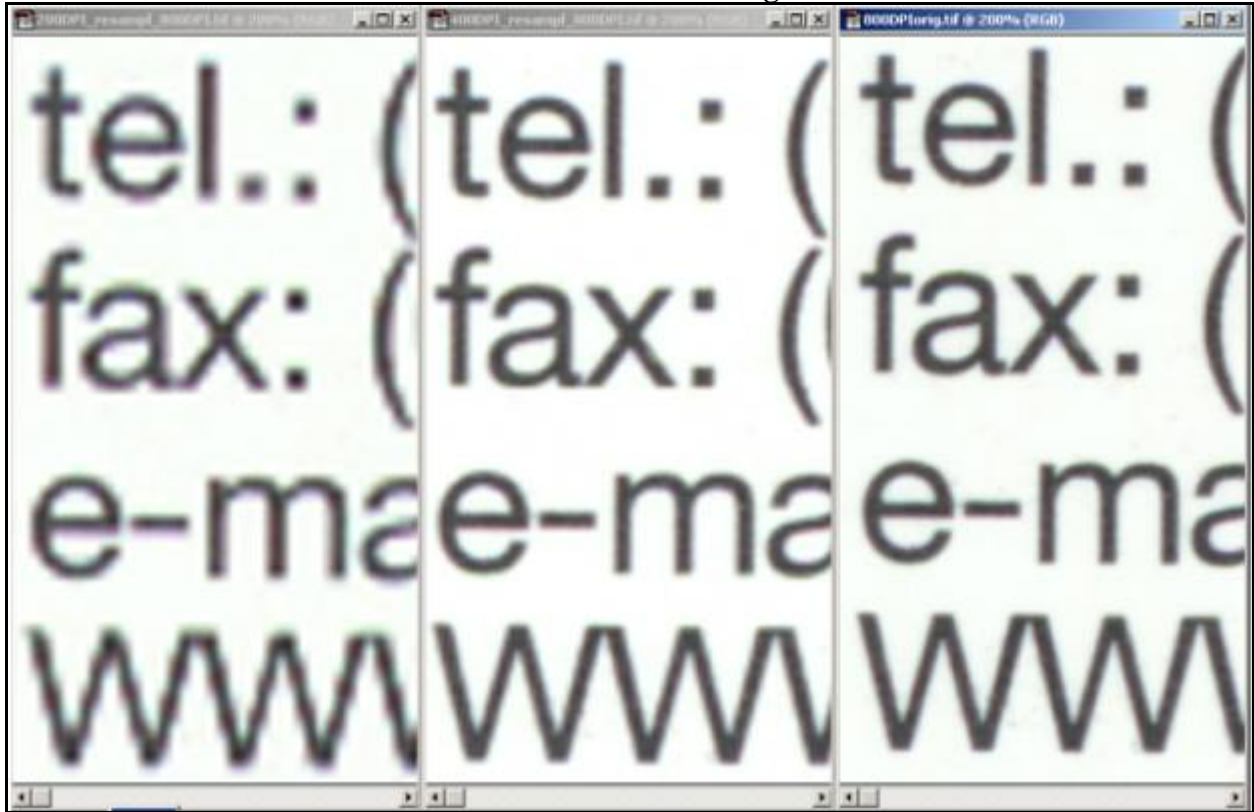
It is the **same three images like in the previous example**.

The images **A** and **B** were only resampled to the same resolution like image C.

A: bicubic 800 DPI

B: bicubic 800 DPI

original: 800DPI



The mathematical operation itself does not create another new information. It only subjectively looks like the images have become better. The first image looks worse than the second but the second is really quite same as the third.

This experiment brings very important conclusion:

Information in image B scanned at 400 DPI is quite same like information in image C at 800 DPI.

In information volume of images B and C is no difference, even though their size is 1:4 !

The difference between images A and B shows that the information volume of B is larger then in image A, but why? Used scanner produces images in 350 DPI based on its built in hardware capacity. The image is immediately recounted to requested resolution (in this particular scanner up to 9600 DPI). No matter what will be the final resolution, the information volume will remain the same as it was in the original data – maximum 350DPI. The difference will be only in subjective perception of the picture and also in the size of the data file. Resample can be used for preprint preparation, but should not be used when creating the original digital data.

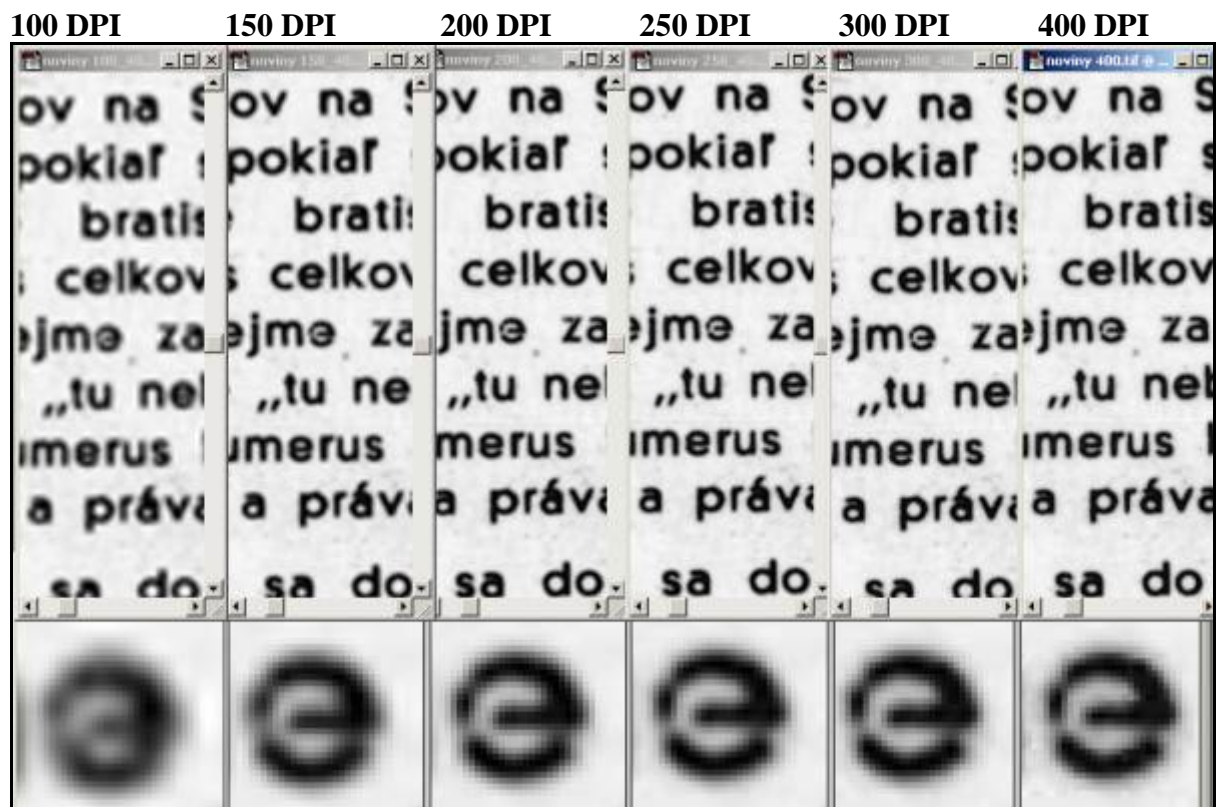
It is important to realize that for perfect holding of the results of digitization it is necessary to save only the original data. If needed in future, it will be always possible to resample image to get higher resolution. Many software applications enable this conversion within the basic functions (MS-WORD for example when pasting image into a text). Therefore it is not necessary to save, take care of and transfer unreasonably large data. It is always effective to handle only the necessary original information and then convert it for the purpose of the best presentation.

We can make one more experiment with these images. Lets take original image C (800 DPI) again. If we resample this image to 400 and to 200DPI and then back to original 800 DPI (similar experiment like the one with the ideal lines), we see what resolution has brought loss of information. If we want to keep all the available information volume, saving this image in 400 DPI is sufficient. Attention, each resample brings some losses, so the best option in this case would be to keep the image in original scanned form (350 DPI). But this does not have to be final conclusion, as you may see from the following example.

Microfilm example

Information volume available on microfilm is influenced by the film characteristics. For digitization of microfilm, the information volume responds to the characteristic of the film material as well as the characteristics of the scanner. Scanner SunRise, for example, can fully use the resolution possibilities of its CD component. When using this device for microfilm digitization, information volume is limited only by the microfilm characteristics. In this case it is totally opposite situation than it was in the first example, when the quality of the original (its information volume) was much higher than the capacity of the scanning device.

The original scanned image in 400 DPI is on the very right side. This image has been resampled to 100,150,200,250,300 and then back to 400 DPI. If there was information in the image that needed higher resolution, this information has been lost when returning the image to 400DPI.



This series of the images shows that the perceptible degradation of the information volume starts when using resolution under 200DPI, when the image is 4 times smaller than the original. It means that for the microfilm definitely more than 150 DPI is needed. The resolution of 200DPI would be quite sufficient. Therefore we can say that using 250 – 300 DPI would be safe to catch the whole available information volume. It would be needless to use 400 DPI or even more.

The conclusion in the question of the resolution

Information volume of original object and the characteristics of digitization device determine the pragmatic resolution, which should be used to save the digital information. Use of lower resolution means loss of information, use of higher resolution leads to unreasonable data size growth, because the information volume stays the same. Specification of resolution for an image does not inform about real information volume of this image, real information volume is always lower.

Practical experience

To find out if the image has the information volume according to the declared resolution, edges of the image can be examined. On a well digitized image containing straight borders of black and white areas (for example text on a business card) nearby pixels must be only black-gray- white.

From my experience same results can be achieved, when printing-out images of 300DPI and 800DPI according to above mentioned criteria on a good laser printer.

Loss compression

I think that it is the word loss which brings the unpopularity of loss compression among the users of digital images. Let's compare the effect of compression (in ratio 1:3 and 1:10) with sharpening of the image.

Saved in JPG format, Photoshop 5.0

the highest quality (10), compression 3x

medium quality (5), compression 11x

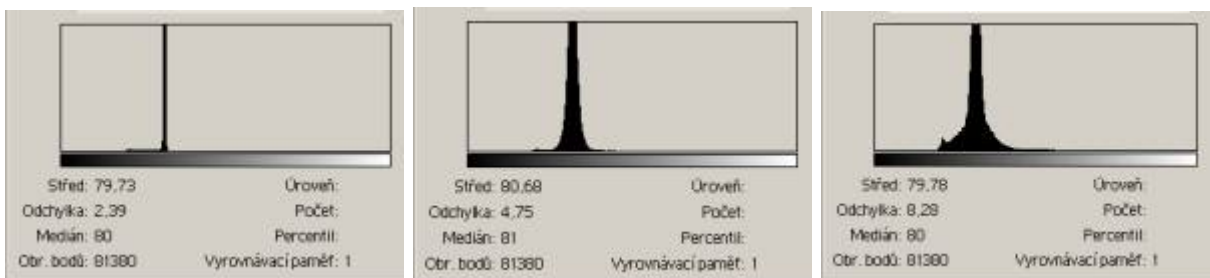
sharpening



The following images have been created by subtracting of the compressed image from the original image. Therefore they show the deformation of the original. The images were edited for better visualization using brightness (+100) contrast (+40). Compressed image identical with the original would lead to plain gray picture.



Corresponding histograms show the change of original caused by compression or sharpening.



We can see that the sharpening has brought greatest changes. Even though the effect of sharpening is higher than reasonable use of compression, sometimes the sharpening functions are implemented in the process of capturing of the data!

Saving the image in JPG format in the highest quality has brought the diversions only around +/- 2 bits (for test with plain gray calibration card), these diversions are mostly on the edges.

When scanning real parchment or paper the diversions in homogenous area range around +/- 5 bits. These effects are therefore concealed by the pattern of the scanned material.

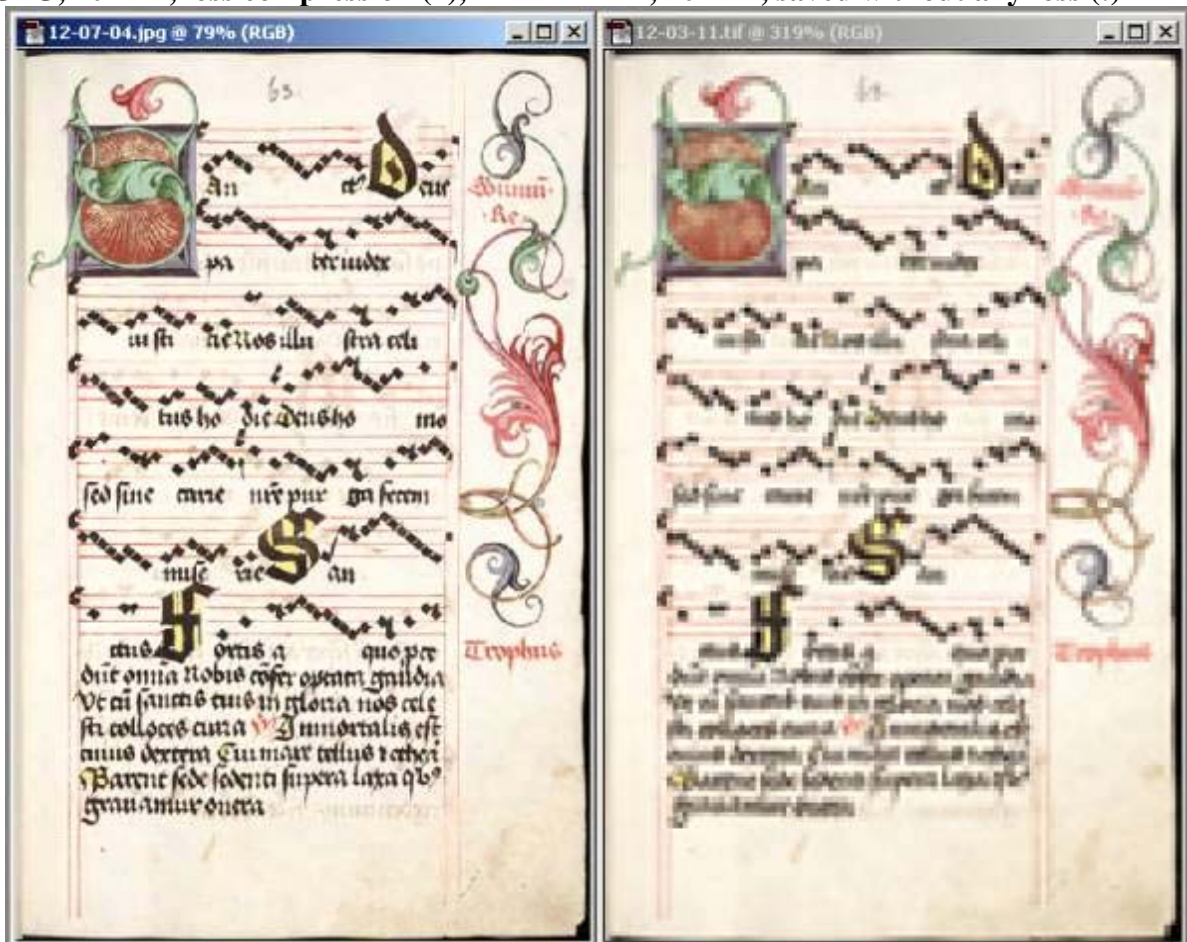
When saving the images using higher compression, experienced eye can spot the typical effects, but those don't lead to the loss of information. The edges and the lines in the picture carry the main information. Only the high compression really brings the loss of information by degradation of the edges and sharpness.

Relation of compression and resolution

Using of high compression can be combined with higher resolution. The result is getting higher information volume, while preserving the same size of the data. The following example shows two images of the same size 65 kB. Let's see which image brings more information.

JPG, 40 DPI, loss compression (2),

TIF, 10 DPI, saved without any loss (!)



Calibration

Calibration requirements still complicate the question of what is the right digital image to be archived. Generally it is possible to say, that the difference between the calibrated image and the one without the calibration is usually greater than the shown ones when using the loss compression. Unfortunately among the digitizing studios providing the images with the ICC profiles is still rather exceptional.

Recapitulation

Using of loss compression when communicating on Internet is a matter of course. New technologies for compression are being developed. These technologies are more effective in

using of the memory space only for the kind of the information which is of real use to the viewer. The digitization is always at the start, in this point it is therefore necessary to consider not only the needs of future user in the terms of information volume, but also memory size of created data in relation to its effective saving.

Reedited by Stanislav Psohlavec 2003-08-05