







DETERMINATION OF OPTIMAL FORMATS FOR DIGITAL IMAGE COMPRESSION

DETERMINACIÓN DE LOS FORMATOS ÓPTIMOS PARA LA COMPRESIÓN DE IMÁGENES DIGITALES

Abimael Adam Francisco Paredes^{1,*} , Heidy Velsy Rivera Vidal de Sánchez¹ ,
 Inés Eusebia Jesús Tolentino¹ , Jimmy Grover Flores Vidal¹ 

Received: 06-03-2024, Received after review: 21-06-2024, Accepted: 16-09-2024, Published: 01-01-2025

Abstract

The objective was to determine the influence of different image formats and tools used for compression on the final size of the images, to know which are the optimal formats for compression. The sample was made up of five digital image files with BMP extension, taken in different scenarios and at different times at the researcher's discretion. The technique used was the analysis of digital image files and as an instrument a double input matrix, where the conversions of BMP files to six different extensions of image files were registered, with four different tools for manipulation of image files. The experimental design was factorial, where the two factors were the image compression formats and tools and the dependent variable the final image file size. Factorial ANOVA statistical analysis was applied with $\alpha = 0.05$. It was obtained that the format of smaller size was the JPG when using as tool the *Illustrator* and the one of greater size the one of greater extension the PSD also obtained with the *Illustrator*. The statistical analysis showed that the format factor significantly influences the final size of the images ($p < 0.05$) and the tool factor does not show significant influence on the size of the images ($p > 0.05$), nor is the interaction between the factors significant. It is concluded that regardless of the tool used, it is the image format that influences the final size.

Keywords: image quality, compression techniques, pixels, image handlers, size reduction

Resumen

El objetivo de este trabajo fue el determinar la influencia de diferentes formatos de imagen y herramientas que se utilizan para la compresión en el tamaño final de las mismas, para conocer cuáles son los formatos óptimos para la compresión. La muestra estuvo conformada por cinco archivos de imágenes digitales con extensión .bmp, tomadas en diferentes escenarios y horas a criterio del investigador. La técnica empleada fue el análisis de archivos de imágenes digitales y como instrumento una matriz de doble entrada, donde se registraron las conversiones de los archivos .bmp a seis diferentes extensiones de archivos de imágenes, con cuatro diferentes herramientas de manipulación de archivos de imágenes. El diseño experimental fue factorial, donde los dos factores fueron los formatos y las herramientas de compresión de imágenes y la variable dependiente, el tamaño final del archivo de imagen. Se aplicó análisis estadístico ANOVA factorial con $\alpha = 0,05$. Se obtuvo que el formato de menor tamaño fue el .jpg al utilizar como herramienta el *Illustrator* y el de mayor tamaño el .psd, también obtenido con el *Illustrator*. El análisis estadístico mostró que el factor formato influye de forma significativa en el tamaño final de las imágenes ($p < 0,05$) y el factor herramienta no muestra influencia significativa en el tamaño de las imágenes ($p > 0,05$), como tampoco es significativa la interacción entre los factores. Se concluye que independientemente de la herramienta que se utilice, es el formato de la imagen lo que influye en el tamaño final.

Palabras clave: calidad de imagen, técnicas de compresión, píxeles, manejadores de imágenes, reducción de tamaño

^{1,*}Universidad Nacional Hermilio Valdizán, Pillco Marca, Perú.
 Corresponding author ✉: adamfp28@hotmail.com.

Suggested citation: Paredes, A.; Rivera Vidal de Sánchez, H. ; Tolentino, I. and Flores Vidal, J. "Determination of optimal formats for digital image compression," *Ingenius, Revista de Ciencia y Tecnología*, N.º 33, pp. 9-14, 2025, DOI: <https://doi.org/10.17163/ings.n33.2025.01>.

1. Introduction

Image compression encompasses a set of techniques applied to digital images that enable efficient storage or transmission [1]. These techniques have been developed to address the significant file sizes that image files can occupy, which often limits their exchange via email and other electronic platforms. Compression methods rely on mathematical algorithms that reduce file size, thereby minimizing resource consumption and transfer time [2].

All image compression algorithms aim to achieve a smaller compressed image size (high compression factor) while maintaining a high-quality reconstructed image (high-quality compression). The efficiency of these algorithms can be evaluated based on the specific application using various criteria [3]. The most important criterion is the compression factor, which compares the image size before and after compression. Therefore, a higher compression factor indicates a more effective compression algorithm [4].

The most common types of compression are lossless and lossy compression. In lossy compression, some image information is discarded during the compression process. Some algorithms combine both techniques to achieve compression [5]. The effectiveness of compression also depends on the type of image. For example, bitmap images are composed of a grid of cells or pixels, each with a specific size, and lose resolution when resized [2]. In contrast, vector images are constructed from mathematically defined objects, such as points and lines, which are controlled through Bézier curves. This structure provides greater flexibility, as vectors can be scaled without losing resolution [6].

Regarding compression, both with and without loss, Ruiz, Yarasca and Ruiz [7] explain that lossless compression employs complex mathematical algorithms to condense code chains while preserving all the information in the image. This ensures that the image can be fully regenerated, without any loss, during decompression, although it requires specific encoding and decoding times. Formats such as Portable Network Graphics (PNG) utilize this type of compression. In contrast, Rojatkari et al. [8] describe lossy compression as a technique in which certain image information, typically deemed minimally perceptible, is discarded during compression, resulting in a loss of some original file data. This method is commonly used in image formats such as Joint Photographic Experts Group (JPEG).

Among the most common types of lossless image compression is Run Length Encoding (RLE), which, as noted by Hardi et al. [9], is one of the simplest compression schemes. It works by replacing sequences of identical bits with a code. The method scans the image to identify pixels of the same color, and when the image is saved, only the color value and the position of the

color pixels are recorded. This technique is particularly effective for images with large areas of uniform color, as it compresses the image without losing quality. The Lempel-Ziv-Welch (LZW) method is similar to RLE but supports a wider range of formats, including TIFF, PDF, and GIF [2]. Like RLE, it is highly effective for images with large areas of uniform color and simple designs, but its efficiency diminishes when compressing images with a broad range of photographic-like colors. Huffman coding assigns shorter bit codes to frequently occurring data and longer codes to less frequent data, making it widely used due to its simplicity and high speed [1]. Arithmetic coding, conversely, represents sequences of symbols in binary form by using intervals of real values between zero and one [10].

The most common lossless compression models include Transform Coding, which uses a discrete Fourier transform to represent the image through transform coefficients. A quantization process is then applied, where coefficients with small, insignificant values are eliminated, resulting in some loss of information without causing noticeable image distortion [11]. Vector quantization involves selecting a representative set of pixels from the original image and discarding the non-representative ones. This is achieved by constructing dynamic tables or through clustering for vector classification [1]. Fractal Compression treats images as fractal objects, meaning they are composed of a repeating fragmented structure. Then, some functions are created to generate transformations that divide the original image into smaller, self-similar parts. The iterative application of these transformations produces an image that closely resembles the original but is smaller in size, as some information is lost during the division process [10].

The compression of digital images has been studied from various perspectives, ranging from research on the algorithms used for compression [12,13] [3,4] to specific applications in fields such as forestry sciences [14], forensic sciences [15,16], medical sciences [17–19], and other disciplines. Most research focuses on algorithms for analysing digital image files, with little reference to comparisons of tools and formats for selecting optimal options. This study aims to address this gap by examining whether image formats, tools, and their interaction influence the size of image files. The primary objective of the research is to determine whether the interaction between formats and tools significantly affects the size of digital image files.

2. Materials and methods

2.1. Methodology

The research followed a quantitative methodological approach, utilizing an experimental design at an explanatory level. The population consisted of photo-

graphic image files in BMP format, and the sample included five digital image files in BMP format, captured in various settings and at different times at the researcher's discretion. The technique employed was the analysis of digital image files, and the instrument used was a double-entry matrix, where the conversions of BMP files to six different file extensions were recorded using four different tools (software) for digital image manipulation.

The statistical design employed was a factorial design, with two factors: image compression formats and tools, and the dependent variable being the final image file size. The experimental design included six levels for the first factor (Format) and four levels for the second factor (Tool), resulting in a total of 24 treatments applied to a sample of five images, yielding 120 measurements of the dependent variable. The formats used were JPG, PNG, PSD, PDF, TIFF, and TGA, while the image editing tools were CorelDraw, Photoshop, Illustrator, and Gimp. The images selected for the sample were coded as presented in Table 1.

Table 1. Sample description

N°	Name	Extension
1	Archive f01_bmp.bmp	11796536 bytes
2	Archive f02_bmp.bmp	47775800 bytes
3	Archive f03_bmp.bmp	47775800 bytes
4	Archive f04_bmp.bmp	53747768 bytes
5	Archive f05_bmp.bmp	59722040 bytes

The statistical analysis was conducted using factorial ANOVA to measure both the individual effects of each factor and the effect of their interaction on the dependent variable, with a 95% confidence level. The statistical software used for the analysis was SPSS 25.

3. Results and Discussion

3.1. Analysis and Results

Of the 120 data points recorded, it was observed that when using the CorelDraw tool, the PDF format produced the smallest files [20], while the TGA format resulted in the largest. With the Photoshop tool, the smallest file size was achieved using the JPG format, while the largest was obtained with TIFF. Similarly, when using Illustrator, the JPG format yielded the smallest file size, whereas the PSD format generated the largest files. Lastly, when using Gimp, the smallest file size was also achieved with the JPG format, while the PSD format resulted in the largest files. As noted, JPG consistently produced the smallest file sizes. Although JPG compression is lossy, the algorithm compensates by softening edges and areas with similar colors, making the loss of information imperceptible to the naked eye. This allows for a high degree

of compression, with image quality degradation only noticeable under significant zoom [2]. In this regard, Tan [19] suggests that the choice of format should primarily be based on the content of the image. Photographic images, or those with soft tones and few sharp edges, are generally best compressed using a lossy format such as JPEG.

The analysis also revealed that, across the four tools used for converting BMP files, the conversion to JPG resulted in the lowest average file size, at 210,152,480 bytes. In contrast, the conversion to PSD produced the highest average file size, at 7,042,890,180 bytes. The smallest file was a JPG generated with the Illustrator tool, with a size of 50,474,220 bytes, while the largest file was a PSD, also generated with Illustrator, with a size of 10,044,601,140 bytes. As noted on the Adobe Photoshop portal, PSD is the default format for Photoshop and is compatible with other tools, such as Illustrator. PSD files can reach a maximum size of 2 GB. The fact that the average PSD file size in this study exceeded that limit suggests that some tools may not be fully optimized for converting BMP images to PSD format. Illustrator, which produced PSD files of approximately 1 GB, proved to be the most effective for this conversion. Additionally, Parmar and Pancholi [21] mention that the JPG format is widely used in photography due to its ability to handle millions of colors while maintaining good quality, even with lossy compression.

A factorial ANOVA test was performed to assess the influence of the factors (Format and Tool) on the dependent variable (Image size), with the results presented in Table 2.

Table 2. Results of the applied factorial ANOVA

Origen	Sum of squares	gl	Mean square	P
Model	8.52022E+20	23	3.70444E+19	0
Intersection	1.00536E+21	1	1.00536E+21	0
A. Formats	5.63869E+20	5	1.12774E+20	0
B. tools	5.86365E+19	3	1.95455E+19	0.09
A*B	2.29516E+20	15	1.53011E+19	0.053
Error	8.4412E+20	96	8.79292E+18	
Total	2.7015E+21	120		
Total	1.69614E+21	119		

In Table 2, the p-value is of primary importance, as it indicates the influence of the factors on the dependent variable. For the Format factor, $p < 0.05$ suggests a statistically significant influence on the size of the final converted file, which is expected given the variation in file sizes across different formats. In contrast, the Tool factor shows no significant influence, as $p > 0.05$, indicating that the final image size is not dependent on the tool used for conversion. Additionally, the interaction between the factors does not exhibit any significant influence on the final image size. Therefore,

it can be concluded that the format is the primary factor affecting the final size of BMP images, regardless of the tool used for conversion.

The differences between format types are expected and consistent with the fact that each format employs a distinct algorithm for compression, which directly influences the final file size [1]. This observation is also supported by Salomón [22], who highlights that each format uses a different compression methodology, and, therefore, conversion and compression tools, when operating based on these methodologies, do not show significant differences between them. The results indicate that compressing an image to a particular format can be achieved using any tool, as the resulting file size will not be statistically different. This is further supported by the non-significance of the interaction between the factors. Similarly, AbuBaker, Eshtay, and AkhoZahia [12] reported differences in the size and quality of digital mammogram images depending on the compression methods used, a trend also evident in the differences between the various output formats, each utilizing distinct methods. Likewise, Wahba and Maghari [23] demonstrated that the compression techniques unique to each format are key determinants of the file size or extension of the compressed image.

When differences between formats were observed, JPG consistently resulted in the smallest file sizes while maintaining acceptable quality, comparable to other formats. This finding aligns with Dhawan's [24] research, which compared the compression of various image formats based on the different algorithms used. The smallest JPG file sizes were obtained using the Illustrator tool, suggesting that, although no statistical difference was found between the tools, Illustrator may be preferable for achieving smaller file sizes. This is further supported by Sakshica and Gupta [25,26], who emphasize that Illustrator is particularly effective for compressing vector images [27–30].

4. Conclusions

The results of this study indicate that the final file size of compressed images is determined primarily by the format chosen for compression rather than the tool employed. The smallest file size was consistently achieved with the JPG format, particularly when using the Illustrator tool, which is notably effective for compressing vector images.

However, while JPG yielded the smallest file size, it employs a lossy compression method, which results in the loss of some image pixels, potentially affecting resolution upon decompression. Therefore, further research and experimentation with alternative tools are recommended to more effectively determine the optimal image format for compression, ensuring a balance between minimal file size and the preservation of

image quality aligned with the intended use.

References

- [1] N. La Serna, L. Pro Concepción, and C. Yañez Durán, “Compresión de imágenes: Fundamentos, técnicas y formatos,” *Revista de Ingeniería de Sistemas e Informática*, vol. 6, no. 1, pp. 21–29, 2009. [Online]. Available: <https://upsalesiana.ec/ing32ar1r01>
- [2] C. A. Ordoñez Santiago, “Formatos de imagen digital,” *Revista Digital Universitaria*, vol. 5, no. 7, 2005. [Online]. Available: <https://upsalesiana.ec/ing32ar1r02>
- [3] M. Al-khassaweneh and O. AlShorman, “Frei-chen bases based lossy digital image compression technique,” *Applied Computing and Informatics*, vol. 20, no. 1/2, pp. 105–118, 2024. [Online]. Available: <https://doi.org/10.1016/j.aci.2019.12.004>
- [4] AlShorman, O. M. Mahmoud, AlKhassaweneh, and Mahmood, “Lossy digital image compression technique using run-length encoding and frei-chen basis,” in *Universidad de Yarmouk*, 2012. [Online]. Available: <https://upsalesiana.ec/ing32ar1r4>
- [5] P. Chamorro-Posada, “A simple method for estimating the fractal dimension from digital images: The compression dimension,” *Chaos, Solitons & Fractals*, vol. 91, pp. 562–572, 2016. [Online]. Available: <https://doi.org/10.1016/j.chaos.2016.08.002>
- [6] L. Arranz, *Vector images and bitmaps*. Recursostic, 2005. [Online]. Available: <https://upsalesiana.ec/ing32ar1r6>
- [7] M. E. Ruiz Rivera and E. Yarasca Carranza, Juan Eduardo Ruiz Lizama, “Análisis de la compresión de imágenes utilizando clustering bajo el enfoque de colonia de hormigas,” *Industrial Data*, vol. 16, no. 2, pp. 118–131, 2013. [Online]. Available: <https://doi.org/10.15381/idata.v16i2.11929>
- [8] D. V. Rojatkhar, N. D. Borkar, B. R. Naik, and R. N. Peddiwar, “Image compression techniques: Lossy and lossless,” in *International Journal of Engineering Research and General Science*, vol. 3, no. 2, 2015, pp. 912–917. [Online]. Available: <https://upsalesiana.ec/ing32ar1r66>
- [9] S. M. Hardi, B. Angga, M. S. Lydia, I. Jaya, and J. T. Tarigan, “Comparative analysis run-length encoding algorithm and fibonacci code algorithm on image compression,” *Journal of Physics: Conference Series*, vol. 1235, no. 1, p. 012107, jun 2019. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1235/1/012107>

- [10] G. E. Belloch, *Introduction to Data Compression*. Computer Science Department. Carnegie Mellon University, 2013. [Online]. Available: <https://upsalesiana.ec/ing32ar1r10>
- [11] R. C. González and R. E. Woods, *Tratamiento digital de imágenes*. Madrid: Díaz de Santos, 1996. [Online]. Available: <https://upsalesiana.ec/ing32ar1r11>
- [12] A. AbuBaker, M. Eshtay, and M. AkhoZahia, “Comparison study of different lossy compression techniques applied on digital mammogram images,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 12, pp. 149–155, 2016. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2016.071220>
- [13] C. Ding, Y. Chen, Z. Liu, and T. Liu, “Implementation of grey image compression algorithm based on variation partial differential equation,” *Alexandria Engineering Journal*, vol. 59, no. 4, pp. 2705–2712, 2020, new trends of numerical and analytical methods for engineering problems. [Online]. Available: <https://doi.org/10.1016/j.aej.2020.05.012>
- [14] X. P. Alaitz Zabala, R. Díaz-Delgado, F. García, F. Auli-Llinas, and J. Serra-Sagrasta, “Effects of jpeg and jpeg2000 lossy compression on remote sensing image classification for mapping crops and forest areas,” *e Ministry of Science and Technology and the FEDER*, 2020. [Online]. Available: <https://upsalesiana.ec/ing32ar1r14>
- [15] M. C. Stamm and K. J. R. Liu, “Anti-forensics of digital image compression,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1050–1065, 2011. [Online]. Available: <http://doi.org/10.1109/TIFS.2011.2119314>
- [16] T. H. Thai, R. Cogramne, F. Retraint, and T.-N.-C. Doan, “Jpeg quantization step estimation and its applications to digital image forensics,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 123–133, 2017. [Online]. Available: <http://doi.org/10.1109/TIFS.2016.2604208>
- [17] L. González, J. Muro, M. del Fresno, and R. Barbuza, *Un enfoque para la compresión de imágenes médicas basado en regiones de interés y compensación de movimiento*. 4to Congreso Argentino de Informatica y Salud, CAIS 2013, 2013. [Online]. Available: <https://upsalesiana.ec/ing32ar1r17>
- [18] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin, and A. Bilgin, “The current role of image compression standards in medical imaging,” *Information*, vol. 8, no. 4, 2017. [Online]. Available: <https://doi.org/10.3390/info8040131>
- [19] M. A. Ameer Kadhum, “Compression the medical images using length coding method,” *Journal of Electrical and Electronics Engineering*, vol. 12, no. 3, pp. 94–98, 2017. [Online]. Available: <http://doi.org/10.9790/1676-1203029498>
- [20] Adobe. (2023) Elección de un formato de archivo. Adobe. All rights reserved. [Online]. Available: <https://upsalesiana.ec/ing32ar1r26>
- [21] C. K. Parmar and K. Pancholi, “A review on image compression techniques,” *Journal of Information, Knowledge and Research in Electrical Engineering*, vol. 2, no. 2, pp. 281–284, 2013. [Online]. Available: <https://upsalesiana.ec/ing32ar1r20>
- [22] D. Salomon, G. Motta, and D. Bryant, *Compresión de datos. La referencia completa*. Springer-Verlag London Limited, 2007. [Online]. Available: <https://upsalesiana.ec/ing32ar1r21>
- [23] W. Wahba and A. Maghari, “Lossless image compression techniques comparative study,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, 02 2016. [Online]. Available: <https://upsalesiana.ec/ing32ar1r22>
- [24] S. Dhawan, “A review of image compression and comparison of its algorithms,” *International Journal of Electronics & Communication Technology*, vol. 2, no. 1, pp. 22–26, 2011. [Online]. Available: <https://upsalesiana.ec/ing32ar1r23>
- [25] K. Sakshica and K. Gupta, “Various raster and vector image file formats,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 3, pp. 268–271, 2015. [Online]. Available: <http://doi.org/10.17148/IJARCCE.2015.4364>
- [26] A. K. Al-Janabi, “Efficient and simple scalable image compression algorithms,” *Ain Shams Engineering Journal*, vol. 10, no. 3, pp. 463–470, 2019. [Online]. Available: <https://doi.org/10.1016/j.asej.2019.01.008>
- [27] V. Barannik, S. Sidchenko, N. Barannik, and V. Barannik, “Development of the method for encoding service data in cryptocompression image representation systems,” *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 9, pp. 103–115, 2021. [Online]. Available: <https://doi.org/10.15587/1729-4061.2021.235521>
- [28] P. K. Pareek, C. Sridhar, R. Kalidoss, M. Aslam, M. Maheshwari, P. K. Shukla, and S. J. Nuagah,

- “Intopmicm: Intelligent medical image size reduction model,” *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 5171016, 2022. [Online]. Available: <https://doi.org/10.1155/2022/5171016>
- [29] X. Gao, J. Mou, S. Banerjee, and Y. Zhang, “Color-gray multi-image hybrid compression–encryption scheme based on bp neural network and knight tour,” *IEEE Transactions on Cybernetics*, vol. 53, no. 8, pp. 5037–5047, 2023. [Online]. Available: <https://doi.org/10.1109/TCYB.2023.3267785>
- [30] R. Kumar, P. Seetharaman, A. Luebs, I. Kumar, and K. Kumar, “High-fidelity audio compression with improved rvqgan,” *Advances in Neural Information Processing Systems*, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2306.06546>