

Procesamiento y gestión digital de la información

Marcos René Álvarez Moreno*

Resumen

Se describen los elementos básicos de los formatos de imágenes digitales, así como sus procesos de captura, almacenamiento y gestión. (FRRE)

PALABRAS CLAVE: imágenes digitales, formatos digitales, procesamiento de imágenes, gestión de imágenes electrónicas, escáneres, hardware, software.

Abstract

The basic elements of digital images formats, capture processes, storage and management are described along the document. (FRRE)

KEYWORDS: Digital Images, Digital Formats, Image Processing, Digital Images Management, Scanners, Hardware, Software.

Introducción

El impacto que la información tiene en el desarrollo de nuevas tecnologías para su procesamiento y gestión, ha llevado a la necesidad de crear procedimientos y/o técnicas apoyados en hardware y software que permitan gestionar y procesar la información.

La información se presenta como datos textuales o imágenes; la diferencia fundamental radica en su interpretación, es decir, en el significado de la información que está representada visualmente.

El procesamiento de la información considera que unas pocas operaciones simbólicas, relativamente básicas, tales como *codificar*, *comparar*, *localizar*, *almacenar* y *gestionar*, pueden crear conocimiento a través de la interpretación de imágenes (las cuales pueden contener texto en sí).

El procesamiento de la información contempla que los datos pueden ser codificados ó interpretados en diferentes modos o formatos. Este procesamiento es el término usado para denominar las operaciones desarrolladas sobre un conjunto de datos de imagen para mejorarlas de alguna forma, para ayudar a su

interpretación o para extraer algún tipo de información útil de ella.

El *formato digital* es el obtenido por dispositivos electrónicos, componentes hardware y software, que hacen posible la interpretación de una imagen (la cual puede contener texto) en formato entendible por éstos, así como para su visualización, una vez codificada la imagen, se le da el atributo de digital.

Las *imágenes digitales* son fotos electrónicas tomadas de una escena o escaneadas de documentos, fotografías, manuscritos, textos impresos e ilustraciones. Se realiza una muestra de la imagen digital y se confecciona un mapa de ella en forma de cuadrícula de puntos o elementos de la figura (píxeles). A cada *píxel* se le asigna un valor

* Departamento de Producción, Dirección General de Bibliotecas (DGB), Universidad Nacional Autónoma de México (UNAM), Edificio de la Biblioteca Central, Circuito Interior, Ciudad Universitaria, 04510 México, D. F. México. Correo electrónico: marcos@dgb.unam.mx



tonal (negro, blanco, matices de gris o color), el cual está representado en un código binario (ceros y unos), como se muestra en la figura 1. Los dígitos binarios ("bits") para cada píxel son almacenados por una computadora en una secuencia, y con frecuencia se los reduce a una representación matemática (comprimida). Luego la computadora interpreta y lee los bits para producir una versión analógica para su visualización o impresión.

1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	1	0	0	0	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	0	0	0	0	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	0	0	0	1	1	0	0	0	1
1	1	1	1	1	1	1	1	1	1

Figura 1. Valores de píxel: 0 para el negro y 1 para el blanco

Indudablemente el hecho de mantener y gestionar los contenidos en formato digital es ventajoso para el mantenimiento, actualización, almacenamiento y acceso a los mismos, lo cual plantea aspectos de calidad y utilidad de quien lo requiera, esperando una recuperación rápida, calidad aceptable, y visualización completa de las imágenes digitales.

Procesamiento de imágenes

La captura de imágenes digitales debe tomar en cuenta los procesos técnicos comprendidos al convertir una representación analógica en digital, así como también los atributos de los documentos fuente en sí mismos: dimensiones físicas y presentación, nivel de detalles, rango tonal, y presencia de color.

El procesamiento digital tiene como objetivo mejorar el aspecto de las imágenes y hacer más evidentes en ellas ciertos detalles que se desean hacer notar. La imagen puede haber sido generada de muchas maneras, por ejemplo, de forma manual, mecánica, fotográficamente representada por una cuadrícula con un esquema de puntos o líneas de diferente tamaño como por ejemplo *microfilms*, o electrónicamente, por medio de un escáner* y/o una computadora. Las clases generales de procesamiento incluyen la edición de la imagen (escalarla, comprimirla, otorgarle nitidez, etc.).

El procesamiento de imágenes no puede producir información a partir de nada; si en el conjunto de datos no existe información concerniente a una aplicación o interpretación en particular, entonces no importa qué cantidad de complicadas rutinas de

procesamiento apliquemos, no se podrá obtener información.

El equipo utilizado y su rendimiento afectarán la calidad de la imagen; diferentes sistemas con las mismas características declaradas (por ejemplo: dpi, profundidad de bits y rango dinámico) pueden producir resultados radicalmente diferentes. El rendimiento del sistema se determina por medio de pruebas que verifican la resolución, reproducción de tonos, calidad de colores, ruido y artefactos.

Los sistemas adquieren los datos pixel a píxel, la imagen es guardada como una gradilla de pixeles. La radiancia (sensores visibles) de cada uno de esos pixeles es registrada por el detector y entonces es digitalizada para permitir su manipulación por parte del computador. Las radiancias podrán tomar cualquier valor entre 0 y R_{max} , que es el número de intervalos digitales que están disponibles y dependen del número de bits del convertidor análogo-digital $R_{max}=2^n-1$, donde n es el número de bits.

Hasta aquí los datos han sido grabados, digitalizados y guardados en la forma de una gradilla de pixeles, si los datos fueron adquiridos por algún sensor capaz de producir imágenes, los datos pueden mostrarse en un monitor como se

* Se utiliza el término escáner para referirse a todos los dispositivos de captura de imágenes, incluyendo las cámaras digitales.



Figura 2. Imagen digital

muestra en la figura 2. Para que los datos sean mostrados en pantalla, el computador asigna a cada valor entero (o rango de ellos) un color en particular, de forma tal que pueda ser distinguido visualmente.

El formato de archivo para las imágenes originales deberá aceptar la resolución, profundidad de bits, información de color que se requieran. Por ejemplo, tiene poco sentido crear una imagen a todo color sólo para guardarla en un formato que no acepta más de 8 bits (por ejemplo: GIF (*Graphics Interchange Format*)).

El formato también deberá ser abierto y bien documentado, ampliamente soportado y compa-

tible en todas las plataformas de hardware-software conocidas.

A pesar de la existencia de otros formatos como PNG, SPIFF, GIF y JPEG, la mayoría confiamos en el formato TIFF para guardar imágenes originales, ya que éste es compatible en cualquier plataforma de software para su visualización. Para acceder a los gráficos se pueden crear imágenes derivadas en otros formatos. Los formatos más comunes se muestran en la tabla 1.

Las operaciones que se realizan en el procesamiento de imágenes pueden ser:

- Edición, retoque, mejora; en algunos casos el software de escaneo de las imágenes realiza estos pasos, en otros, se utilizan herramientas comerciales de edición de imágenes separadas, por ejemplo:

Nombre y versión actual	TIFF (Tagged Image File Format)	GIF (Graphics Interchange Format)	JPEG (Joint Photographic Expert Group) /JFIF (JPEG File Interchange Format)	Image Pac, Photo CD	PNG 1.2 (Portable Network Graphics)	PDF 1.3 (Portable Document Format)
Extensiones	.tif, .tiff	.gif	.jpeg, .jpg, .jif, .jfif	.pcd	.png	.pdf
Profundidad (es) de bits	Bitonal a 1 bit; escala de grises o color de paleta de 4 u 8 bits; hasta color de 64 bits	Bitonal, escala de grises o color entre 1 y 8 bits	Escala de grises a 8 bits; color a 24 bits	Color a 24 bits	1-48 bits; color a 8 bits, escala de grises a 16 bits, color a 48 para color	Escala de grises a 4 bits; color a 8 bits; soporta hasta 64 bits

Tabla 1. Formatos comunes de archivo de imágenes

Adobe Photoshop, Adobe Illustrator, Corel Photo Paint, Corel-draw X3, Paint Shop PRO, Macromedia Fireworks y herramientas libres como XnView, NConvert, GFL SDK y GFLAX.

- Compresión, (en ocasiones las imágenes pueden ser muy grandes lo que puede ocasionar problemas en el procesamiento de las mismas) algunas veces llevada a cabo por firmware (hardware-software) del escáner o hardware de la computadora. La compresión también puede ser una operación sólo de software, a pesar de que el hardware dedicado es más rápido y se lo debería considerar cuando se crean archivos muy grandes o grandes cantidades de archivos.
- Conversión de formato de archivo, el escaneado original de la imagen puede no estar en un formato adecuado para todos los usos previstos, por lo que requiere conversión.
- Escala, es probable que los escaneados capturados a alta resolución no sean adecuados para la visualización en la pantalla. Con frecuencia se necesita aplicar escala (es decir, reducción de resolución a través de eliminación de bits) para poder crear imágenes y presentarlas en cualquier aplicación como por ejemplo en la web a través de un visualizador (Internet Explorer, Netscape, Mozilla, Opera, etc.).

- OCR (reconocimiento óptico de caracteres), conversión de texto escaneado a texto legible por medio de una máquina, que se puede buscar o indexar y ser interpretado como una imagen.

La digitalización deja los datos con un formato adecuado para su manipulación por parte del computador. Cada sistema de procesamiento de imágenes tiene su propio software asociado, cada uno para una tarea diferente.

La utilidad de las imágenes digitales queda garantizada cuando se definen claramente las necesidades, se conocen los atributos de los documentos, la infraestructura técnica de apoyo del procesamiento y posterior gestión, y entrega del contenido sea la apropiada para las necesidades del proyecto.

La infraestructura técnica se refiere en forma general a los componentes que hacen posible la digitalización de imágenes. Con frecuencia, el proceso completo se lleva a cabo mediante técnicas de digitalización, sugiriendo así una serie de pasos lógicamente ordenados. En la práctica real, estas técnicas de digitalización pueden ser simples o complejas dependiendo de las herramientas que se tengan para llevarlas a cabo.

La tecnología necesaria para navegar desde un extremo de las técnicas de digitalización al otro consta

principalmente de hardware y software como ya se ha mencionado anteriormente. Una perspectiva integral de la infraestructura técnica también incluye protocolos y normas, políticas y procedimientos.

Las decisiones en lo que respecta a la infraestructura técnica requieren una planificación cuidadosa debido a que la tecnología de la digitalización de imágenes cambia rápidamente. Si los equipos y materiales elegidos son los indicados para los usos previstos y los resultados esperados, el rendimiento, presentación y visualización de la información se maximizará en todos los aspectos.

Gestión digital de la información

Una vez digitalizadas las imágenes, viene la parte de la gestión; la gestión de archivos que contienen la información digitalizada consiste en una serie de pasos interrelacionados, diseñados para asegurar la fácil identificación, organización, acceso y mantenimiento de los archivos dependiendo de los fines de la información. Comprende el proceso de hacer llegar las imágenes al usuario y abarca redes, software, dispositivos de visualización, impresoras, escáner, etc.

La gestión digital de la información incluye:

Procesamiento y gestión digital de la información

Soluciones de gestión de imágenes (software especial para organizar archivos de imágenes); como por ejemplo, a partir de impresión en papel de imágenes (que pueden tener texto) contenidos en materiales plásticos llamados "microfilms" (ver figura 3) capaces de proyectarse a través de un lente y un haz de luz luminoso contenidos en un tipo escáner, como por ejemplo en un *PC printer* (ver figura 4).



Figura 3. Microfilms



Figura 4. PC Printer 80

Almacenamiento (dispositivos y medios); como por ejemplo, soportes accesibles a una computadora



Figura 5. Almacenamiento en CD-ROM de la información

(CD-ROM, ver figura 5) de modo que pueda ser de fácil edición y acceso.

La innovación tecnológica ha permitido crear dispositivos llamados escáner de quinta generación capaces de procesar imágenes (incluyendo texto) y almacenarlo en un formato definido por el usuario. Escáner como el Canon DR5020 (ver figura 6), Fujitsu fi 5750C (ver figura 7) ó el más reciente microfilm escáner 350 de Canon que permite a través de la visualización de un cuerpo opaco (microfilm) la digitalización directa en computadora en determinado formato (ver figura 8 y 9).

Esta tecnología permite la rápida recuperación en formato electrónico de imágenes; que son fuente de información digital (una base de datos, una publicación electrónica, etc.), esta a su vez puede ser distribuida de diversas formas sin necesidad de cambiar sus contenidos ni estructura internas.

Las formas básicas son dos, aunque ambas puedan subdividirse en algunas más:

1. En línea (online) o vía telemática, que a su vez puede adoptar las siguientes variaciones:
 - a. En formato web o html, que es la forma mayoritaria.
 - b. En formato ASCII puro.
2. Fuera de línea o en soporte material
 - a. CD-ROM (ver figura 5).



Figura 6. Scanner Canon DR5020



Figura 7. Scanner Fujitsu 5750C



Figura 8. Scanner 350 de Canon

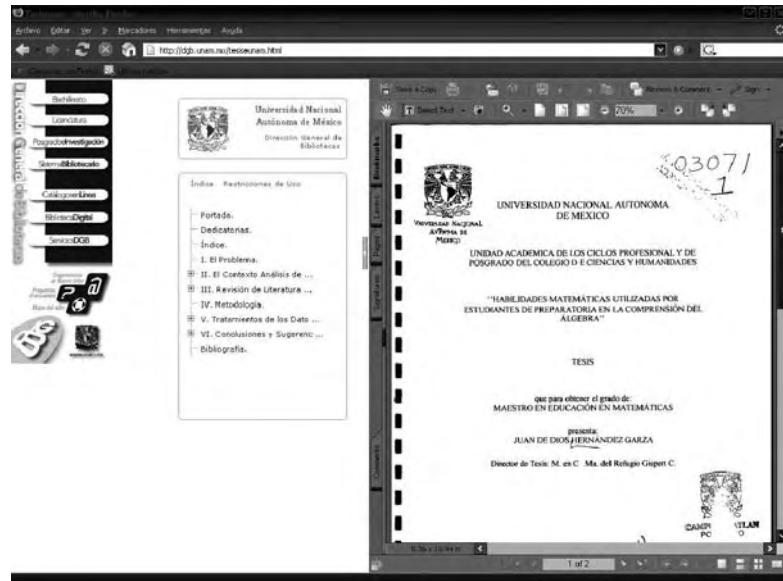


Figura 10. Búsqueda en línea de la información



Figura 9. Digitalización desde microfilm



Figura 11. Consulta electrónica de la información

Como ejemplo de aplicación de estas formas básicas para la gestión digital de la información se tienen: interfaces gráficas en web que apoyadas de un lenguaje de programación permiten realizar la búsqueda y visualización de la información en diferentes formatos, como PDF (ver figura 10 y 11).

Para la distribución fuera de línea, se manejan soportes en medios ópticos (CD-ROM, ver figura 5) de la informa-

ción, que solo pueden ser de consulta y no distribuidas para su reproducción, por los derechos de autor.

En la figura 12 se puede apreciar de forma general como se da la digitalización y de que forma se puede gestionar la información para que el usuario final pueda visualizarla.

Procesamiento y gestión digital de la información

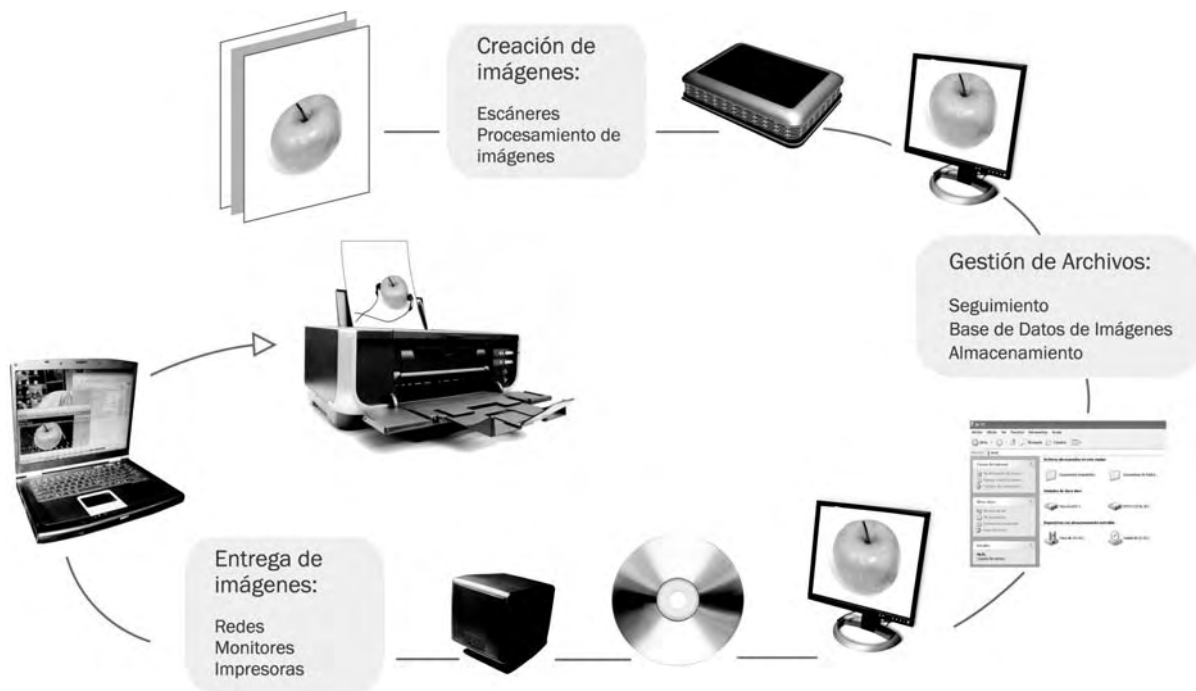


Figura 12. Gestión y digitalización de la información

Estos soportes informáticos para el procesamiento de la información son mucho más rápidos y eficientes para la organización y la creación de métodos de acceso a las imágenes cuya finalidad es apoyar en proyectos enfocados a la edición electrónica así como a la gestión digital.

El procesamiento de imágenes no pretende "copiar" la producción impresa, sino que debe generar una nueva estructura de la información. El uso de documentos de hipertexto hacen que una imagen ya no sea lineal, como ocurre con la confección en papel, sino que la imagen ahora también posee "profundidad". El usuario tiene acceso a la información

de formas muy variadas, cuenta con la posibilidad de proveer vínculos a documentos de modo instantáneo, lo cual permite explicar dichos contenidos de formas diversas.

El procesamiento y gestión digital de la información cambia la forma tradicional de hacer las cosas, mediante el uso de equipo de cómputo, lenguajes de programación, las telecomunicaciones y las bases de datos; se ha constituido el primer producto de información electrónica.

Utilizando eficientemente el procesamiento de la información así como su gestión, se pueden obtener ventajas como la implementación de bibliote-

cas digitales, sitios de Internet que permitan consultar artículos, libros, etc., cuyos elementos base, son de carácter electrónico.

Conclusiones

En el procesamiento y gestión digital de la información se debe llevar a cabo un análisis funcional y de operación que contemple los diferentes elementos lógicos y físicos así como de su relación entre estos; lenguajes de alto nivel, programas comerciales, software libre, y una arquitectura de cómputo que permita la accesibilidad de la información en medios electrónicos.

Obras consultadas

ABADAL FALGUERAS, Ernest. *Sistemas y servicios de información digital*. Gijón : Trea, 2001. 147 p.

BOLOGNA, Jack, WALSH, Anthony M. *The Accountant's Handbook of Information Technology*. New York : Wiley, c1997. 368 p.

BORGMAN, Christine L. *From Gutenberg to the Global Information Infrastructure : access to information in the networked world*. Cambridge, Massachusetts : MIT, 2003. 324 p.

MORRIS, Tim. *Computer Vision and Image Processing*. Basingstoke, Hampshire : Palgrave Macmillan, 2004. 300 p.

PADOVA, Ted. *Adobe Acrobat 7 PDF Bible*. Indianapolis, Indiana : Wiley, c2005. 896 p.

PEÑA Rosalía. *Gestión digital de la información : de bits a bibliotecas digitales y la web*. México : Alfaomega, 2003. 440 p.

RUSS, John C. *The Image Processing Handbook*. Boca Raton : CRC, 2004. 732p.

WITTEN Ian H., BAINBRIDGE David. *How to Build a Digital Library*. Amsterdam : M. Kaufmann, c2003. 518 p.

