

Repositorios digitales institucionales

Diseño, implementación y optimización de un recurso estratégico para las Universidades - 2012/2013

Módulo 3: Aspectos técnicos y tecnológicos del repositorio

Bloque 3.1: Generación y tratamiento de objetos digitales

Lic. Martín Williman FCEN-UBA

36

GIMP

PARTE 1: APRENDIENDO LO TECNICO

36

36A

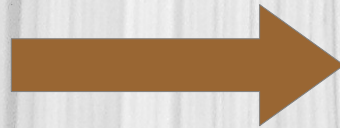
DIGITALIZACIÓN

**OBJETO
ANALÓGICO**



Digitalizar

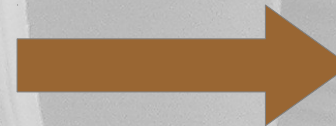
**ARCHIVO
INFORMATICO**



Representado

NÚMEROS

Representado



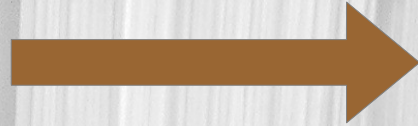
**Soporte con
secuencias de
0 y 1**

“Digitalizar es representar una información mediante números”



¿Por que digitalizar?

**DOCUMENTOS
DIGITALES**



“Son ubicuos”



Facilitan

Acceso

Almacenamiento

Reproducción

Consulta

Conservación

Gestión

Difusión

Imagen digital

Dos sistemas para la representación interna de una imagen:

•Raster

- Matriz de puntos (píxeles) - Trama
- Cada punto tiene información de posición y de valores de luz
- Representan cualquier imagen (+)
- Sus archivos son pesados (-)

•Vectorial

- Cada línea es un vector
- Posee coordenadas y comandos
- Sus archivos son livianos (+)
- Solo representa objetos geométricos (-)

Imagen digital

Raster

Vectorial



Imagen digital

Raster



Vectorial



Raster: Captura de imagen

Dos fases:

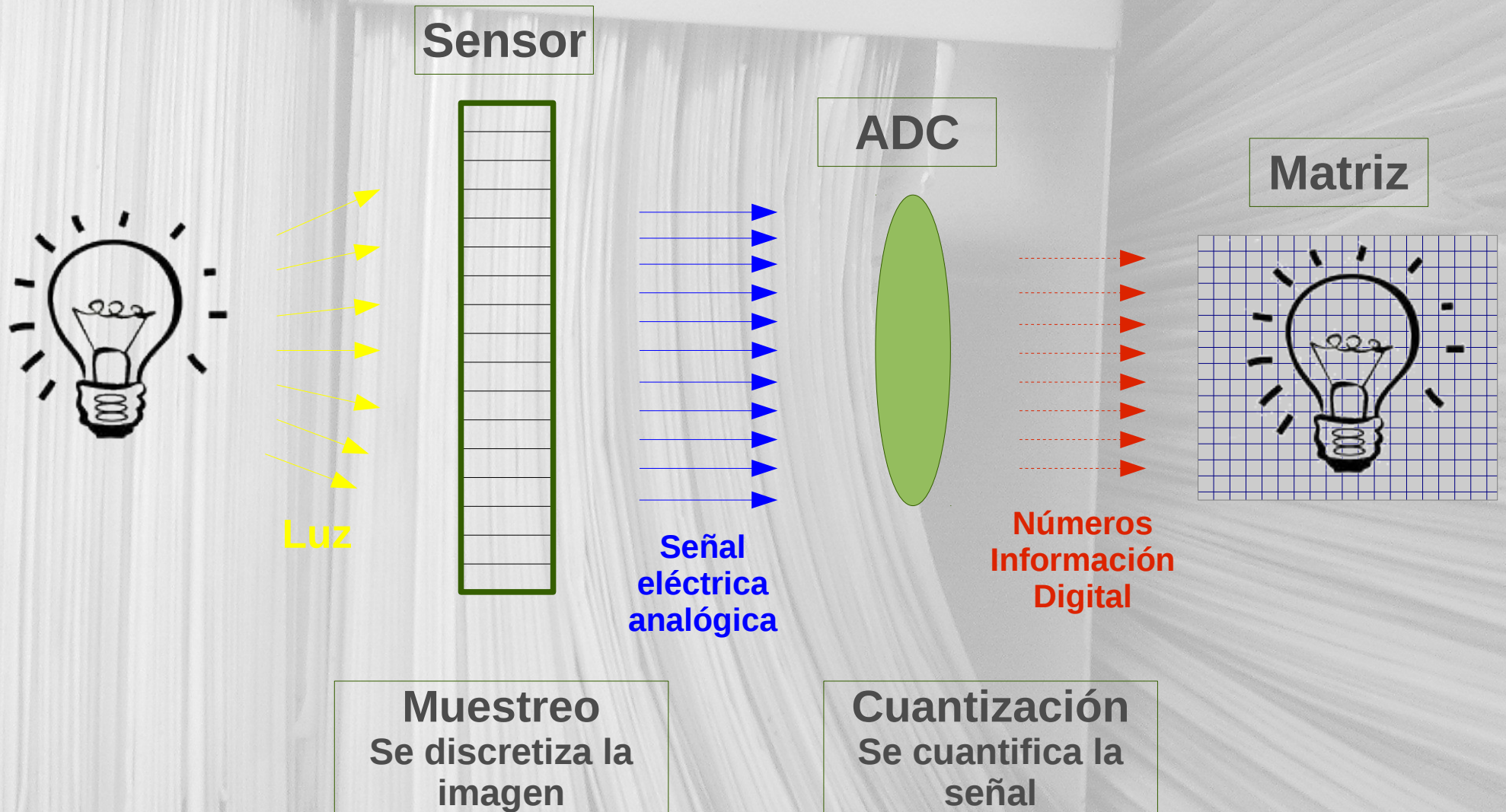
- **Muestreo**

- Descomponer la imagen natural continua en un número finito de puntos
- Realizado por el Sensor - Celdillas fotoeléctricas

- **Cuantización**

- A cada punto en el que fue descompuesta la imagen, se le asigna un valor numérico que representa la intensidad de la luz recibida
- Realizado por ADC (Convertor analógico digital)

Raster: Captura de imagen



Raster: Parámetros

Resolución espacial: Numero de puntos que conforman la imagen (píxeles).

- **Absoluta**

- Número total de píxeles
- Ej 600 X 600 px = 360.000 de px

- **Relativa**

- Número de píxeles por unidad de medida
- Normalmente se expresa en píxeles por pulgada (ppp)
- Puede se de Captura o de Salida
- Ej 300 ppp

Raster: Parámetros

Resolución espacial

- **Óptica**

- Número de puntos reales tomados por el dispositivo de captura

- **Interpolada**

- Mediante algoritmos de interpolación, se agregan nuevos píxeles como promedio de los píxeles reales, que no poseen nueva información.
 - Para impresión y visualización por pantalla
 - Produce deterioro en la imagen

Raster: Parámetros

¿Que significa que un escáner tenga una RO de 600 ppp?

1 pulgada es 2,54 cm.

$600 \text{ px} / 2,54 \text{ cm} = 236 \text{ píxeles por centímetro}$

Por cada centímetro de la imagen el escáner toma información sobre la luz en 236 puntos

Considerando que el escáner tiene $RH = RV$, por cada cm^2 el escáner toma 55696

Cada pixel tiene una área de $0,000017955 \text{ cm}^2$

Raster: Cálculos



1920
Estudiante
ingeniería
Perú 222

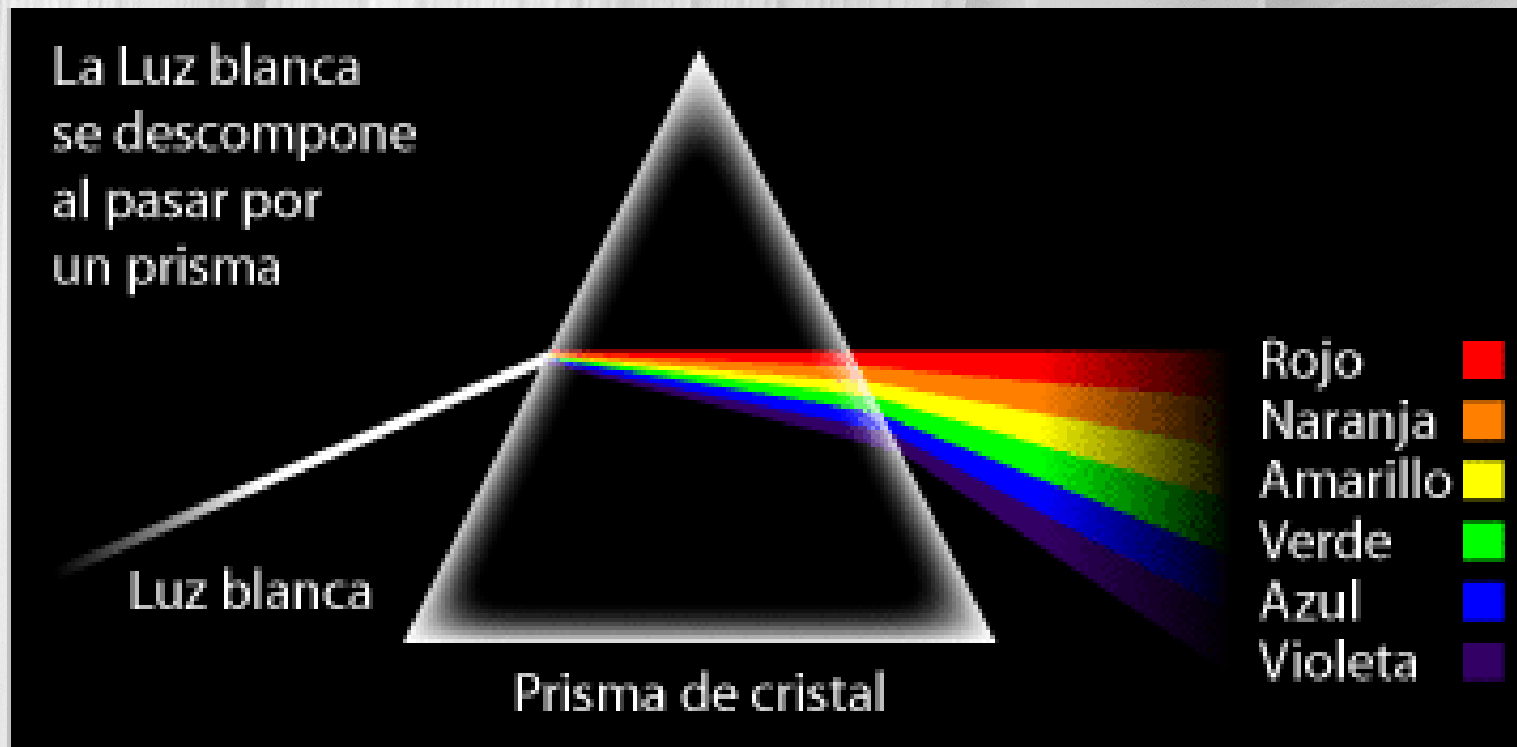
15 cm lado largo
Escaneada a 300 ppp
 $15 \text{ cm} / 2,5 \text{ cm} = 6 \text{ pulgadas}$
 $6 \text{ p} \times 300 \text{ ppp} = 1800 \text{ px}$
Monitor 1280 X 1024

Regla: 3000 px en el lado largo

1. Mido el lado largo de la fotografía
2. En este caso 15 cm
3. Paso a pulgadas $15 \text{ cm} / 2,5 \text{ cm} = 6 \text{ pulgadas}$
4. $3000 \text{ px} / 6 \text{ pulgadas} = 500 \text{ ppp}$

COLOR

Naturaleza de la luz



COLOR

- **Bit**

- Bit es un acrónimo del término Binary digit (dígito binario)
- El Bit es la unidad mínima de representación de información digital empleada en dispositivos informáticos
- Un bit es un biestable, que permite dos posibles estados:
 - activado / desactivado
 - 0 / 1
 - Blanco / Negro

COLOR

¿Como representar con números el color de un píxel?

Cuatro niveles de escala de color

- **Imágenes bitonales**

- 1 bits para la representación del color en cada píxel => dos colores (blanco/negro)
- Archivos muy livianos
- Limitadas para la representación de imágenes reales
- Con la utilización de tramas se pueden representar grises, pero pierden capacidad de compresión

COLOR

Imágenes bitonales

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	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



COLOR

- **Escala de grises**

- Se representan niveles de luminosidad, no de colores
- La cantidad de bits utilizados para representar el nivel de luminosidad en cada píxel - **“Profundidad de bits”**
- 4 bits ==> 16 niveles (2^4)
- 8 bits ==> 256 niveles (2^8)

COLOR

Escala de grises



Nº de canales
Profundidad

=1 (gama de grises)
=8 bit

00110111

COLOR

- **Paletas de colores**

- Se define una paleta de colores para cada imagen (LUT)
- Cada píxel tiene un número que referencia a un color de la paleta
- La cantidad de bits utilizados para representar el color en cada píxel - **“Profundidad de bits”**
- 4 bits ==> 16 colores (2^4)
- 8 bits ==> 256 colores (2^8)

COLOR

- **Color Real**

- Puede registrar el color del espectro visible para el ojo humano y más
- Cada píxel esta representado por un triplete de distintas intensidades de tres colores (canales) red - green - blue (RGB)
- Cada canal 8 bits ==> 24 bits de profundidad (fotografía, escaneado)
- 1 pixel ==> (8 bits R; 8 bits G; 8 bits B)
(256 R; 256 G; 256 B)
(**16,7 millones de colores**)

COLOR

- **Color Real**



00110010

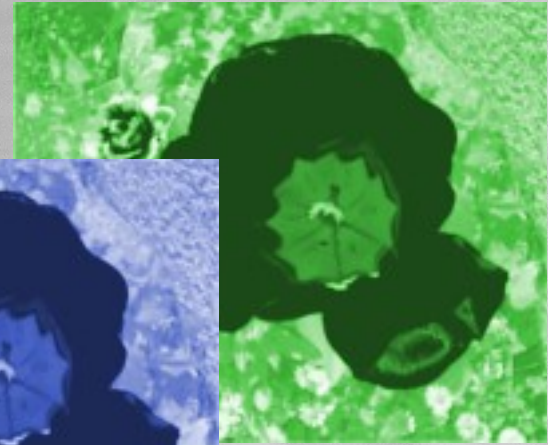
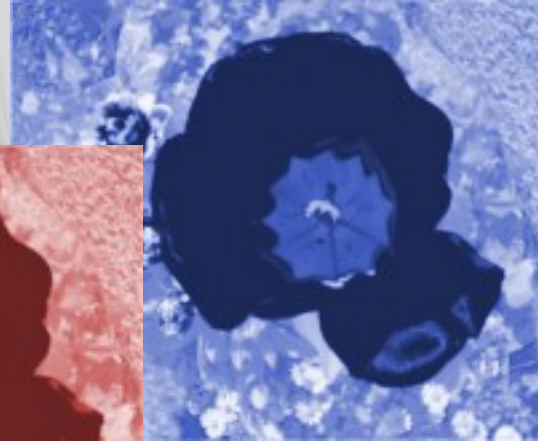
Canal R

11001010

Canal B

11000111

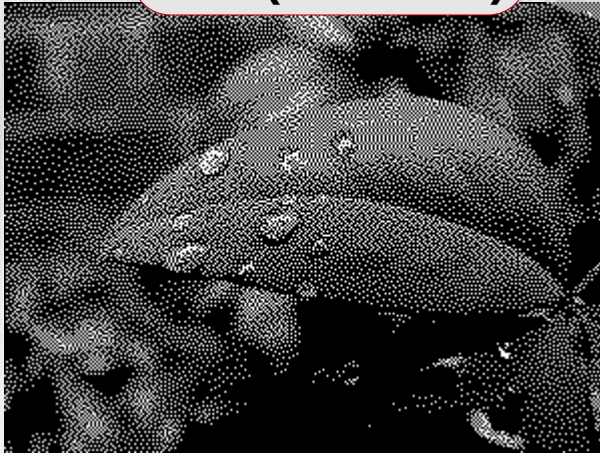
Canal G



COLOR

- Bitonal – Escala de grises – Paleta - Color Real

1 bit (2 colores)



2 bit (4 colores)



4 bit (16 colores)



8 bit (256 colores)



24 bit (16 M de colores)



COLOR

¡Más Cálculos!

15 cm lado largo X 10cm lado corto
Escaneada a 600 ppp

$15 \text{ cm} / 2,5 \text{ cm} = 6 \text{ pulgadas}$
 $6 \text{ p} \times 500 \text{ ppp} = 3000 \text{ px}$
 $10 \text{ cm} / 2,5 \text{ cm} = 4 \text{ pulgadas}$
 $4 \text{ p} \times 500 \text{ ppp} = 2000 \text{ px}$

$3000\text{px} \times 2000\text{px} = 6000000 \text{ px}$

$b/n \Rightarrow 6000000 \times 1 \text{ bits} = 732,4 \text{ KB}$

$\text{Grises} \Rightarrow 6000000 \times 8 \text{ bits} = 5,7 \text{ MB}$

$\text{Color RGB} \Rightarrow 6000000 \times 24 \text{ bits} = 17,2 \text{ MB}$

$\text{Color CMYK} \Rightarrow 6000000 \times 32 \text{ bits} = 22,9 \text{ MB}$



1920
Estudiante
ingeniería
Perú 222

COLOR

¡Más Calculos!

Tamaño de archivo en bits = (Número de píxeles de ancho \times número de píxeles de alto) \times número de bits de profundidad

El resultado sale en bits, para pasarlo a Bytes, dividirlo entre 8; para pasarlo a KB dividir el resultado en Bytes entre 1024 KB; para pasarlo a MB, dividir el resultado en KB entre 1024.



1920
Estudiante
ingeniería
Perú 222

COLOR

Gestion de color (CMS)



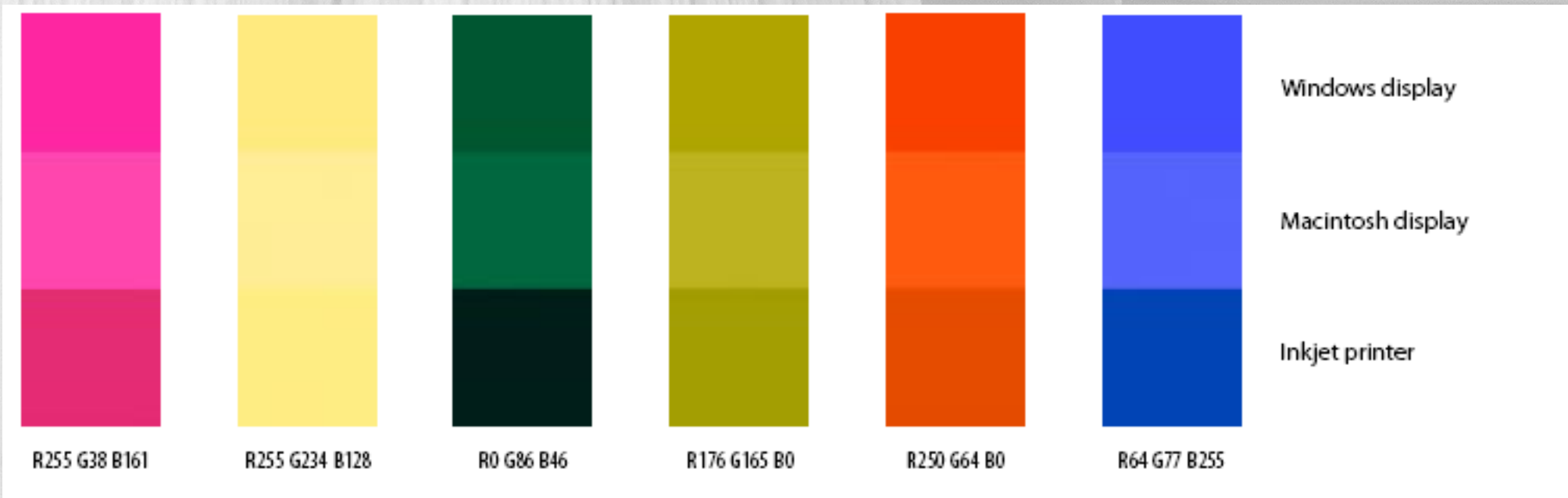
Busca alcanzar la correspondencia de color entre los distintos dispositivos



COLOR

Gestión de color (CMS)

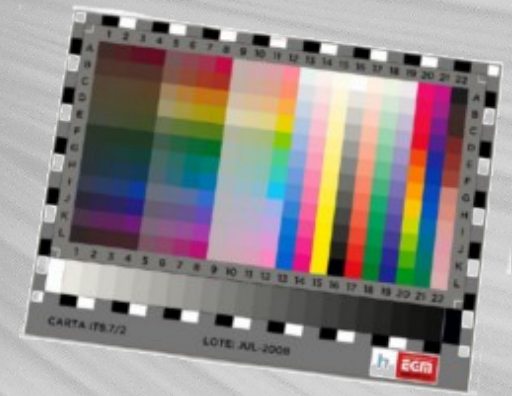
- Diferentes dispositivos
- Diferentes sistemas de color
- Calibración



COLOR

Perfiles de Color

- Son específicos de cada dispositivo en un lugar y momento determinado
 - Monitor ==> Espectrofotómetro
 - Escaner ==> Cartas de color
 - Camara ==> Cartas de color
 - Impresora ==> Hardware \$\$\$ - servicio
- Archivo con extensión .icc / .icm
- Hay perfiles por emulación del funcionamiento de un dispositivo



EQUIPOS DE CAPTURA

- **Escáneres**

- Mas usado en documentos textuales y gráficos
- Están sufriendo un retroceso en el mercado
 - No hay innovaciones
 - Muchas marcas no producen mas los modelos profesionales
 - Los documentos ya nacen digitales
 - Muy costosos

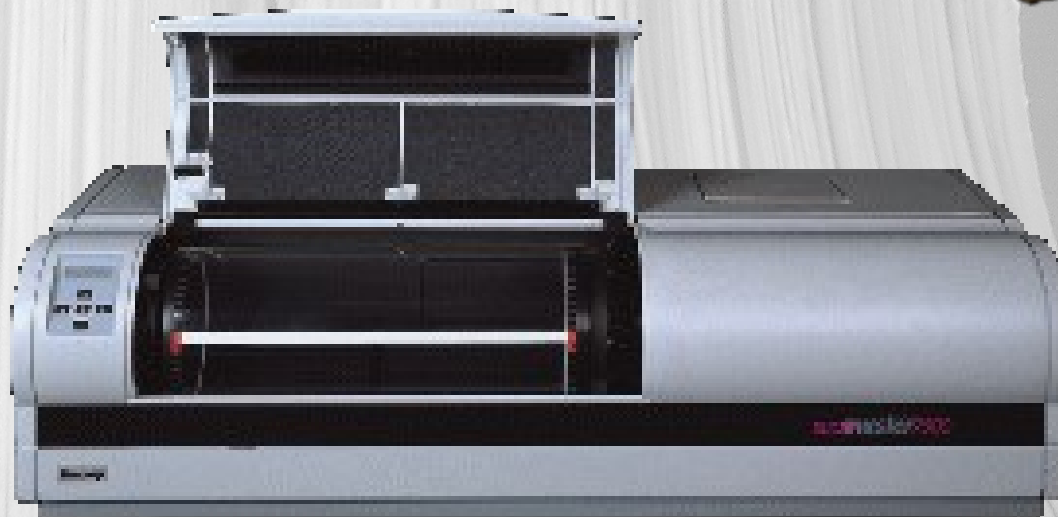
EQUIPOS DE CAPTURA

- **Escaneres de Tambor**

- Obtiene las imagenes de mayor calidad
- Ya casi no se fabrican, muy costosos
- Requieren operadores calificados
- RO de 8000 ppp
- No sirven para materiales rígidos
- Es necesario el contacto completo entre el material y el tamborno. Se utilizan productos aceitosos que

EQUIPOS DE CAPTURA

Escáneres de Tambor



EQUIPOS DE CAPTURA

- **Escáneres Plano**

- Para materiales opacos y transparentes
- Menos costosos
- Fáciles de manejar
- En el mercado se los encuentra en tres gamas
 - Baja: Uso domestico
 - Media: RO de 1200 ppp – 2400 ppp
 - Alta: Ya no hay en mercado

EQUIPOS DE CAPTURA

Escaneres Plano



EQUIPOS DE CAPTURA

- **Cámaras Fotográficas**

- Están ocupando el espacio dejado por los escáneres
- Excelente alternativa al escáner planetario
- Difícil control de la luz
- Necesidad de un soporte estable



EQUIPOS DE CAPTURA

- **Evaluación del equipo**
 - Resolución espacial real
 - Rango dinámico
 - Relación señal/ruido
 - Luz: componente uv
 - Temperatura
 - Velocidad
 - Costo
 - Mantenimiento
 - Garantía
 - Servicio técnico

EQUIPOS DE CAPTURA

Evaluación del equipo

	Bookeye ® 4 Professional	CopiBook™ HD
Resolución espacial	600 X 400 dpi	600 X 600 dpi
Profundidad de bit de captura y salida	36 bit color 12 bit escala de grises	36 bit color 12 bit escala de grises
Tipo de material que puede digitalizar	Periódicos, revistas, catálogos, carpetas de archivos, documentos encuadernados y con grapas, libros y documentos históricos frágiles. Escaneado del libro en V. Formato A2	Documentos frágiles, libros antiguos, material delicado, formatos grandes, encuadernación rígida. Formato A2
Rango dinámico (también margen de densidades o dMax)	No se informa	No se informa
Rendimiento, en cuanto a velocidad de escaneado, transferencia y almacenamiento	A2+300 dpi 4.0 s A2+400 dpi 4.6 s A2+600 dpi 6.8 s	A2+300 dpi color 3,4 s A2+400 dpi color 4,7 s A2+600 dpi color 7,7 s

EQUIPOS DE CAPTURA

Evaluación del equipo

Conectividad a PC	Puerto USB Conexión a red Conexión inalámbrica	Puerto USB Conexión a red
Generación de calor durante el escaneado	+5 a +40 °C	+10°C a +30°C
Tipo de iluminación (fría, con radiación UV...)	Lampara LED fría, sin emisión IR/UV Tiempo de vida: 50.000 h	Sin iluminación adicional
Posibilidad de acoplar adaptadores (alimentadores automáticos, adaptadores de transparencia)	Interruptor de pie para accionar el disparo.	Se le puede agregar lamparas. Interruptor de pie para accionar el disparo. Ratón.
Garantía y Soporte técnico	12 meses	12 meses extensible a 5 años

EQUIPOS DE CAPTURA

Evaluación del equipo

Estabilidad ante fallos	No se informa	No se informa
Precio	No se informa	No se informa
Sistemas operativos con los que funciona	Windows, Linux, Mac, OS, Unix	Linux
Otros aspectos que consideres interesantes para su posible aplicación en una digitalización de documentos	Camera CCD, 22.500 píxeles Perfil de color ICC Humedad relativa: 20 to 80 % (sin condensar) Ruido: < 42 dB(A) (escaneando)	Camara CCD: 600 dpi. Gestión de Perfil de color ICC. 1 terabytes de disco duro. Ajustes totalmente automáticos: foco, tiempo de exposición, balance de blancos. Apertura de vidrio automático con ajuste de presión sin riesgo.

36

GIMP

PARTE 2: PLANIFICANDO

36

36A

Almacenamiento

- **Aspectos a valorar**

- Volumen: Total, master y derivados
- Modo de almacenamiento: on line, off line, near line
- Velocidad: Lectura, escritura, modificación y transferencia
- Infraestructura informática y de red
- Espacio físico: área, control de temperatura y humedad, aislamiento
- Seguridad
- Recursos humanos
- Presupuesto: almacenamiento, mantenimiento
- Vida útil del soporte
- Fiabilidad / fallo

Almacenamiento

→ Máster

- Contiene toda la información del original relevante para el objetivo del proyecto
- Fiel al original para ser soporte de seguridad en caso de pérdida del original
- Archivos muy grandes
- Formatos
 - TIFF: Todo tipo de imágenes, compresión sin pérdida G3 y G4
 - JPG2000: Múltiples canales, hasta 32 bits. Sistema de compresión c/s pérdida mejorado. Con creciente aceptación

Almacenamiento

→ Derivados

- Obtenidos automáticamente a partir del máster
- Objetivo: Transmisión rápida y lectura agil
- Se aplican procesos de compresión sobre el máster
- Archivos entre 15 y 30 veces mas pequeños que el máster
- Formatos
 - JPEG: Son una serie de algoritmos de compresión
 - PNG: Código abierto. Permite transferencia y descarga progresiva
 - PDF/a: Preservación a largo plazo. No admite audio, vídeo ni ejecutables

Almacenamiento

→ Compresión

- Reducción del tamaño del fichero facilitando el almacenamiento y la transmisión
- Reducir a la cantidad de datos necesarios para describir eficazmente la imagen
- Eliminación de datos redundantes
- Ratio – Razón -Taza de compresión
Expresa la diferencia de tamaño entre el fichero antes y después de la compresión
Ej 2:1
- Dos tipos
 - Sin pérdida
 - Con pérdida

Almacenamiento

→ Compresión sin pérdida

- Luego de descomprimir el archivo, este queda igual al original
- Ratios bajos, menores a 2:1 en imágenes complejas
- Muy buenos en imágenes bitonales y texto
- Algoritmos
 - LZW
 - Huffman (TIFF G3 y G4)
 - RLE
 - JBIG

Almacenamiento

Compresión con pérdida

- Hay destrucción de información durante el proceso de compresión
- Cuando se descomprime, el archivo es de menor calidad que el original
- Alcanza altas tasas de compresión
- JPEG (mas usado)
 - Elimina lo redundante para el ojo humano
 - Produce distorsiones en la imagen: Halos al rededor de los bordes
 - No recomendable en dibujos de linea
 - JPEG2000 corrige estos problemas

Almacenamiento

- **Tipos de soporte**

- Dos tipologías

- Naturaleza magnética

- Naturaleza óptica

Almacenamiento

- **RAID (redundant array of independent disks)**
 - On line
 - Conjunto de discos duros acoplados
 - Contempla la duplicación de la información:
Almacenamiento redundante
 - Mayor tolerancia a fallo: recuperación automática de datos
 - Ocupan poco espacio físico
 - Controlador: software o hardware
 - Sustitución y exención en caliente
 - No evita la necesidad de backup

Almacenamiento

- **Cintas electromagnéticas**

- Off line / Near line (mecanismo robotizado)
- Muy económicas
- Librerías de cintas: Muchas capacidad (PB)
- Acceso secuencial a datos: Lentas
- Problemas de compatibilidad
- Usados como backup

Almacenamiento

- **Sistemas ópticos**

- Off line / Near line
- Evolución continua: cd, dvd, blu-ray
- Muy económicos, también las unidades de lectura
- Librerías de discos: hasta 600 discos
- Lentos
- Los datos no son editables
- En desuso
- CD : 700MB
- CD-RW : regrabable no aleatorio
- DVD : 1,4 GB – 10 GB
- Blue-ray : 25 GB – 50GB
- Vida media corta

Almacenamiento

- **Almacenamiento en la nube (Cloud storage)**
 - Creciente aceptación
 - Almacenar, compartir, sincronizar dispositivos
 - Obtención inmediata de un sistema de almacenamiento gestionado y con seguridad
 - Redundancia geográfica
 - Delegación de los objetos y su control
 - Continuidad del servicio
 - Algunos servicios, uso domestico gratuito
 - Dropbox
 - SugarSync
 - MegaCloud
 - Google Drive

Almacenamiento

- **Volumen**

$N^{\circ} \text{ de doc.} \times \text{ peso medio del doc.} \times \text{ Factor de ponderación} / \text{ capacidad del soporte} = \text{ unidades de soporte}$

Confusión de la capacidad del soporte por sistema en base decimal o en base dos ==> 7% menos

Almacenamiento

- **Considerar**

- No es posible señalar un soporte infalible a fallo
- Las tecnologías de almacenamiento cambian constantemente diseñados para durar un tiempo limitado en un momento de tecnología cambiante (5 años)
- No pensar en un sistema para un período de tiempo largo, sino un sistema FIABLE y POTENTE (velocidad y capacidad)
- En el presupuesto hay que reflejar la necesidad de renovación tecnológica

OAIS

