

# Competência Informacional

## 3º Módulo – Preservação Digital

### Tema 2 - Digitalização

Denis Novaes Lopes

*Engenheiro Químico / Analista de Sistemas*



Ministério da  
**Ciência, Tecnologia  
e Inovação**



# SUMÁRIO

- ▶ Conceitos Digitais;
- ▶ Primórdios digitalização;
- ▶ Digitalização;
- ▶ Digitalização Gráfica;
- ▶ Sensores;
- ▶ Interface para escâner;
- ▶ Estrutura Física;
- ▶ OCR(Optical Character Recognition)



# CONCEITOS DIGITAIS



# CONCEITOS DIGITAIS

- Que são Twain?
- Que são Cores?
- Como as imagens digitais são definidas?



# CONCEITOS DIGITAIS

## Drivers específicos (Twain)

- Controle de Resolução
- Controle de Cor
- Ferramentas Auxiliares



# CONCEITOS DIGITAIS

Driver: **TWAIN** (**T**echnology **W**ithout **A**n Interesting **N**ame):

- softwares específicos que viabilizam o reconhecimento do equipamento scanner pelo computador,





# CONCEITOS DIGITAIS

## RESOLUÇÃO:

- ▶ A resolução de uma imagem está associada ao seu nível de detalhes, e pode ser mensurado em dpi (dots per inch).
- ▶ Em documentos administrativos normalmente varia de 200 a 400 dpi, enquanto que no setor de artes gráficas e engenharia as resoluções usadas podem passar de 10000 dpi.



# CONCEITOS DIGITAIS

## RESOLUÇÃO:

- Para obtenção de uma imagem de melhor qualidade é aconselhável que seja utilizado no *reconhecimento ótico* - OCR (optical character recognition), uma resolução maior que 300 dpi.
- O aumento dos detalhes da imagem é proporcional à elevação da quantidade de pontos por polegada (dpi),





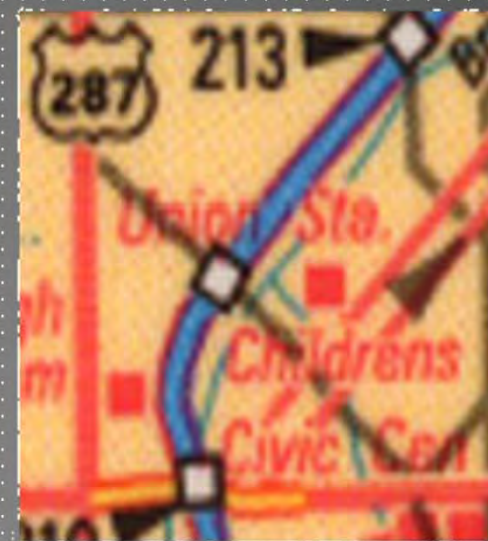
# CONCEITOS DIGITAIS

RESOLUÇÃO:

300 DPI



72 DPI



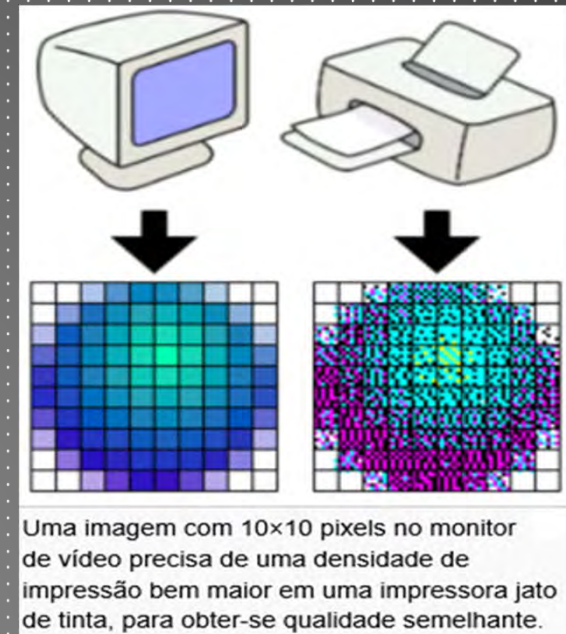
# PPI OU (PIXELS POR POLEGADA)

- Pixel é abreviação de "picture element". *É o menor elemento físico de um dispositivo de vídeo digital que o olho pode discernir.*
- Aproximando-se muito da tela do seu monitor, nota-se micro pontos, estes são os pixels. É uma descrição do número lógico de pixels da imagem original (X pixels por Y pixels)



# DPI (PONTOS POR POLEGADA) VS PPI (PIXELS POR POLEGADA)

- As impressoras processam quatro canais de cores (CMYK), cada um deles com uma quantidade fixa de tinta por canal, enquanto um monitor de vídeo forma um pixel utilizando três canais de cores (RGB), cada um deles com 256 níveis de luminosidade diferentes.
- O número de cores que podem ser gerados pela impressora em um ponto é  $2^4 = 16$ , enquanto um monitor de vídeo pode gerar  $256^3 = 16.777.216$  pixels diferentes.



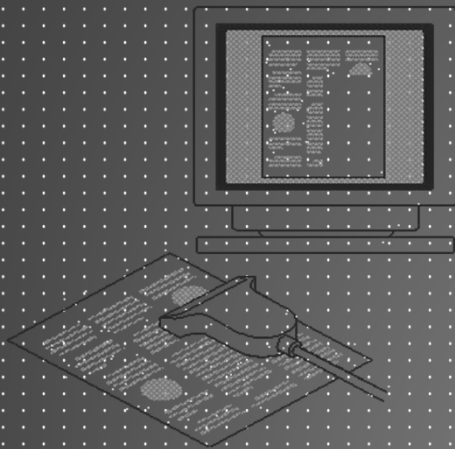


# As Imagens digitais são definidas por um conjunto de parâmetros

- A profundidade (*bit depth*), (tons), (cores)
- Dimensões, (altura) e (largura)
- Modo de cor, (CMYK), (RGB)
- Resolução, (dpi)
- Compressão, (redução de imagem)
- Formato, (tiff<sup>1</sup>), (pdf), (jpeg). A geração de formatos pelos escâneres deverá ser sempre em tiff, embora os demais formatos – JPEG 2000 e PNG – podem ser utilizados livremente para a geração dos vários tipos de imagens visado principalmente a internet



# DIGITALIZAÇÃO



“Conversão de um dado qualquer (gráfico, texto, palavras, objetos . . .) para o formato digital”(Bits)



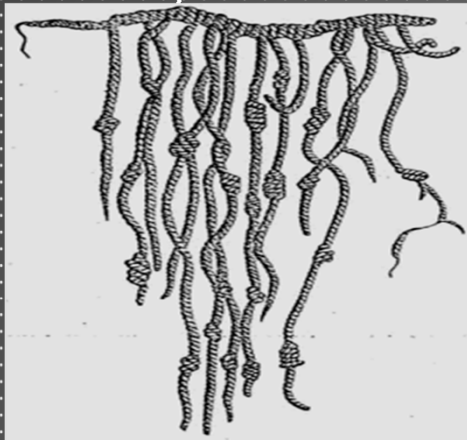


# PRIMÓRDIOS DIGITALIZAÇÃO

O QUE CHAMAMOS DE PASSADO É CONSTRUÍDO SOBRE BITS.

—JOHN ARCHIBALD WHEELER

Khipu Inca



Teletipo



CÓDIGO MORSE						U	..-
A	·-	F	··-	K	-·-	V	··-
B	-··	G	-·-	L	·-·	W	·--
C	-·-	H	···	M	--	X	-·-
D	-·	I	··	N	-·	S	··
E	·	J	·- - -	O	- - - -	T	-
						Z	- - ·

Sinais de Fumaça



Friktores Grego



Tambor



	Υ	Υ	Υ		
Υ	E	O	Θ	Z	
Υ	A	B	Δ	Γ	Σ
Υ	I	N	M	Φ	Λ
	H	T	P	K	Ω
	Υ	Π	Ξ	X	Ψ



# DIGITALIZAÇÃO GRÁFICA





# DIGITALIZAÇÃO GRÁFICA

## Alguns software de edição

Limpeza, recuperação e gerencia:

- PaperPort

OCR:

- OminiPage
- Readiris
- FineReader
- Acrobat

Edição de imagens:

- Corel Draw
- Photo Shop
- Print Shop Pro e outros





# SENSORES

Você já viu algum sensor?

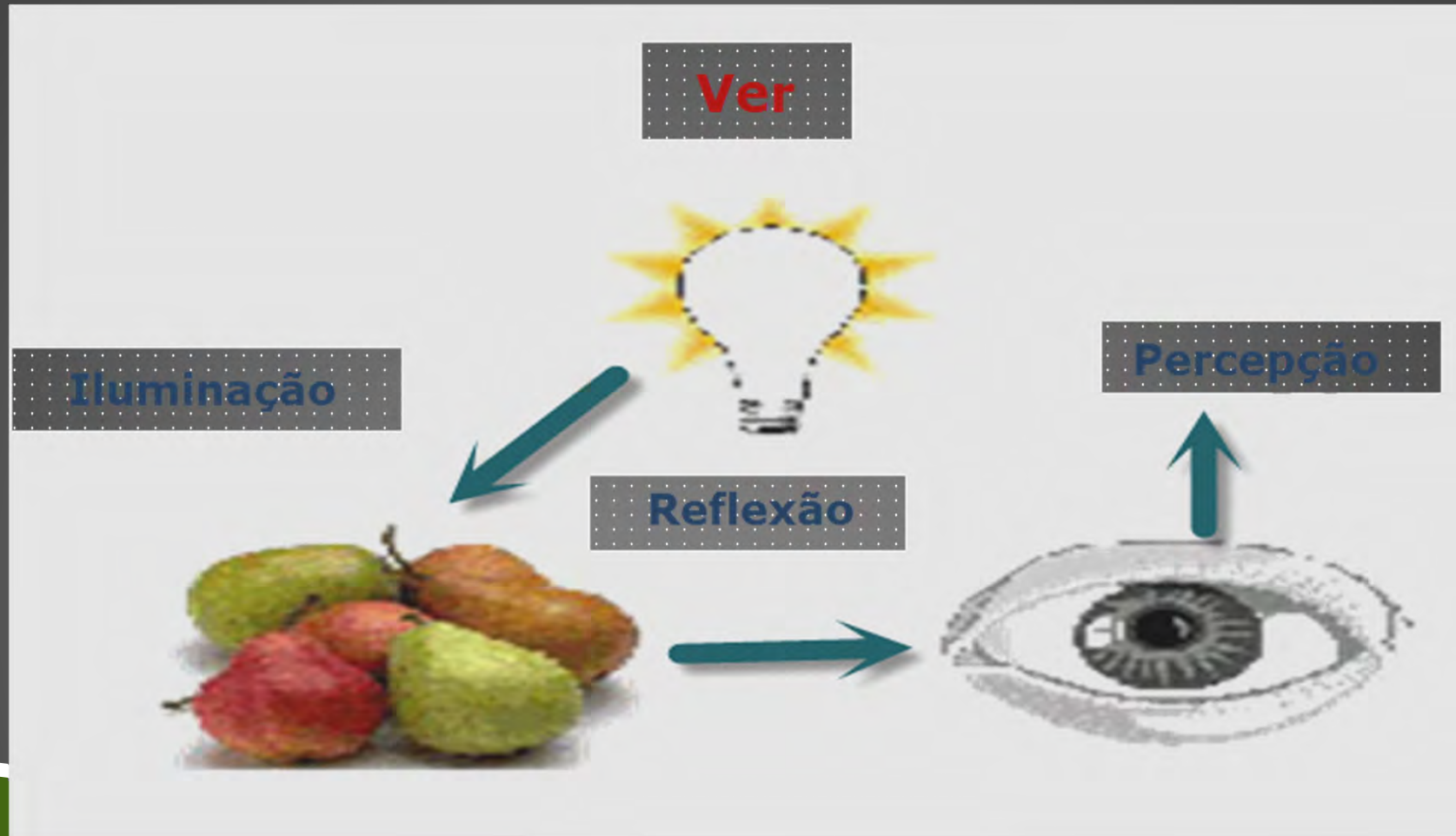
O que seria um sensor ?

Qual sensor que se mais usa no final do mês?





# SENSORES



# SENSORES



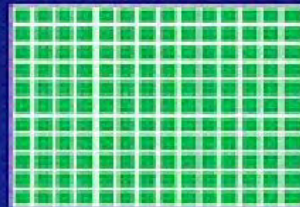
Captura



Motivo

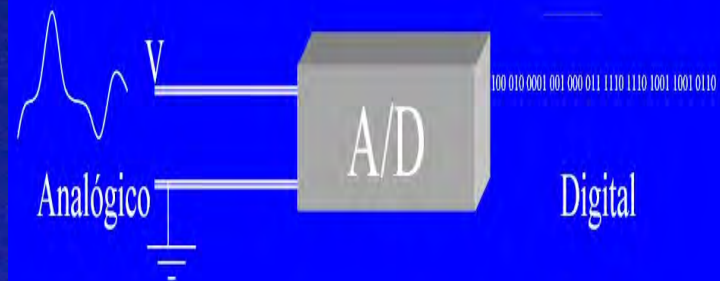


Lente



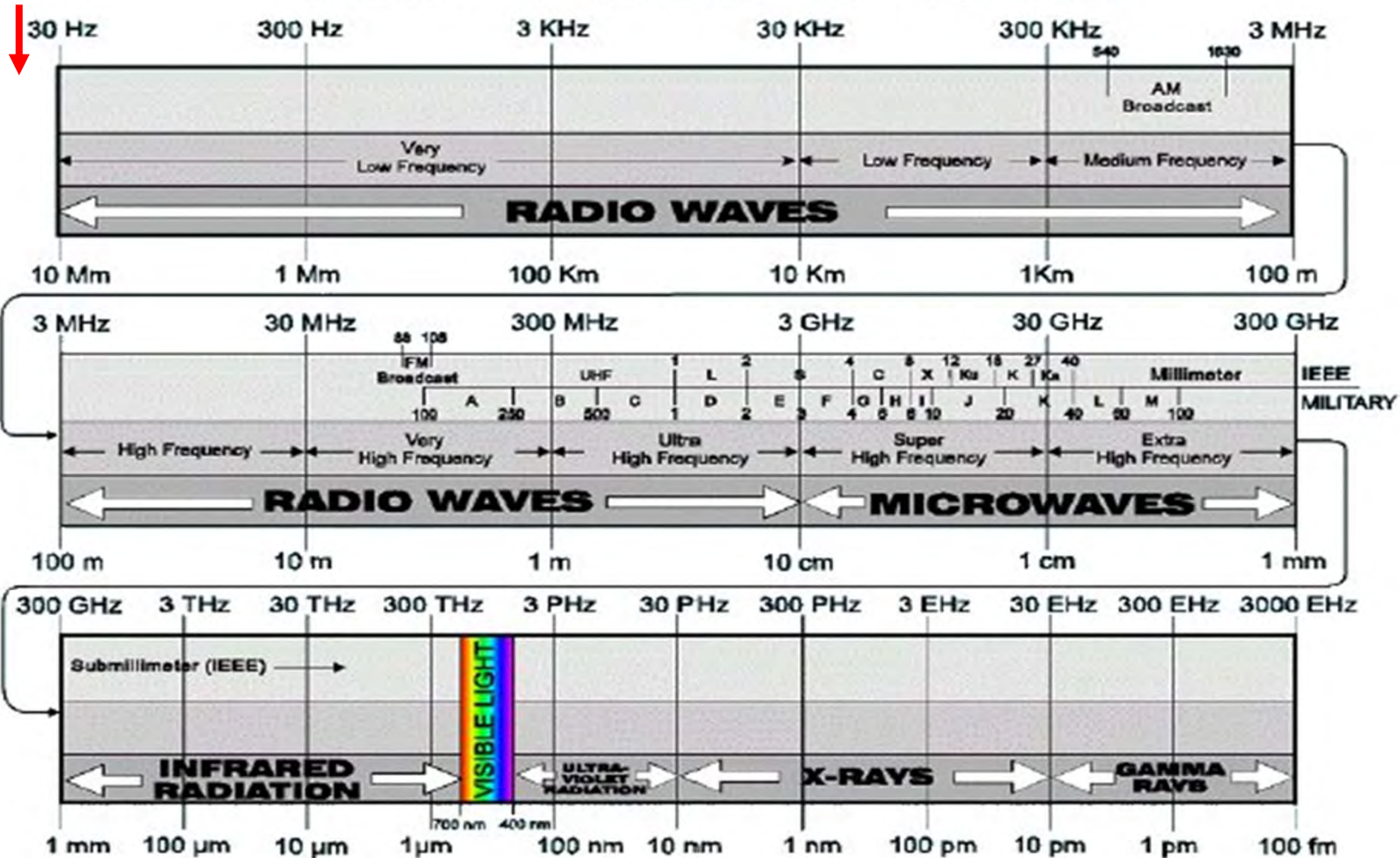
Sensor

## Conversor A/D



# Espectro da Radio Frequência

Audição



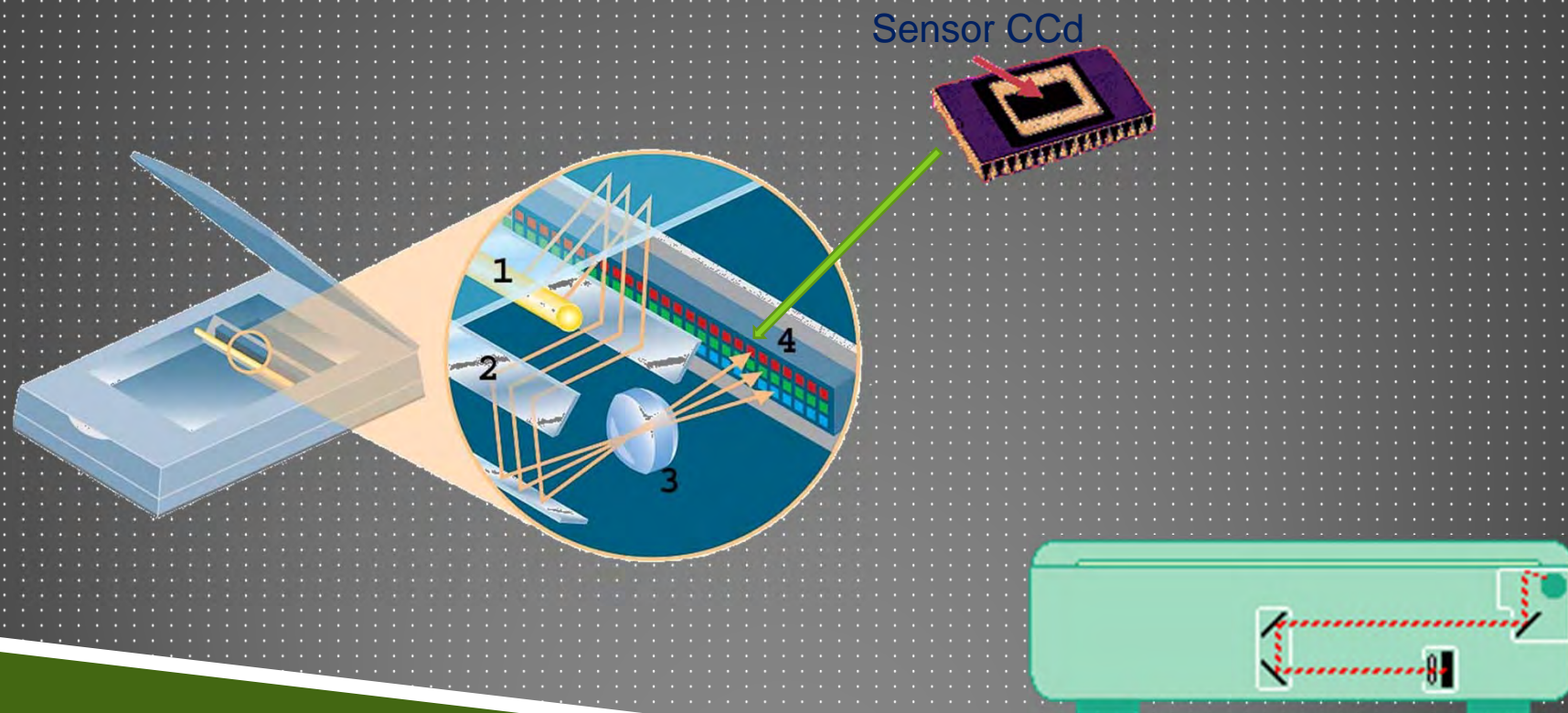


# TIPOS SENSORES PARA SCANNER

- CCD (Charge Coupled Device)
- CIS (Contact Image Sensor)
- PMT (Photomultiplier Tube )
- CMOS (Complementary Metal Oxide Semiconductor )
- PSD(Position Sensitive Diodes)

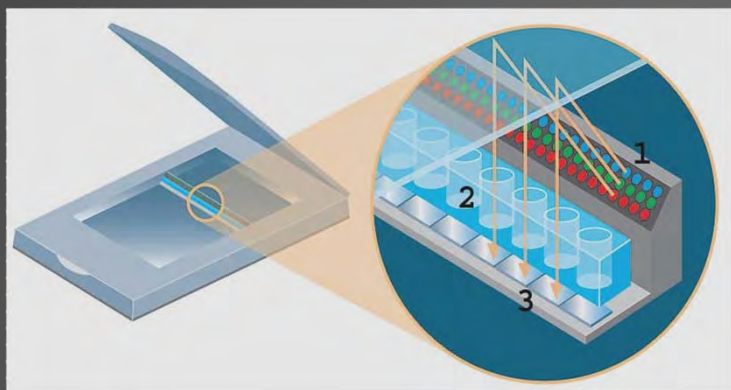


# CCD (CHARGE COUPLED DEVICE) DISPOSITIVO DE CARGA EMPARELHADA

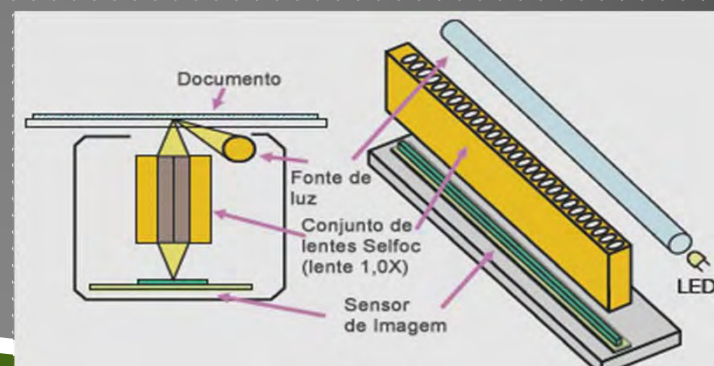
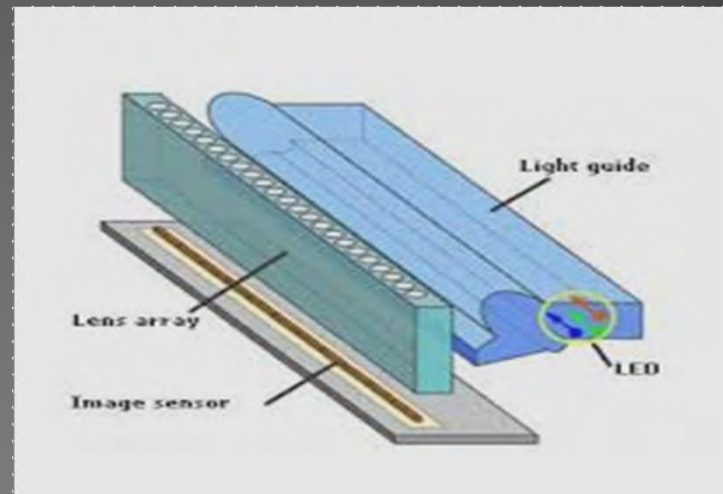




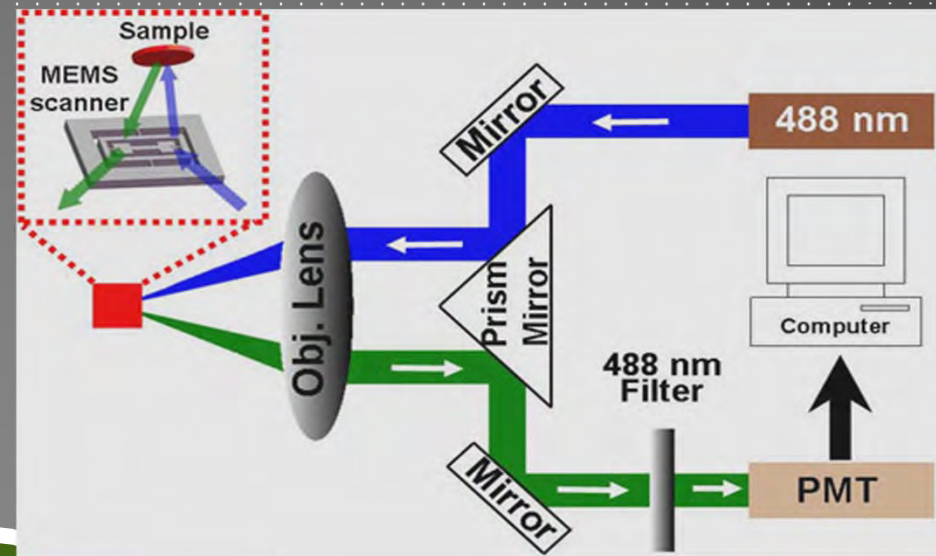
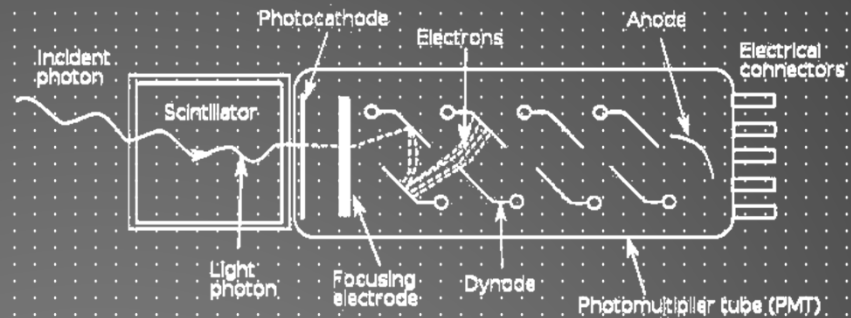
# CIS (CONTACT IMAGE SENSOR) SENSOR DE CONTATO DE IMAGEM



Sensor CIS

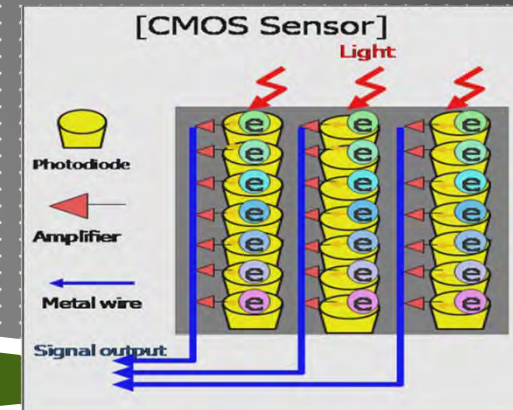
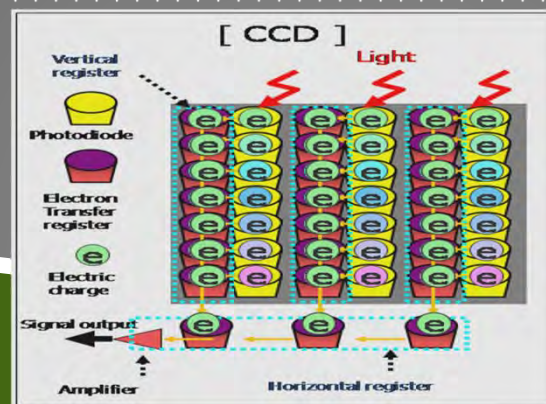
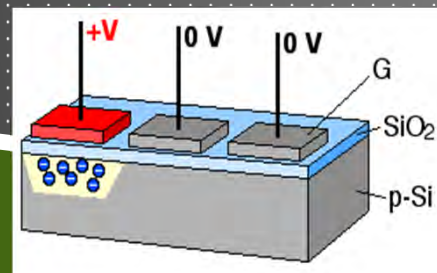
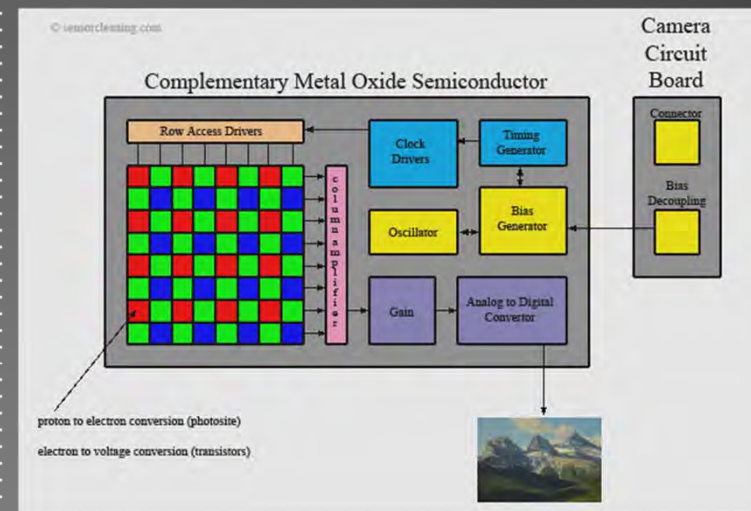
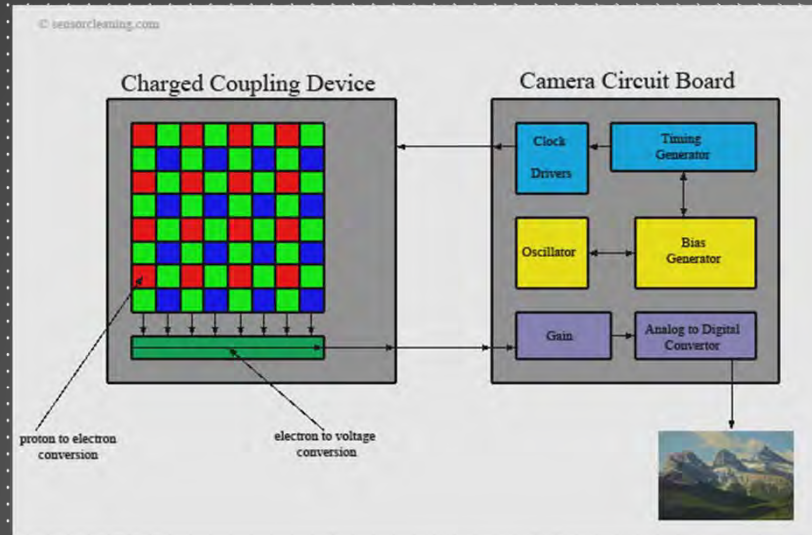


# PMT (PHOTOMULTIPLIER TUBE) TUBO OTOMULTPLICADOR



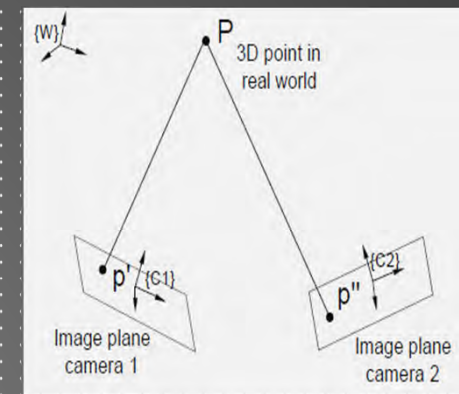
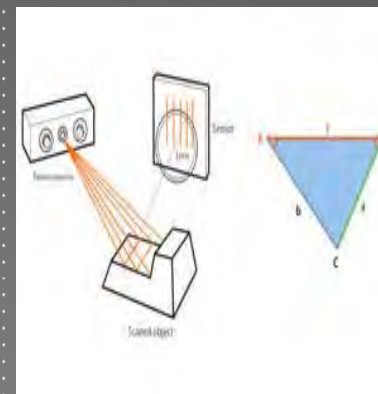
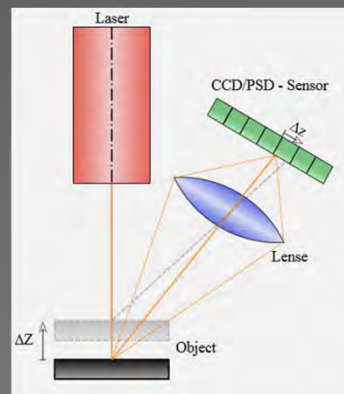
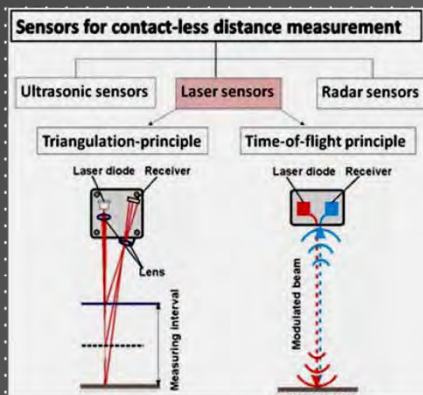


# CCD vs CMOS

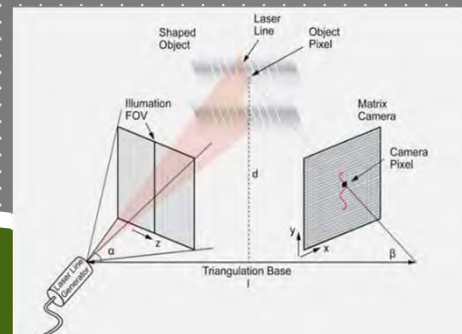
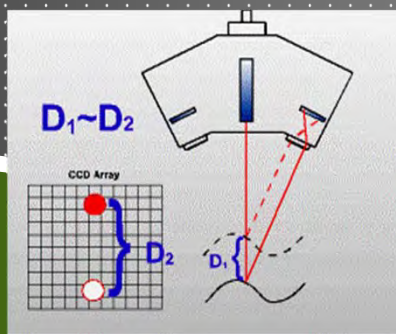


# PSD vs CCD

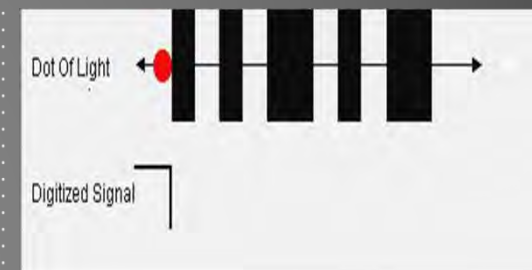
## Operação a Laser



## Triangulação em 3D

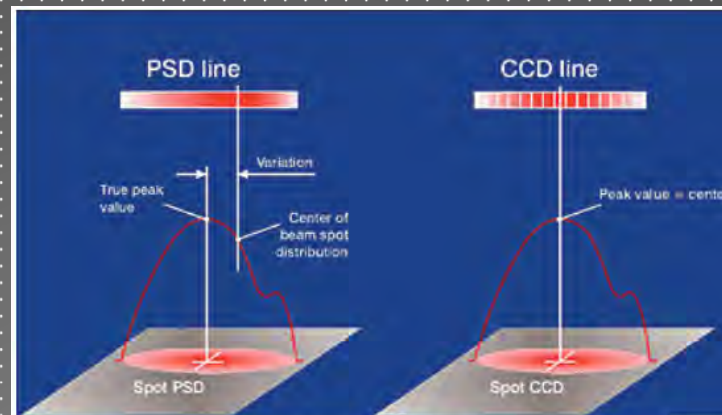


## Leitura do código de Barra





# PSD vs CCD



Esta ilustração mostra a distribuição de intensidade de luz e detecção de posição com um elemento CCD e um elemento PSD analógico. As alterações de cores e reflexos variáveis resultam em erros de detecção de posição quando um elemento PSD é usado. Por outro lado, o elemento CCD lhe dá a posição exata, independentemente da cor de destino, inclinação, ou reflexos variáveis.

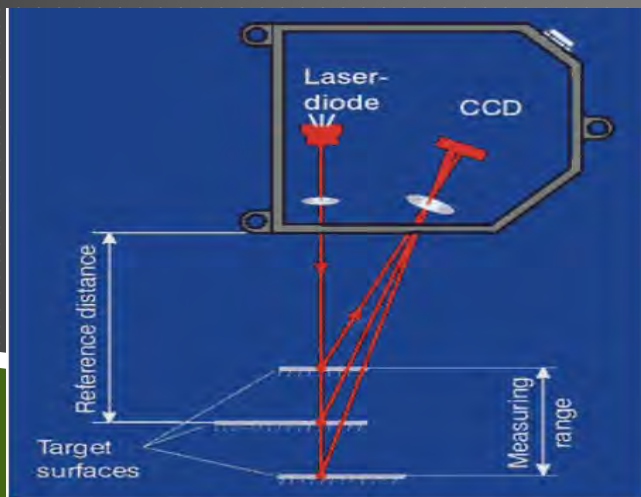
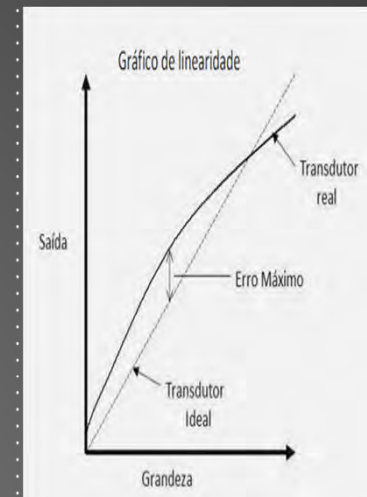
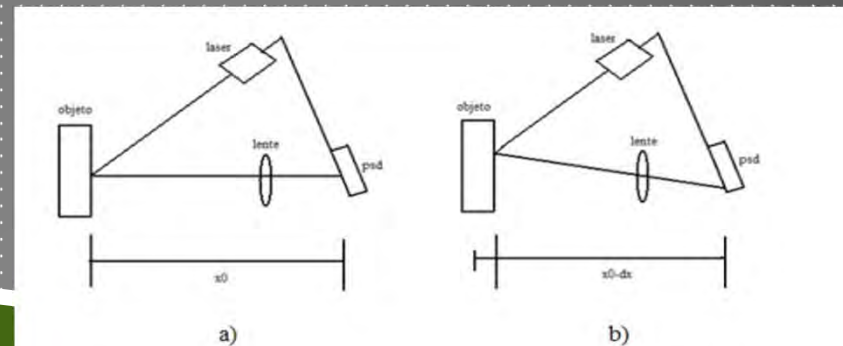


Figura 1. Agora você pode realizar medição de deslocamento sem contato com sensores de triangulação. Com um sensor de laser óptico (classe 2), estes sensores podem ler a uma distância de referência qualquer posição alvo mudar para baixo para microns. Os sensores estão disponíveis, que podem operar a varia de 2 a 700 mm.



Medida de distância por triangulação: a) spot resultante de um objeto no ponto  $x_0$ ; b) spot resultante do objeto no ponto  $x_0-dx$



# TIPOS DE ESCÂNERS

Portátil de Mão



Portátil Código de Barra



(Flatbed) de Mesa



PMT (Photomultiplier Tube) Tubo fotomultiplicador



Documento



Fotografia, Micro Filmes...Etc.

Escâner Duplo





# TIPOS DE ESCÂNERS



Herbário



Maquina  
fotográfica



Planetário de Livros



Laser de Mão



Laser de Mesa.



Laser de Campo



# INTERFACE SCANNER

- ▶ As velocidades de transmissões de dados:

- Serial: 115 Kbps
- Paralela: 2 Mbps
- USB 1.1: 1.5 a 12 Mbps
- USB 2: 480 Mbps
- USB 3: 4,8 Gbps
- SCSI: 5 a 160 Mbps
- FireWire: 100 a 400 Mbps





# CORES

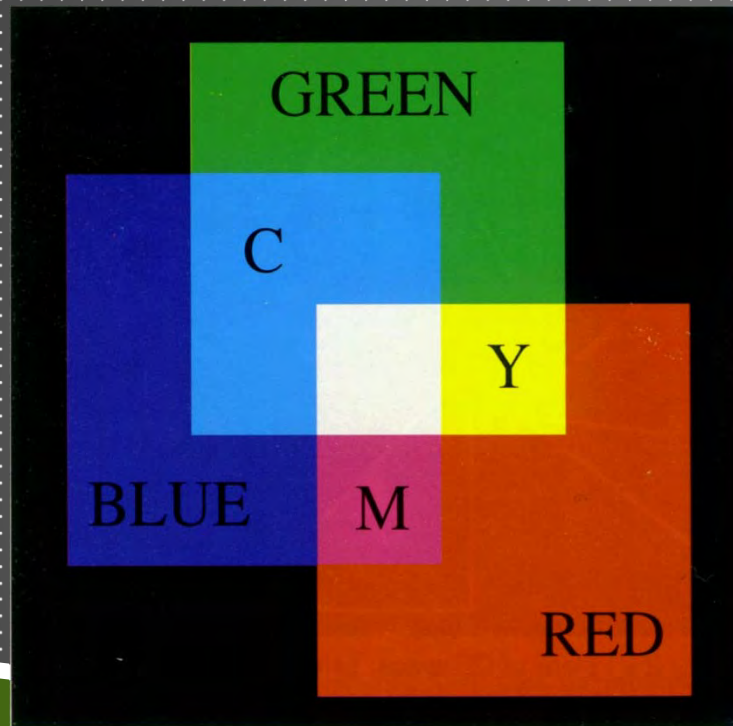
Que são Cores Aditivas ou RGB ?

Que são Cores Subtrativas ou CMYK ?



# CORES

Cores Aditivas – RGB



Cores Subtrativas - CMYK



# CORES

## Profundidade:

- 1 bit ( $2^1$ ) = 2 níveis
- 4 bits ( $2^3$ ) = 16 níveis
- 8 bits ( $2^8$ ) = 256 níveis x 3 cores = 24 bits
- 12 bits ( $2^{12}$ ) = 4096 níveis x 3 cores = (36 bits)
- 16 bits ( $2^{16}$ ) = 65536 níveis x 3 cores = 48 bits
- 24 bits ( $2^{24}$ ) = 16,7 milhões níveis x 3 cores = 72 bits



# ESTRUTURA FÍSICA E OPERACIONAL



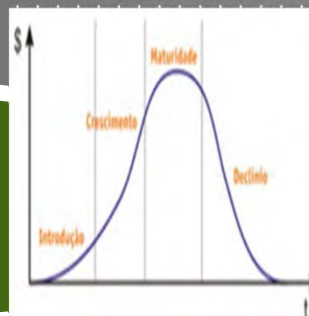


# ESTRUTURA FÍSICA E OPERACIONAL

## ► Armazenamento e cópias de segurança (backups):

- Não há uma solução centralizada universal e consistente, que garanta o armazenamento a longo prazo de um documento digital;
- Os avanços tecnológicos e as políticas em um ciclo vital muito curto e custo alto;
- Formatos: revistos com frequência e investimentos contínuos;

Custo Auto



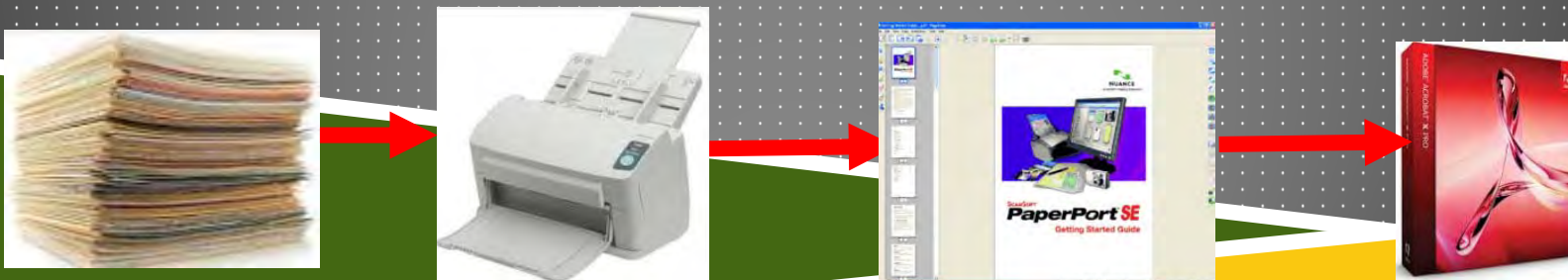
Avaliação contínua



# ESTRUTURA FÍSICA E OPERACIONAL

## ► Organização:

- Os equipamentos, os computadores, as mesas e as prateleiras, os equipamentos de calibração dos scanners deverão estar em seus locais apropriados.
- Os materiais que serão digitalizados deverão estar organizados na sequência cronológica.
- Uma Ficha de Controle de Digitalização e seu processo de elaboração, devem ser elaboradas com detalhamento dos fluxos e dos processos de cada projeto.



# ESTRUTURA FÍSICA E OPERACIONAL

- ▶ Planejamento visando a sustentabilidade dos formatos digitais:
  - formato TIFF para imagens de alta qualidade e impressão profissional principalmente no modo CMYK, pode ser arquivado sem qualquer compressão.
  - Atualmente, está se estudando uma versão do formato JPEG 2000, é arquivado sem qualquer compressão.
  - O formato PDF/A, subtipo do PDF, busca incrementar a sua longevidade arquivista. Foram removidas algumas capacidades do formato PDF, como a inclusão de áudio, vídeo, Java script e referenciamento de fontes.



# ESTRUTURA FÍSICA E OPERACIONAL

Planejamento visando a sustentabilidade dos formatos digitais:

U.S. National Archives and Records Administration (NARA)	PDF/A	Aceita
European Commission (MoReq)	PDF/A	Recomenda
German government (SAGA v4.)	PDF/A	Recomenda
French government	PDF/A	Recomenda
Dutch government	PDF/A	Mandatário
National Archives of Sweden	PDF/A	Aceita
Austrian National Library	PDF/A	Recomenda
National Archives of Norway	PDF/A	Recomenda
Organization for the Promotion of Automated Accounting	PDF/A	Recomenda
Brazilian Federal Legislature	PDF/A	Mandatário







## Bem-vindo aos Guias Práticos do PaperPort 11

O que deseja fazer?



Digitalizar documentos



Editar e aperfeiçoar imagens



Organizar arquivos



Montar documentos



Procurar itens



Adicionar notas



Processar formulários

Não exibir novamente os Guias Práticos.

Usar o menu Ajuda no modo de exibição de área de trabalho para restaurar os Guias Práticos.



The screenshot shows a software interface with a menu bar (Arquivo, Editar, Exibir, Item, Ferramentas, Ajuda) and a toolbar. The interface is divided into several panels:

- Pastas:** A file explorer showing a tree structure of folders, including 'Documents', 'Anti-Malware', 'Fax', 'Meus documentos do PaperPort', and various sub-folders like 'Amostras', 'Fotografias', 'Apresentações', etc.
- Digitalizar ou obter foto:** A central panel with a 'Scanner:' section and a 'Selecionar...' button. Below it, a 'Selecione um perfil:' section lists options: 'Documento em PB', 'Documento em tons de cinza', 'Documento em cores', and 'Fotografia colorida'. At the bottom, there are buttons for 'Novo...', 'Excluir...', and 'Configurações...', and a 'Digitalizar' button.
- Desempilhar a página atual / Desempilhar todas as páginas:** A panel showing a preview of a scanned document page with a red box highlighting a specific area.
- Confirmar várias páginas Endireitamento a...:** A dialog box with a question mark icon, asking 'Deseja modificar todas as páginas?' and providing 'Sim' and 'Não' buttons. A checkbox at the bottom reads 'Não exibir esta mensagem novamente.'

Red arrows indicate the workflow: one arrow points from the 'Arquivo' menu to the 'Pastas' panel; another points from the 'Selecionar...' button to the 'Digitalizar ou obter foto' panel; a third points from the 'Selecionar...' button to the 'Desempilhar a página atual' panel; and a fourth points from the 'Digitalizar' button to the 'Confirmar várias páginas' dialog box.





Conversão de arquivo - OCR0002.txt

Escolha a codificação que faz com que o seu documento fique legível.

Codificação de texto:

Windows (padrão)  MS-DOS  Outra codificação:

- Croata (Mac)
- Europeu Ocidental (DOS)
- Europeu Ocidental (IA5)
- Europeu Ocidental (ISO)
- Europeu Ocidental (Mac)
- Europeu Ocidental (Windows)

Visualização:

Conserve O Gram  
July 1993 Number 18/1  
NATIONAL PARK SER VICE  
4w~  
OEPAIUMENTAIQF  
HF INTERIOR  
Polyester Fibn Book Supports  
Open books ideally should not be exhibited on a long term or permanent basis because of the cumulative damage to them caused by light and the physical distortion that is likely to occur. The hook support described here can be constructed by park staff for temporary display at a reasonable cost.

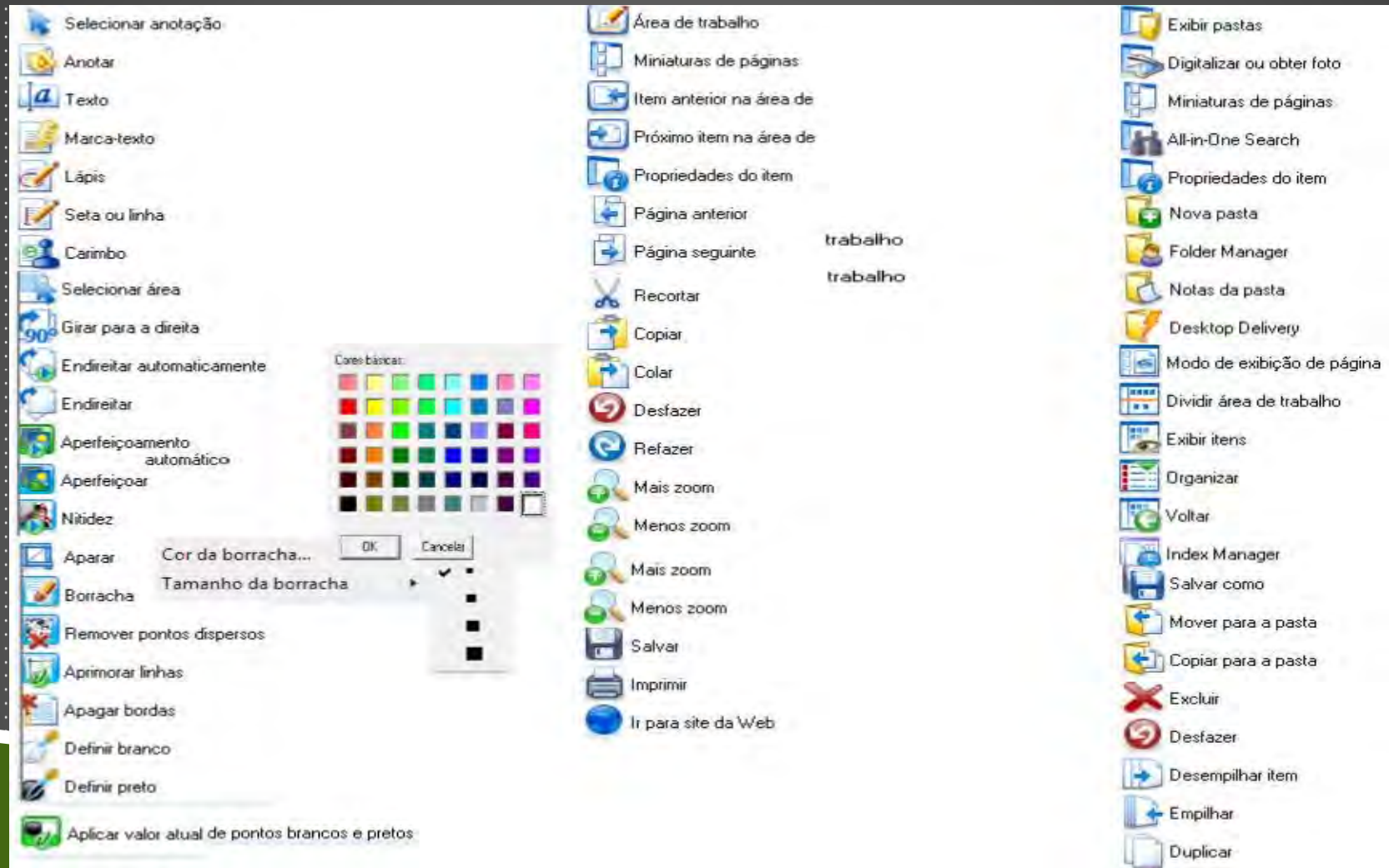
OK Cancelar

Enviar para  
Enviando item para winword...





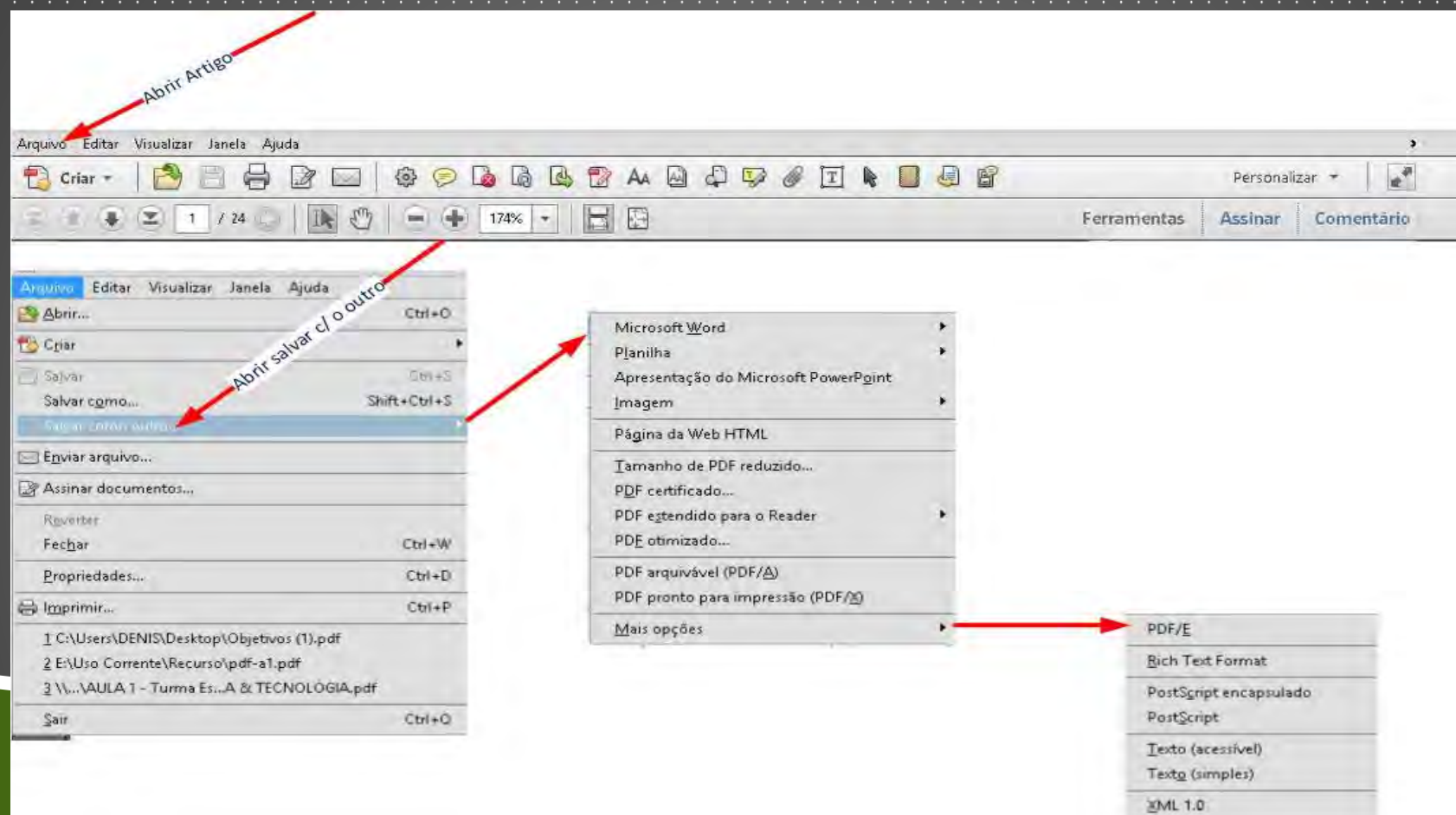
# COMANDOS do PAPERPORT





# ACROBAT

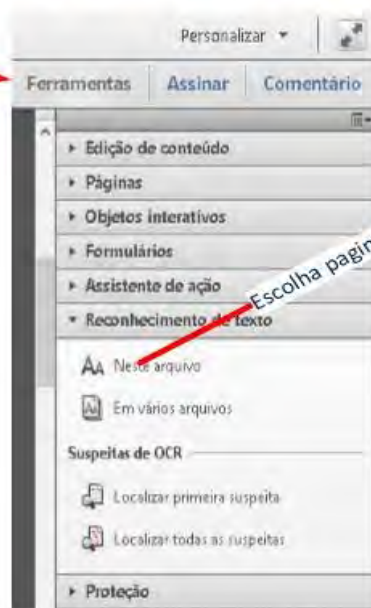
Salvar em vários formatos PDF:



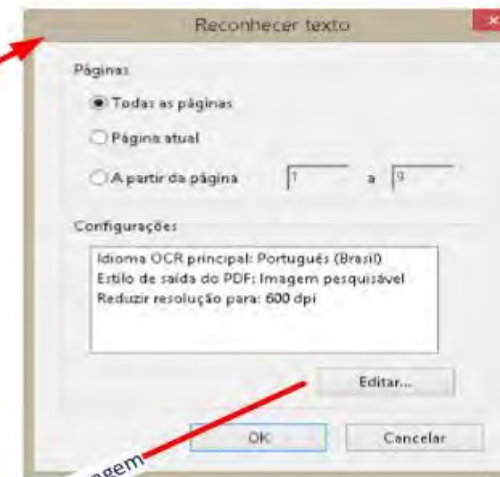
# ACROBAT

## Como executar o OCR no Acrobat:

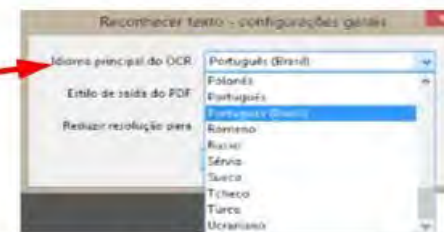
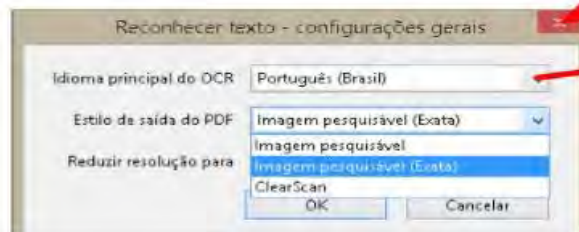
Abra Ferramentas



Escolha páginas e em seguida editar



Escolha o idioma e tipo de imagem



Dê um OK nos finais



# REFERÊNCIAS:

A FRAMEWORK of guidance for building good digital collections, 3 rd edition December 2007.

<http://www.niso.org/publications/rp/framework3.pdf>. PDF Association: <http://www.pdfa.org/>

ANDERSSON, Henrik. Position Sensitive Detectors - *Device Technology and Applications in Spectroscopy*. Thesis for the degree of Doctor of Technology. Sundsvall 2008. <http://miun.diva-portal.org/smash/get/diva2:1939/FULLTEXT01.pdf>. Acessado em: 17/02/2014

CONSELHO NACIONAL DE ARQUIVOS - CONARQ *Recomendações para Digitalização de Documentos Arquivísticos Permanentes-2010*. [http://www.conarq.arquivonacional.gov.br/media/publicacoes/recomenda/recomendaes\\_para\\_digitalizao.pdf](http://www.conarq.arquivonacional.gov.br/media/publicacoes/recomenda/recomendaes_para_digitalizao.pdf). Acessado em: agosto/2013

FRAGA, Marcelo Noce. *Padrões de qualidade para o serviço de digitalização, baseados em normas e procedimentos adotados*. Instituto Brasileiro de Informação em Ciência e Tecnologia. Dezembro 2013

HOU, Bo; Wu, Z. Y.; de la Tognaye, J. L.; de Bougrenet, P.; Grosso, P. *Charge-coupled devices combined with centroid algorithm for laser beam deviation measurements compared to position sensitive device*. <http://publib.upol.cz/~obd/fulltext/Physica-40-41/acta7.pdf>

IDEIAS de Negócio- digitalização de documentos-Sebrae <http://www.sebrae.com.br/setor/servicos/acesse/ideias-de-negocio/ideias-de-negocio/visualizar-ideias/documento/AA077030A89C4A8183257998004CE163/campb/impNeg>

INTERNATIONAL COLOUR CONSORTIUM. Specification ICC.1:2010 – *Image Technology Colour*

KENNEY, A., e Chapman, S. *Requisitos de resolução digital para textos: Métodos para o estabelecimento de critérios de qualidade de imagem*. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico no 51). <http://siarq49.siarq.unicamp.br/cpba>. Acessado em: 08/ 2013

LIBRARY OF CONGRESS. *Sustainability of Digital Formats Planning for the Library of Congress*. Washington, DC., 2013. <http://www.digitalpreservation.gov/formats>. Acessado em: 08/2013)





# REFERÊNCIAS:

*Making the case for PDF/A and Adobe® Acrobat.*

<http://web.archive.org/web/20131210165012/http://www.adobe.com/enterprise/pdfs/pdfaforAcrobat.pdf>... Acessado: 15/02/2014.

Acessado em: 17/02/2014

*Management – Architecture, Profile Format, and Data Structure* [http://www.color.org/specification/ICC1v43\\_2010-12.pdf](http://www.color.org/specification/ICC1v43_2010-12.pdf) Acessado em: 08/2013

MOVING theory into practice digital imaging tutorial - [http://www.library.cornell.edu/preservation/tutorial/tutorial\\_English.pdf](http://www.library.cornell.edu/preservation/tutorial/tutorial_English.pdf)

*National Information Standards Organization. Understanding Metadata.*

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf> Acessado: 17/02/2014

OGDEN, S. (Ed). "Cleaning Books and Shelves", *Preservation of Library and Archival Materials: a Manual*. Andover, MA: Northeast Document Conservation Center, 1999.

OSBORNE, Rebecca and Scott, Catherine; IBM; Griggs, Karen; Rhodes, Erin and Puglia Steven. Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files.

[http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf) Acessado em: 27/10/2013

PARDUE, Diana R., "Volcanic Ash: Cleaning Museum Objects", *Conserve O Gram*, 3/5, Washington, DC: National Park Service, 1993.

QUEENSLAND STATE ARCHIVES: *Digitisation Disposal Policy Toolkit*. The State of Queensland Department of Public Works), 2010.

RIVERA, Rogelio. Vargas, Gustavo. Vázquez, Marco. **IBM® System Storage® LTO Ultrium 6 Tape. Drive Tape DRIVE**

**PERFORMANCE WHITE PAPER**. [http://www-01.ibm.com/common/ssi/cgi-](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=SP&infotype=PM&appname=STGE_TS_TO_USEN&htmlfid=TSD03160USEN&attachment=TSD03160USEN.PDF)

[bin/ssialias?subtype=SP&infotype=PM&appname=STGE\\_TS\\_TO\\_USEN&htmlfid=TSD03160USEN&attachment=TSD03160USEN.PDF](http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=SP&infotype=PM&appname=STGE_TS_TO_USEN&htmlfid=TSD03160USEN&attachment=TSD03160USEN.PDF) Acessado em: 02/03/2014





# REFERÊNCIAS:

SCHWARZ, Rosângela Rodrigues; Taga, Vitor; Falcão, Eleonora Milano *Vieira-Repositórios Educacionais: estudos preliminares para a Universidade Aberta do Brasil*. [http://www.scielo.br/scielo.php?script=sci\\_serial&pid=1413-9936&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_serial&pid=1413-9936&lng=en&nrm=iso) Acessado em: 27/02/2014

SERGIYENKO, Oleg. Balbuena, Daniel. Hernández Rivas, Moisés. Flores, Wendy. Rivera, Javier and Bueno, Alejandro Sánchez- *A Method and Electronic Device to Detect the Optoelectronic Scanning Signal Energy Centre*- <http://www.intechopen.com/download/get/type/pdfs/id/40008> \_Acessado em: 17/02/2014

SILVA, Rubens Ribeiro Gonçalves da. *Manual de digitalização de acervos : textos, mapas e imagens fixas* - Salvador : UFBA, 2005. 56p <https://repositorio.ufba.br/ri/bitstream/ufba/141/1/Manual%20de%20digitalizacao%20de%20acervos.pdf> . Acessado em: 09 /2013.

SITTS, Maxine. (ed). *Handbook for Digital Projects: A Management Tool for Preservation and Access*. <http://www.nedcc.org/assets/media/documents/dman.pdf> (Acessado em Agosto de 2013) University of Colorado Digital Library. Digitization Best Practices, Version 1.0. Denver, CO, EUA, 2009.

STALLINGS William, *Supplement to Computer Organization and Architecture, Eighth Edition*- <http://williamstallings.com/COA/COA8e.html> Acessado em: 27/02/2014

STALLINGS, William. *Computer Organization and Architecture, Eighth Edition*

*TECHNICAL guidelines for digitizing archival materials for electronic access: creation of production master files, U.S. National Archives and Records Administration(NARA) June 2004 -Guia para digitalização de documentos Versão 2.0 –2006*. <http://www.archives.gov/preservation/technical/guidelines.html>

THE LIBRARY: *of Congress Technical Standards for Digital Conversion Of Text and Graphic Materials*. <http://memory.loc.gov/ammem/about/techStandards.pdf> . Acessado em: 16/02/2014

WATERS, Donald J. *Do microfilme à imagem digital: Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos; Arquivo Nacional, 2001. (Caderno Técnico no 49)*. <http://siarg49.siarq.unicamp.br/cpba> Acessado em: 08/2013

