



INTER
FACES
CIENTÍFICAS

EXATAS E TECNOLÓGICAS

ISSN IMPRESSO - 2359-4934

ISSN ELETRÔNICO - 2359-4942

<http://dx.doi.org/10.17564/2359-4942.2018v3n2>

AQUISIÇÃO DE IMAGEM PARA HEMEROTECA DIGITAL

IMAGE ACQUISITION FOR DIGITAL LIBRARY

ADQUISICIÓN DE IMÁGENES PARA HEMEROTECA VIRTUAL

Efraim Santana Leite Filho¹
Thiciane Suely Couto Silva³

Victor Brito Villar²
Fábio Gomes Rocha⁴

RESUMO

Este trabalho teve o objetivo de verificar quais dispositivos tinham melhor custo-benefício para a digitalização de documentos e apontar qual o melhor formato de arquivo de imagem a fim de armazená-los em um sistema de hemeroteca virtual. A partir das amostras coletadas num jornal, com diferentes fontes, qualidade de impressão e deterioramento; uma leitura OCR foi realizada com o objetivo de avaliar a precisão do reconhecimento das palavras no texto. Em seguida, uma amostra foi convertida em diferentes formatos,

ponderando-se a qualidade final da imagem e o espaço ocupado. Concluiu-se que o dispositivo LG G4 apresentou o melhor custo-benefício e o formato de imagem tido como o mais adequado foi o JPEG.

PALAVRAS-CHAVE

Digitalização. Hemeroteca Digital. Câmera. Smartphones. Formato de Imagem.

ABSTRACT

This paper intended to define which of the available devices was considered to be the most cost-effective one to scan documents and point out which of the image file formats was the most adequate one for storing these files on a virtual library. Samples collected from a newspaper, which expressed different types of fonts, print quality and deterioration, were put through an OCR reading for testing the accuracy of the recognized words. Following that, one of the samples had been converted to different image file formats and

evaluated the overall image quality and storage space occupied. As a conclusion, the LG G4 was considered the most cost-effective device and JPEG was selected as the most adequate file format for this application.

KEYWORDS

Scanning. Virtual Library. Camera. Smartphones. Image file format.

RESUMEN

Este trabajo tuvo el objetivo de comprobar qué dispositivos tenían mejor costo-beneficio para escanear documentos y elegir cuál es el mejor formato de archivo de imagen para almacenarlos en un sistema de hemeroteca virtual. A partir de las muestras recogidas en un periódico, con diferentes fuentes, calidad de impresión y deterioro; una lectura OCR fue realizada con el objetivo de evaluar la precisión del reconocimiento de las palabras en el texto. A continuación, una muestra se convirtió en diferentes formatos, ponderando la

calidad final de la imagen y el espacio ocupado. Se concluyó que el dispositivo LG G4 presentó el mejor costo-beneficio y el formato de imagen como el más adecuado fue el JPEG.

PALABRAS CLAVE

escaneado de documentos. hemeroteca digital. cámara. smartphones. formato de imagen.

1 INTRODUÇÃO

Na década de 1990, a indústria dos serviços on-line apresentou um tremendo crescimento, uma porcentagem de aproximadamente 200% ao ano (KEAVENEY; PARTHASARATHY, 2001). Instituições como a Universidade de Harvard, a Universidade de Cambridge e o Instituto de Massachusetts possuem seus acervos digitalizados e acessíveis a usuários do mundo inteiro. Tentativas como o Projeto Gutenberg também contribuíram com o acesso à literatura, esse disponibilizando em sua base de dados livros cujos direitos autorais tenham expirado ou foram concedidos pelos autores (LEVACOV, 1997).

No Brasil, esta demanda ocorreu também em órgãos públicos e empresas sem fins lucrativos, assim que a Lei nº 12.527/2011, posta em vigor em 16 de maio de 2012 (conhecida como lei do acesso à informação) os obrigou a disponibilizar a qualquer pessoa, física e jurídica, o acesso a informações públicas.

Esta pesquisa, considerando os aspectos aqui citados, se faz importante pela tentativa de encontrar uma ferramenta que apresente um bom custo-benefício para as digitalizações de documentos e registros, bem como indicar qual o formato de arquivo de imagem mais adequado para armazenamento que apresente uma legibilidade satisfatória. Este trabalho se divide em seções de Teoria de Aquisição de Imagens, capítulo que expõe a fundamentação teórica do artigo, Estudo de Caso, que lida com a metodologia executada e os dispositivos testados, Resultados, que trabalha com critérios objetivos de avaliação dos dados colhidos e Conclusão.

2 ESTUDO TEÓRICO: TEORIA DE AQUISIÇÃO DE IMAGENS DIGITAIS

As imagens digitais são representadas digitalmente por uma matriz finita S de tamanho m por n (onde m é quantidade de linhas e n é a quantidade de colunas) (FRERY; PERCIANO, 2013). Os elementos desta

matriz são chamados de coordenadas ou posições e seus valores são os dados responsáveis por formar a imagem digitalmente.

Os dados da imagem se utilizam de modelos que definem o formato dos valores representativos da imagem, sendo o mais comum para dispositivos eletrônicos como TVs, monitores de computadores, scanners e câmeras digitais o modelo RGB (FRERY; PERCIANO, 2013). O RGB Model (ou modelo vermelho, verde e azul) representa os detalhes em cada pixel, armazenando as diferentes intensidades das cores vermelho, verde e azul, onde estas variam num intervalo de 0 (zero) a 255, caso utilizem o padrão de 8 bits. Uma vez abordada a base para uma imagem digital, o próximo passo é a busca pelo formato que concilie sua reprodução virtual fiel aos detalhes com o mais compacto tipo de armazenamento.

A criação de uma imagem a partir de um scanner se dá por digitalização de uma informação presente em um documento ou foto. O processo de digitalizar uma informação parte do princípio de se definir alguns parâmetros do objeto a ser digitalizado, como paleta de cores (colorido, escala de cinzas ou preto e branco), o tipo de objeto (uma foto ou um documento) e a resolução do “scan” (BURGUER; HANSON; VAN OS, 2002). Estes parâmetros podem ser definidos por meio de um software ou no próprio aparelho scanner. O resultado é uma imagem mapeada bit a bit, cujas informações foram capturadas pelas variações de cores captadas no sensor. Seu formato pode ser alguns dos conhecidos formatos de imagem (abordados posteriormente neste trabalho) ou um formato de documento como o formato PDF.

Já a formação de uma imagem numa câmera digital se dá pela transformação do que é capturado pelo sensor e lente da câmera em sinais de imagem eletrônicos de forma repetida; isto acontece quando o obturador é acionado. Até que o obturador seja fechado, os sinais eletrônicos são processados em dados de um arquivo de imagem pelo conversor analógico-digital e ao concluir a captura de imagem, o arquivo pode ser armazenado e exibido em uma tela (KAZAMI *et al.*, 2003). Parâmetros como exposição,

foco, abertura de lente e até o resultado da imagem dependerão do hardware da câmera e do software/processador utilizados para processamento das informações capturadas.

2.1 FORMATOS DE IMAGENS

JPEG: O formato criado pelo **Joint Photographic Experts Group** e definido pela ISO 10918, é um dos mais usados para armazenar imagens em tamanhos reduzidos. A compressão do JPEG visa eliminar dados e detalhes da imagem redundantes para compactar o arquivo e tenta compensar isso com brilho dos pixels; a aplicação recorrente de compressão pode, inevitavelmente, gerar perda de detalhes.

De acordo Zimmerle e Pinheiro (2011), para o olho humano, mudanças individuais e sutis no brilho de cada pixel pouco alteram a percepção de qualidade da imagem, o que serve de grande artifício para a compressão e permitindo a eliminação de informações pouco importantes. Uma das poucas desvantagens é inflexibilidade quanto a manipulação recorrente da foto, já que a compressão acontece todas as vezes que se salva o arquivo, o que o torna mais adequado como formato final e não intermediário de imagem.

TIFF: A **Tagged Image File Format** é considerada uma família de formato binário de imagem, com 8 ou 16 bits para cada cor RGB (totalizando 24 ou 48 bits). A TIFF ou TIF permite a compressão de sua imagem com ou sem perda de dados, sendo que algumas câmeras fotográficas podem armazenar a foto diretamente em TIFF, usando o algoritmo de compressão de LZW (Lempel-Ziv-Welch), sem perda alguma. Alguns dos grandes benefícios do TIFF em relação a outros formatos é possibilidade de salvar várias imagens em um único arquivo e a facilidade de manipulação por ter uma compressão sem perdas de detalhes, sendo ele um dos mais utilizados não só em câmeras, mas também em editores de fotos de diversas plataformas. Contudo, seus arquivos ocupam grande espaço em memórias, não sendo tão eficiente como JPEG, na maioria das aplicações.

PNG: O **Portable Network Graphics** é um formato de imagem cujas informações são armazenadas como números inteiros em multi-bytes, com o byte mais importante primeiro. Esses bytes contêm as informações dos detalhes e podem ser classificados como fundamental ou não para a imagem, público ou privado (visto somente por programas específicos) e “Safe-to-copy” ou não. Esse tipo de formato tem como vantagens não possuir taxas de licença para seu uso e um método de compressão em que não ocorre perda de detalhes, sendo o PNG um formato interessante para aplicações em que uma imagem é editada e salva repetidas vezes. O contra-ponto do uso deste formato é o fato de cada arquivo armazenar somente uma imagem.

BMP: O formato BMP é o formato nativo dos sistemas operacionais Windows e as imagens suportadas por ele possuem entre 1 e 24 bits por pixels (quantidades de 16 e 32 bits são exceções). No BMP, o método de compressão “Run-Length” não gera perdas e pode ser usado para arquivos com 4 ou 8 bits por pixels. Considerando que as imagens salvam suas informações em binário, o autor do livro “Formatos de arquivos de Imagens Comprimidas”, o John Miano (1999), afirma que elas também precisam ser abertas em “modo binário” para serem manipuladas; caso contrário, o seu conteúdo pode ser visualizado de maneira inesperada ou incorreta. Este formato está entrando em desuso, principalmente em aplicações web, já que, apesar de não haver perdas na imagem durante a compressão, os arquivos produzidos são consideravelmente maiores em comparação a outros formatos.

3 ESTUDO DE CASO

Esta seção apresenta o estudo de caso, expondo o local, preparação e realização dos estudos, bem como os resultados obtidos, tendo como foco analisar as técnicas, procedimentos e equipamentos, com a finalidade de avaliação, em relação ao processo de digitalização de documentos antigos em português, do ponto de vista dos desenvolvedores, no contexto

da implementação da Hemeroteca Digital. As questões que norteiam esta pesquisa são:

Q1. Quais os equipamentos de melhor custo benefício disponíveis no mercado para produzir material digital de qualidade?

Q2. Qual melhor formato de imagem para armazenamento?

3.1 DESCRIÇÃO DO LOCAL

O projeto foi realizado por meio de um convênio entre a Universidade Tiradentes (UNIT) e o Serviços Gráficos de Sergipe (SEGRASE), ambos da cidade de Aracaju/SE/Brasil, sendo os responsáveis, o Grupo de Pesquisa em Tecnologia da Informação e Comunicação (GPITIC). A pesquisa foi desenvolvida por três alunos do curso de ciência da computação da Universidade Tiradentes, com idades de 21 e 22 anos, cursando o 6º e 8º período, sendo dois do gênero masculino e uma do gênero feminino, sendo orientados por um professor orientador. O objetivo principal do projeto é destacar um modelo acessível de fotografia de documentos antigos para o desenvolvimento de uma biblioteca digital de fácil acesso a dados históricos. A diretoria da SEGRASE é a responsável por avaliar o procedimento, contribuindo para a validação deste artigo.

3.2 MÉTODO E EXECUÇÃO

Para execução, foram utilizados exemplares de jornais do Diário Oficial do estado de Sergipe, os documentos datam de anos de 1978 até 2012, sendo este o objeto de estudo mais relevante, por se tratar de um documento deteriorado pelo tempo, tornando-o propício ao objetivo definido. Mais precisamente, para análise deste artigo, utilizamos uma amostra do dia 24 de abril de 1978 para os testes de fotografia. A técnica utilizada na amostra consistiu em uma foto por página, de 3 páginas diferentes, buscando capturar condições diferentes dos textos encontrados, variando fontes, qualidade da impressão e deterioramento. O equipamento fotográfico usado e suas respectivas configurações foram, conforme o Quadro 1.

Quadro 1 – Tabela de equipamentos

EQUIPAMENTO	LG G4	Apple iPhone 6 Plus	Nikon d5300
PIXELS (em MegaPixels)	16	8	24.2
RESOLUÇÃO DE IMAGEM	5312 x 2988	3264 x 2448	6016 x 4000
ABERTURA	F1.8	F2.2	f3.5-5.6
TAMANHO DO SENSOR	1 / 2.6"	1 / 3"	
TAMANHO DOS PIXELS	1.12	1.5	
FORMATO DE IMAGEM	JPEG	JPEG	JPEG e NEF

Fonte: Dados da pesquisa.

O uso de diferentes configurações de câmeras de smartphones disponíveis no mercado, comparado com uma câmera fotográfica DSLR, é justificado pelo objetivo de apresentar uma alternativa acessível para a execução do projeto, garantindo uma qualidade de material suficiente para a coleta dos dados. As condições mínimas para tal serão abordadas na conclusão deste artigo.

Para análise, buscou-se destacar um parágrafo por página, dentro das condições buscadas, usando amostras das imagens fotografadas por cada câmera. O parágrafo destacado foi então passado por um OCR – para este trabalho foi usado o Tesseract – e o critério de avaliação foi a porcentagem de reconhecimento de palavras e sua correspondência ao texto original.

Cada foto foi tirada, usando uma base inclinada apoiada sobre uma mesa, de forma que a fonte de luz

fosse posicionada bem acima da amostra. Assim, a sombra do fotógrafo não interfere na imagem e a luz consegue incidir em todo o documento, aumentando a qualidade de captura de imagem. Mais detalhes sobre técnica de captura de imagem na seção 2 deste artigo.

Figura 1 – Fotografia dos documentos



Fonte: Elaborado pelos autores

Vale ressaltar que, as fotos não foram tratadas por softwares de processamento de imagens, sendo avaliadas somente pela sua qualidade de captura. Sendo assim, o tratamento de imagens é um fator que poderia modificar a resposta do OCR para cada amostra, podendo melhorar ou piorar o texto reconhecido e, por isso, não se tornou objeto de estudo. Outro fator importante é o conhecimento necessário para o tratamento das imagens em softwares ou via linguagens de programação, questão que também não será tratada neste trabalho.

4 RESULTADOS E DISCUSSÕES

Esta seção destaca os resultados obtidos em cada processo do projeto, notando observações, di-

ferenças e problemas encontrados no decorrer da execução deste artigo.

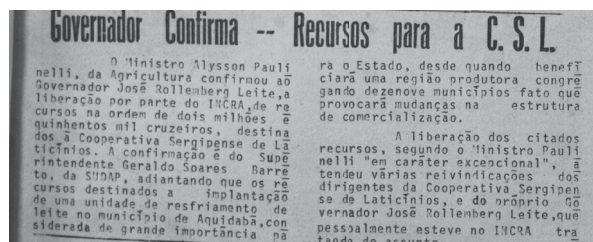
4.1 AQUISIÇÃO DAS IMAGENS

Logo no decorrer da comparação das imagens, a primeira diferença notada encontra-se na captura de luz em cada câmera, como mostrado na Figura 1.

A análise do artigo segue, avaliando trechos das páginas do jornal do dia 24 de abril, iniciando os testes com um parágrafo em boas condições, para avaliar as capacidades individuais de cada câmera e em seguida os testes em dois parágrafos deteriorados, baixa qualidade de impressão e tamanho de fonte muito pequena.

A câmera d5300 da Nikon, com 24.2 megapixels, consegue capturar mais luz e, por isso, alcança uma qualidade maior de imagem. Em hipótese, a iluminação da foto deveria ser maior, tornando a imagem mais clara, entretanto, nas configurações do modo automático, a captura da imagem foi a mais escura entre as amostras. Neste teste, o OCR conseguiu detectar 128 palavras, ou seja, 97.7% do texto.

Figura 2.1 – Recorte da página 1 feita com a d5300

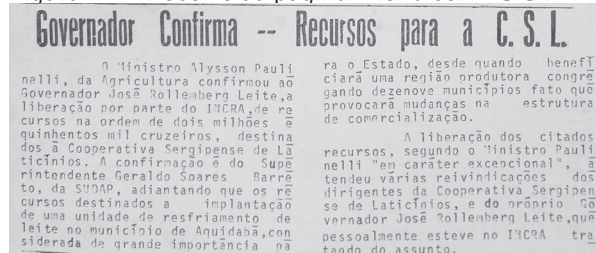


Fonte: Elaborado pelos autores

A fotografia feita pela câmera do LG G4 conseguiu ter uma iluminação melhor, ao mesmo tempo em que captura uma imagem de boa qualidade, os 16 megapixels da câmera traseira tiveram um bom recorte, sem pixelamento da silhueta das letras e um ótimo contraste, permitindo a identificação das letras separadamente, bem como a divisão de colunas do texto. No Tesseract, a resposta foi ótima, poucas palavras não foram

identificadas e dados importantes como substantivos próprios e números dos documentos não foram perdidos. Além disso, pequenos detalhes passaram despercebidos, a exemplo da letra i que foi trocada pelo número 1. O resultado da avaliação mostrou que das 131 palavras que constam no parágrafo original, 127 foram reconhecidas, um aproveitamento de 97%.

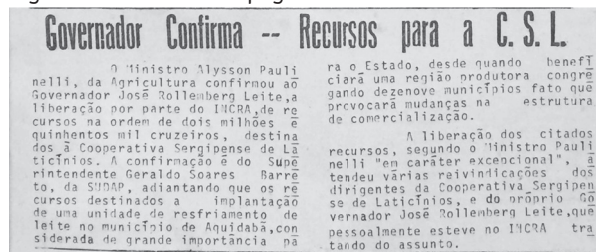
Figura 2.2 – Recorte da página 1 feita com LG G4



Fonte: Elaborado pelos autores

Apesar de possuir apenas 8 megapixels, observamos que a capacidade de captura da luz do iPhone 6 Plus é bem controlada. A amostra possui boa qualidade e permitiu um bom recorte do parágrafo selecionado. O Tesseract conseguiu ler 114 palavras, 87%.

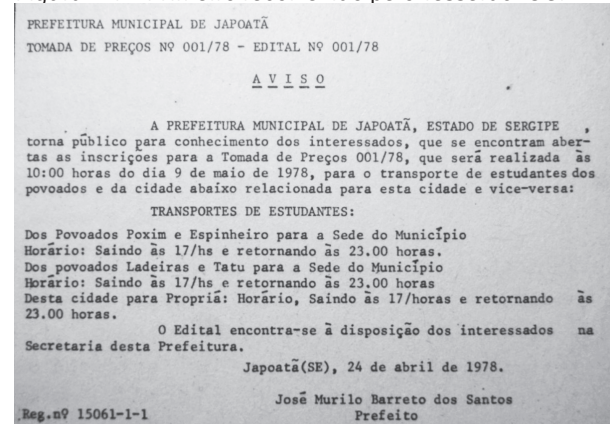
Figura 2.3 – Recorte da página 1 feita com iPhone 6 Plus



Fonte: Elaborado pelos autores

Ao avaliar um mesmo parágrafo, de baixa qualidade de impressão, seguem os resultados:

Figura 2.4 – Primeiro recorte lido pelo Tesseract OCR



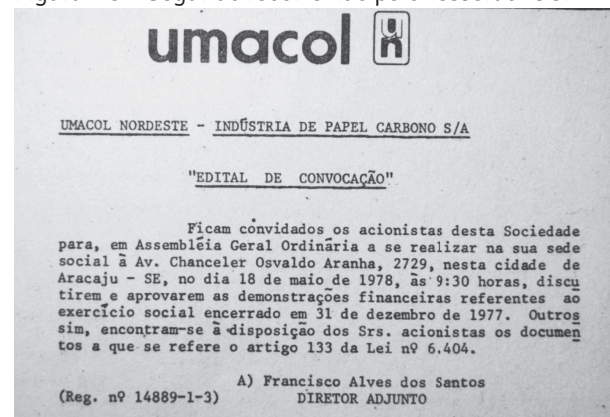
Fonte: Elaborado pelos autores

Quadro 2 – Análise da amostra 1, com total de 152 palavras

Equipamento	Quantidade de palavras reconhecidas	Porcentagem de aproveitamento do OCR
D5300	132	87%
LG G4	140	92%
iPhone 6 Plus	94	61.8%

Fonte: Dados da pesquisa.

Figura 2.5 – Segundo recorte lido pelo Tesseract OCR



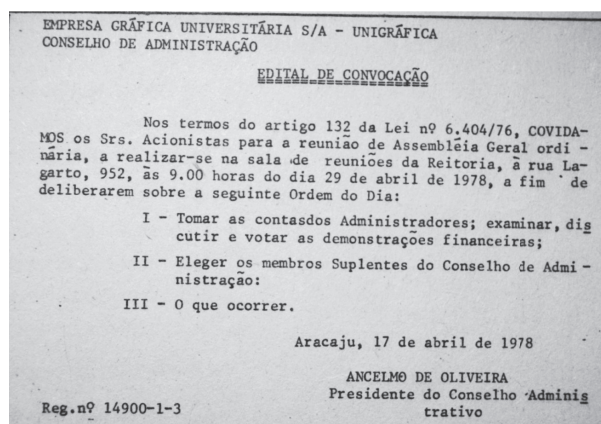
Fonte: Elaborado pelos autores

Quadro 3 – Análise da amostra 2, total de 105 palavras

Equipamento	Quantidade de palavras reconhecidas	Porcentagem de aproveitamento do OCR
D5300	88	84%
LG G4	79	75%
iPhone 6 Plus	72	68.5%

Fonte: Dados da pesquisa.

Figura 2.6 – Terceiro recorte lido pelo Tesseract OCR



Fonte: Elaborado pelos autores

Quadro 4 – Análise da amostra 3, total de 98 palavras

Equipamento	Quantidade de palavras reconhecidas	Porcentagem de aproveitamento do OCR
D5300	77	78%
LG G4	69	70%
iPhone 6 Plus	47	47%

Fonte: Dados da pesquisa

As diferenças dos resultados da leitura de cada câmera demonstram a limitação física de cada equipamento. Todavia, isso não descarta que, com o devido tratamento de imagem e softwares espe-

cíficos, o pesquisador possa obter resultados satisfatórios com equipamentos diferentes. Notando que a influência que o processamento da imagem possa afetar a leitura do OCR, propusemos avaliar os casos de diferentes formatos de imagem e técnicas de compressão, para os casos de câmeras que utilizam tecnologias diferentes.

Buscamos, analisando os diferentes formatos de imagem, verificar se as técnicas de compressão influenciam na leitura adequada do OCR. Além disso, o tamanho da imagem precisa permitir economia de espaço devido a possibilidade da necessidade de armazenamento de uma grande quantidade de arquivos fotográficos. Assim, elegemos para teste a Figura 2.1 para ser avaliada em diversos formatos e verificar o tamanho e os resultados da saída do Tesseract. Como resultado obtivemos a seguinte tabela:

Quadro 5 – Tamanho das imagens em diferentes formatos

Formato	Tamanho
JPEG	418 KB
BMP	8.88 MB
PNG	6.66 MB
TIFF	6.68 MB

Fonte: Dados da pesquisa.

Independente do formato da imagem, o OCR desenvolveu a mesma leitura, sem divergências.

5 CONCLUSÃO

Conseguimos, decorrendo dos resultados, responder as questões levantadas neste trabalho, elegendo a melhor configuração de equipamento, definindo o procedimento necessário para as fotografias e a elucidação dos conhecimentos necessários para o desenvolvimento de uma hemeroteca digital.

Tem-se como resposta, respondendo às perguntas pertinentes a pesquisa, feitas anteriormente:

Q1: O smartphone LG G4 provou ser o dispositivo de melhor custo-benefício para a aplicação explorada neste trabalho;

Q2: O formato de imagem tido como o mais adequado para a aquisição de imagem é o formato JPEG, baseado nos critérios abordados neste artigo, como tamanho de armazenamento e qualidade de imagem resultante.

Primeiramente, avaliando os três tipos de configurações de câmeras, 16 megapixels e abertura de lente de f1.8 foi tido como requisito mínimo para uma câmera capaz de realizar a função. Isso se dá pela quantidade de luz que a câmera pode capturar, permitindo recortes de qualidade das fotos, sem pixelar a imagem, o que perderia detalhes, importante para a leitura do OCR. Outra condição foi o modo automático da câmera, ela foi suficiente para configurar a lente para cada página de maneira eficiente, diminuindo o tempo necessário para as fotos e permitindo que um pesquisador sem conhecimento em fotografia possa capturar imagens satisfatórias.

As fotos foram realizadas sem flash, pois o uso deste recurso iria aumentar ainda mais o brilho refletido pelo papel, conseqüentemente perdendo boa capacidade de captura das letras do documento. O resultado disto são letras confundidas com outros símbolos. A abertura da lente recomendada é de f1.8, grande suficiente para capturar uma folha A3 (297 milímetros por 420 milímetros), ou maior.

Dentre os equipamentos avaliados, o smartphone da LG, o G4, foi selecionado por este trabalho, devido a suas configurações responderem bem a todos os requisitos e ao seu custo-benefício ser o mais adequado, quando comparado com uma câmera DSLR como a da Nikon, d5300, que traz resultados mais consistentes. Para este trabalho, o custo da d5300 a coloca como um equipamento caro, recomendado somente para casos em que a qualidade desejada seja a melhor e o orçamento não seja um problema. Este trabalho não visa exaltar uma determinada marca de equipamento e sim as configurações mínimas para realização do objetivo desta pesquisa, sendo assim, qualquer dispositivo que se encaixe no levantamento feito por este artigo deve ser considerado capaz suficiente para a função.

O ambiente precisa ser bem iluminado, a luz incidente no documento precisa eliminar qualquer sombra na página e não pode causar brilho no papel. Neste trabalho, a sala onde as fotografias foram realizadas contavam com lâmpadas fluorescentes em todo o teto, permitindo luz de todas as direções e, pela distância, não causou brilho excessivo no papel. Outra recomendação é posicionar o documento em um ângulo de 45° com a base (mesa), pois, apoiar o documento paralelo a mesa, pode causar sombra do fotógrafo no objeto fotografado. Não houve necessidade de conhecimento prévio ou estudo de assuntos relacionados a computação ou fotografia somente de ambiente.

Em diversos formatos, usando como teste a amostra de parágrafo da Figura 2.1, obtivemos uma variação muito grande dos tamanhos das imagens, notando que o formato JPEG é o mais recomendado por ser capaz de armazenar uma imagem em ótima qualidade com o menor tamanho. Considerando o formato JPEG, os testes apontaram que o tamanho da imagem não influencia na leitura do OCR, ou seja, JPEG ou BMP (o formato com o maior tamanho) resultaram na mesma saída do Tesseract. Assim, recomenda-se para uma construção de base de dados fotográficos, o uso do formato JPEG, por permitir a qualidade da imagem com economia de espaço e boa leitura do OCR.

Por fim, concluímos que os problemas para a extração de dados e digitalização de documentos por meio de fotografia estão relacionados, essencialmente, com o processo fotográfico. É importante salientar que, o equipamento e conhecimento profissional não é requisito; formatos de compressão de imagem como JPEG auxiliam na criação de uma base de dados com economia de espaço e proporciona qualidade de imagem para o uso de um OCR; e tratamento de imagem não é necessário, porém, casos específicos podem exigir o uso de algum software de processamento de imagens para a obtenção de melhores resultados.

Os autores deste artigo manifestam aqui seu agradecimento à Universidade Tiradentes por possibilitar a realização deste trabalho.

REFERÊNCIAS

FRERY, Alejandro C.; PERCIANO, Talita. **Introduction to image processing using R**. Springer, London, 2013.

KAZAMI, Kazuyuki *et al.* **Digital camera system**. U.S. Patent Application n. 10/413,142, 11 dez. 2003.

KEAVENEY, Susan M.; PARTHASARATHY, Madhavan. Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. **Journal of the academy of marketing science**, v.29, n.4, p.374-390, 2001.

Lei nº 12.527/2011. **Lei do Acesso à Informação**. Disponível em: <<http://www.planalto.gov.br/>

ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 4 maio 2018.

LEVACOV, Marília. Bibliotecas virtuais:(r) evolução? **Ciência da Informação**, v.26, n.2, 1997.

MIANO, John. **Compressed image file formats: JPEG, PNG, GIF, XBM, BMP**. Addison-Wesley Professional, 1999.

NÓBREGA COSTA, Felipe Zimmerle da; SANTOS, Wellington Pinheiro dos. Image File Formats. **Digital Document Analysis and Processing**, p. 35, 2011.

VAN OS, Ron; HANSON, Jay W.; BURGER, Rudolph E. **Scanning system and method**. U.S. Patent n. 6,480,304, 12 nov. 2002.

1 Graduando em Ciência da Computação; Membro do GPITIC.
Email: efraimleite@gmail.com

2 Bacharel em Ciência da Computação; Membro do GPITIC.
Email: victor.brito.villar@gmail.com

3 Bacharel em Ciência da Computação; Membro do GPITIC.
Email: thicianecouto@gmail.com

4 Mestre em Ciências da Computação – UFS; Doutorando em Educação; Especialista em Engenharia de Sistemas; Bacharel em Sistemas de Informação; Líder do GPITIC; Professor da UNIT. Email: gomesrocha@gmail.com

Recebido em: 10 de Fevereiro de 2018

Avaliado em: 20 de Março de 2018

Aceito em: 5 de Abril de 2018

Produção

Editora Universitária Tiradentes

Impressão

Gráfica Gutemberg

Tiragem

200

Tipografia

Signika
Bebas Neue
Arial

Papel

Capa - Papel Supremo 250g/m2

Miolo - Papel Offset 90g/m2