

Toth: uma abordagem para extração de elementos textuais em imagens com variação de iluminação e linhas de texto inclinadas

Daniel M. Kuhn¹, Cristiano R. Cervi¹, Edimar Manica²

¹Instituto de Ciências Exatas e Geociências – UPF - Passo Fundo - RS

²Campus Ibirubá - IFRS - Ibirubá - RS

138714@upf.br, cervi@upf.br, edimar.manica@ibiruba.ifrs.edu.br

Abstract. *Optical character recognition software is designed to convert textual elements of documents into editable and searchable text. This task presents specific challenges when submitted to images captured by smartphone cameras. This work presents an approach to extract textual elements in images with variation of lighting and inclined text lines. The experiments demonstrated that the approach obtained a significant increase of effectiveness in relation to the baseline, at the moment that it also presented superior efficiency.*

Resumo. *Softwares de reconhecimento óptico de caracteres têm como propósito converter elementos textuais de documentos em texto editável e pesquisável. Essa tarefa apresenta desafios específicos quando submetida a imagens capturadas por câmeras de smartphones. Este trabalho apresenta uma abordagem para extrair elementos textuais em imagens com variação de iluminação e linhas de texto inclinadas. Os experimentos demonstram que a abordagem obteve um aumento de eficácia significativo em relação ao baseline, ao instante em que também apresentou eficiência superior.*

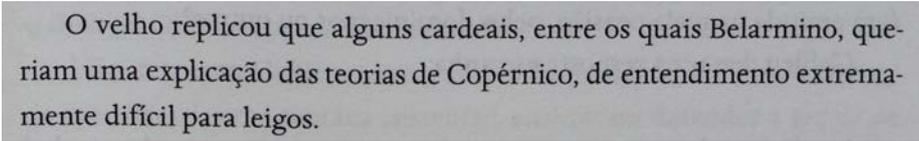
1. Introdução

A análise de *Big Data* é um aspecto chave da sociedade moderna uma vez que permite criar conhecimento a partir de dados. Essa análise traz o conhecimento para o indivíduo de uma forma direta e facilitada permitindo a emancipação das pessoas e as habilitando a agirem e tomarem decisões com mais embasamento [Manica, Dorneles and Galante 2017]. Problemas de heterogeneidade, escalabilidade, complexidade e privacidade impedem o progresso de todos os estágios do *pipeline* que extrai valor a partir de dados [Labrinidis and Jagadish 2012]. Nesse contexto, os problemas iniciam durante a aquisição de dados porque muitos dados não estão nativamente em um formato estruturado e estruturar tal conteúdo para análise futura é o principal desafio [Agrawal et al 2012].

Um exemplo de dados relevantes em um formato não estruturado é observado em textos presentes em imagens postadas nas redes sociais. Estima-se que só no instagram - atualmente a maior rede social de fotografias - sejam postadas em média 52 milhões de fotografias todos os dias [Statistic Brain, 2017]. Um percentual dessas imagens contém elementos textuais. Esse percentual, embora pequeno, representa um

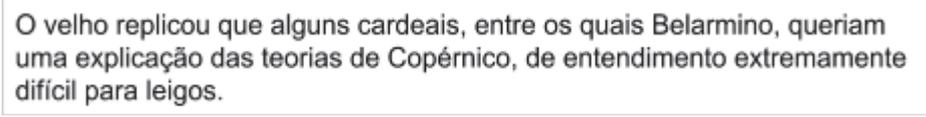
grande número de postagens, dada a escalabilidade da rede social. Extrair esses elementos textuais é útil para aprender mais sobre o usuário e fornecer recomendações mais precisas de produtos e serviços.

A extração de conteúdos textuais em imagens é realizada através do uso de softwares de Reconhecimento Óptico de Caracteres (OCR – *Optical Character Recognition*). O *OCR* é um processo de reconhecimento visual que converte documentos de texto em texto editável e pesquisável [Berchmans and Kumar 2014]. A Figura 1.a ilustra uma imagem de um trecho de um livro capturada por um *smartphone* submetida a um software de *OCR* (entrada). A Figura 1.b apresenta os elementos textuais extraídos da Figura 1.a (saída).



O velho replicou que alguns cardeais, entre os quais Belarmino, queriam uma explicação das teorias de Copérnico, de entendimento extremamente difícil para leigos.

(a)



O velho replicou que alguns cardeais, entre os quais Belarmino, queriam uma explicação das teorias de Copérnico, de entendimento extremamente difícil para leigos.

(a)

Figura 1. Exemplo de uma imagem de entrada e os elementos textuais resultantes do processo de extração.

Nos últimos trinta anos, um número substancial de pesquisas acerca de mecanismos de *OCR* foram realizadas [Islam and Noor 2016]. Em suma, a grande maioria dos esforços destinou-se a solucionar problemas decorrentes da digitalização de documentos de texto através do uso de dispositivos de *scanner*, o que resultou na obtenção de altas taxas de precisão de extração em documentos desta natureza [Asad et al 2016].

Entretanto, os métodos de pré-processamento de imagens aplicados em documentos escaneados são em diversos casos inapropriados ou insuficientes quando destinados a otimizar o reconhecimento de caracteres de imagens capturadas por câmeras de *smartphones*. Isso ocorre porque as características encontradas em imagens escaneadas são, em sua grande maioria, distintas das características presentes em arquivos de imagens obtidas através da câmera de *smartphones*. As imagens capturadas por câmeras podem apresentar variação de iluminação, linhas de texto inclinadas, baixa resolução, desfocagem e distorção de perspectiva, apresentando layouts complexos e interação entre o conteúdo e o plano de fundo [Liang, Doermann and Li 2005] [Kuhn, Cervi and Manica 2018].

Considerando o contexto apresentado, este trabalho tem como objetivo o desenvolvimento de uma abordagem denominada *Toth*, cujo propósito é extrair elementos textuais de imagens capturadas por *smartphones*, onde as imagens sofreram variação de iluminação e apresentaram linhas de texto inclinadas. Como contribuições principais, este trabalho contempla duas soluções: (i) realiza o tratamento para a variação de iluminação; e (ii) implementa um pré-processamento para rotacionar imagens com linhas de texto inclinadas. Para avaliar as soluções propostas, foram realizados experimentos com imagens reais e sintéticas contendo trechos de livros,

capturadas por *smartphones*. Comparando os resultados com um baseline, a abordagem proposta obteve um aumento de eficácia significativo, ao instante que também foi capaz de concluir a extração em tempos significativamente inferiores quando comparado ao baseline.

Este artigo está organizado da seguinte forma. Na Seção 2 são discutidos os trabalhos relacionados. A Seção 3 descreve a abordagem proposta. Na Seção 4 são apresentados os experimentos realizados, bem como são discutidos seus resultados. Finalmente, a Seção 5 apresenta as considerações finais e os trabalhos futuros.

2. Trabalhos relacionados

Esta seção descreve cinco trabalhos que possuem relação ao contexto no qual este trabalho está inserido.

No trabalho [Asad et al 2016] apresenta-se um sistema de *OCR* baseado em redes LSTM (*Long Short Term Term*), capaz de reconhecer caracteres borrados decorrentes de movimentos indesejados. Redes LSTM, são um tipo especial de redes neurais recorrentes (RNN – *Recurrent neural network*) com capacidade de recordar informações por longos períodos de tempo [Olah 2015].

Em [Kil et al 2018] apresenta-se uma abordagem para corrigir adversidades referentes à perspectiva das imagens, tratando linhas que apresentam aspecto curvilíneo. A abordagem se baseia na premissa de que a maioria dos segmentos de linha das imagens são alinhados horizontalmente ou verticalmente, codificando essas propriedades em uma função de custo. Minimizando a função, obtém-se os parâmetros de transformação para posição da câmera, curva da página e comprimento focal da câmera, utilizando-os para realizar a correção de perspectiva das imagens. Esta abordagem apresenta resultados satisfatórios tanto para documentos de texto quanto imagens com fundo rico em detalhes, como rótulos de produtos.

Em [Smith 1987] foi proposto uma nova abordagem para reconhecimento de caracteres. Esse trabalho deu origem ao motor de *OCR Tesseract* [Tesseract 2015]. Em 2005, *Tesseract* passou a ser um projeto *Open Source* e desde 2006 vem sendo desenvolvido pela *Google Inc*¹. *Tesseract* provê suporte a Unicode, capaz de reconhecer mais de 100 linguagens diferentes. Este motor de *OCR* é totalmente treinável, sendo possível adicionar novos símbolos e até mesmo novos idiomas inteiros. *Tesseract* regularmente é citado entre os *softwares* de *OCR Open Source* com maior acurácia [Gabasio 2013] [Dhiman 2013]. Considerando sua acurácia, o fato de ser *Open Source* e a possibilidade de aplicar processos de treinamento, entendemos que tais fatores são relevantes para o contexto da nossa proposta, o que foi determinante para sua adoção como baseline e também como motor de extração em nosso trabalho.

Em um trabalho prévio [Kuhn, Cervi and Manica 2018], foi avaliada a relação entre as características das imagens e a eficácia da extração de elementos textuais em imagens capturadas por *smartphones* submetidas ao *Tesseract*. O experimento avaliou o *Tesseract* com o modo de segmentação de página padrão para o cliente *Tesseract* para a interface de linha de comando (PSM_AUTO). Este trabalho identificou que a variação de iluminação e as imagens com linhas de texto inclinadas afetam a eficácia da

¹Disponível em: <<https://www.google.com/>>. Acessado em: 10/10/2018.

extração. Tais problemas foram essenciais para a motivação de se criar a abordagem *Toth*, uma vez que possibilitou a busca por solução diante das dificuldades encontradas.

A abordagem *Toth* diferencia-se dos trabalhos acima citados por implementar etapas de pré-processamento para tratar imagens com variação de iluminação e imagens com linhas de texto inclinadas.

3. Abordagem Toth

Esta seção apresenta a abordagem *Toth*, que visa a extração de elementos textuais em imagens capturadas por *smartphones*. A subseção 3.1 apresenta a visão geral da abordagem. Nas seções que seguem, são especificadas cada etapa da abordagem *Toth*.

3.1. Visão Geral

Este trabalho tem como objetivo o **desenvolvimento de uma abordagem para extração de elementos textuais em imagens capturadas por smartphones**. A Figura 2 apresenta as principais etapas da abordagem *Toth*.

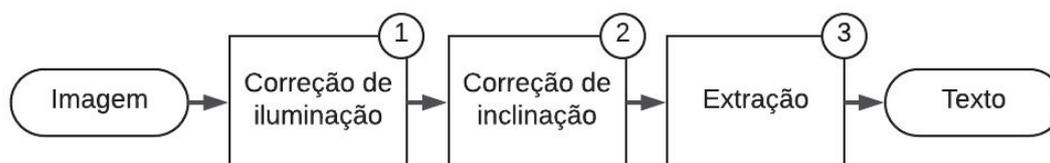


Figura 2. Fluxo de execução da abordagem *Toth*.

Toth é constituído das seguintes etapas: (1) *Correção de iluminação* - tratamento para corrigir adversidades referentes a variações de iluminação; (2) *Correção de inclinação* - pré-processamento responsável por rotacionar imagens com linhas de texto inclinadas; e (3) *Extração* - obtenção dos elementos textuais contidos nas imagens. O processo inicia com a submissão de uma imagem, passa pelas etapas 1, 2 e 3, e tem como resultado o conteúdo textual da imagem em formato editável.

3.2. Correção de iluminação

O objetivo dessa etapa é eliminar adversidades referente à iluminação da imagem que possam prejudicar a eficácia da extração. Para tal, aplica-se métodos de processamento de imagens digitais [Gonzalez 2000].

Algorithm 1: PLC (Process For Lighting Correction - Processo para Correção de Iluminação)

Input: I_i

Output: I_b

```
1  $I_i \leftarrow \text{converte\_escala\_cinza}(I_i);$   
2  $I_i \leftarrow \text{suaviza}(I_i);$   
3  $I_b \leftarrow \text{binarizacao\_adaptativa}(I_i, \text{THRESH\_BINARY});$   
4  $I_b \leftarrow \text{dilata}(I_b);$   
5  $I_b \leftarrow \text{contraí}(I_b);$   
6  $I_b \leftarrow \text{suaviza}(I_i);$   
7 return  $I_b;$ 
```

O Algoritmo 1 descreve o tratamento realizado para correção de iluminação. A entrada é uma imagem normal, já a saída é o resultado desta imagem processada. Na

linha 1, converte-se a imagem para escala de cinza. Na linha 2, aplica-se um filtro para suavizar os pixels da imagem de entrada. Em seguida (linha 3), aplica-se um algoritmo de binarização adaptativo² que segmenta a imagem em dois possíveis planos: (i) os pixels pertencentes à página do livro para a cor branca; e (ii) para os caracteres que correspondem aos elementos de interesse da imagem aplica-se a cor preta. Para a abordagem proposta, definiu-se o método de binarização adaptativa gaussiana, pois ele utiliza um limiar dinâmico para determinar se um pixel será convertido para branco ou preto. Esse limiar é definido automaticamente para cada *pixel*, considerando a cor de seus *pixels* vizinhos.

Na linha 4, aplica-se uma dilatação sobre os *pixels* brancos do fundo da imagem, fazendo com que os pixel pretos sejam *encolhidos*. O objetivo dessa etapa é remover *pixels* ruidos. Como consequência, os *pixels* que compõem os caracteres dos elementos textuais também são parcialmente *encolhidos*. Na linha 5, o processo inverso é aplicado e os *pixels* brancos são *contraídos*, fazendo com que os *pixels* que compõem os caracteres do texto voltem parcialmente ao seu estado inicial, com exceção dos *pixels* ruidos que já foram eliminados no processo realizado na Linha 4.

Quando a imagem é binarizada, os *pixels* que fazem o limite entre os caracteres e o fundo da imagem adquirem um aspecto serrilhado. Na linha 6, aplica-se um filtro de suavização para minimizar esse efeito, além de auxiliar a remoção de ruídos remanescentes.

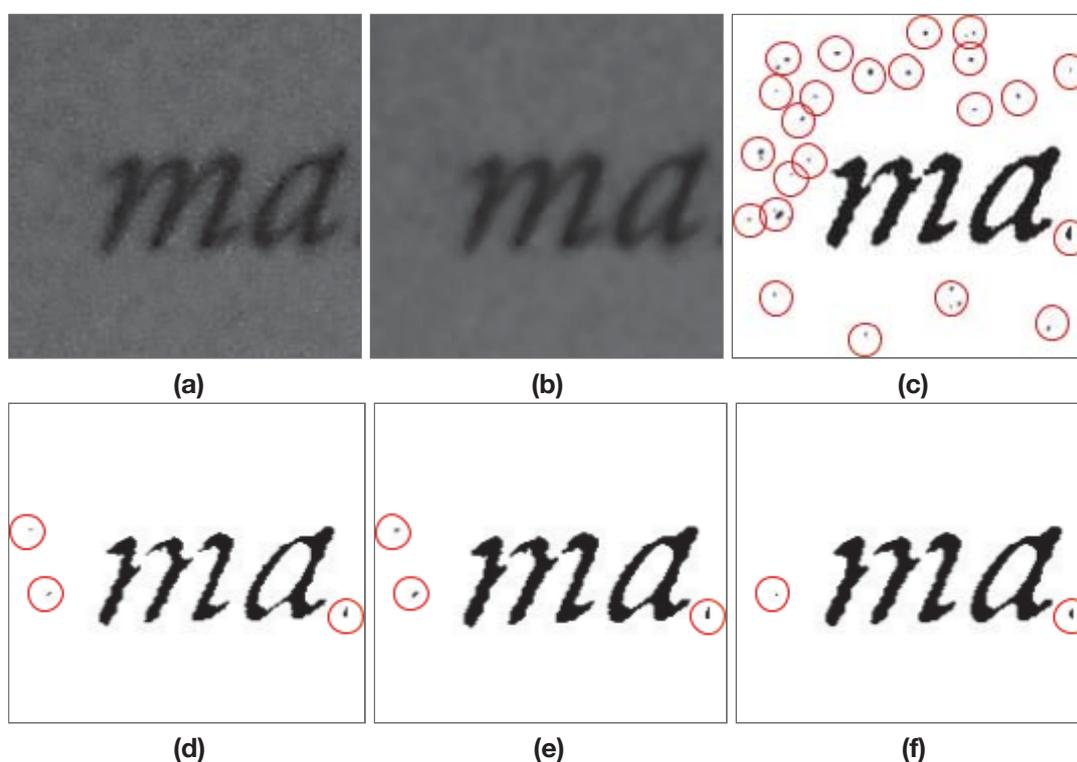


Figura 3. Etapas de processamento do algoritmo 1.

² Documentação dos métodos adaptativos disponível na biblioteca OpenCv: Disponível em: <https://docs.opencv.org/3.4/d7/d1b/group__imgproc__misc.html#ga72b913f352e4a1b1b397736707afcd3>. Acessado em 7 de novembro de 2018.

A Figura 3 ilustra as etapas de processamento do algoritmo 1. A Figura 3.(a) apresenta a imagem de entrada. A Figura 3.(b) apresenta o resultado da suavização. A Figura 3.(c) apresenta o resultado da binarização adaptativa. Observa-se a presença de vários agrupamentos de pixels ruidos, os quais foram destacados com círculos vermelhos. A Figura 3.(d) apresenta o resultado da dilatação do fundo da imagem, observa-se que grande parte dos ruidos foram removidos. Como efeito colateral desse processamento, é possível observar que os pixels que compõem os caracteres também foram parcialmente *encolhidos*. A Figura 3.(e) apresenta o resultado da contração. Observa-se que os pixels pretos foram *aumentados*, fazendo os caracteres retornarem parcialmente à sua forma inicial. Por fim, a Figura 3.(f) apresenta o resultado do filtro de suavização final. Observa-se que os pixels pretos apresentam aspecto mais uniforme, minimizando o efeito serrilhado, além de minimizar os ruidos restantes.

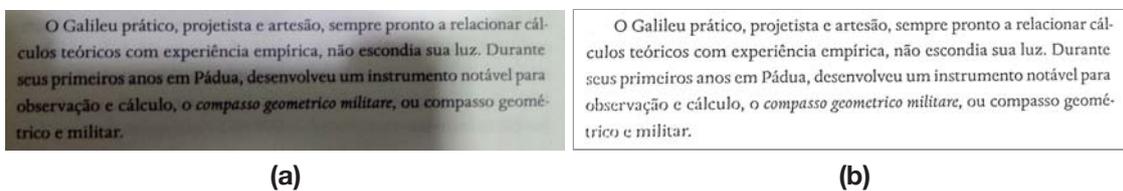


Figura 4. Exemplo de uma imagem submetida ao algoritmo PLC e a respectiva imagem processada.

A Figura 4.(a) apresenta um exemplo de imagem com variação de iluminação (entrada) e a imagem de saída obtida após a etapa de correção de iluminação (Figura 4.b). Observa-se que a imagem foi devidamente tratada, eliminando-se as adversidades relacionadas à iluminação.

3.3. Correção de inclinação

A correção da inclinação rotaciona as imagens que possuem linhas de texto inclinadas. Para realizar a rotação das imagens foi proposto o algoritmo 2.

Algorithm 2: RI (Rotation of Images - Rotação de Imagens)

Input: I_i
Output: I_r

- 1 $I_i \leftarrow \text{converte_escala_cinza}(I_i);$
- 2 $Angle \leftarrow \text{PDAL}(I_i);$
- 3 $I_r \leftarrow \text{PRI}(I_i, Angle);$
- 4 **return** $I_r;$

Como entrada, o algoritmo 2 recebe uma imagem que já passou pela etapa de correção de iluminação. A saída é uma imagem rotacionada. Inicialmente, converte-se a imagem para escala de cinza (linha 1). Na linha 2 o algoritmo PDAL (Algoritmo 3) retorna o ângulo de rotação para a imagem. Por fim (linha 3), o algoritmo PRI (Algoritmo 4) realiza a rotação da imagem.

Algorithm 3: PDAL (Process to Detect the Angle of Lines - Processo para detectar o ângulo das Linhas)

Input: I_i, Min_line
Output: $Angle$

- 1 $I_b \leftarrow \text{binarizacao_adaptativa}(I_i, THRESH_BINARY_INV);$
- 2 $I_b \leftarrow \text{dilata}(I_b);$
- 3 $I_b \leftarrow \text{contraí}(I_b);$
- 4 $L \leftarrow \text{identifica_linhas}(I_b, Min_lines);$
- 5 $sum \leftarrow 0;$
- 6 **for** $L_i \in L$ **do**
- 7 $sum += \text{calcula_angulo}(L_i);$
- 8 **end**
- 9 **return** $\text{media}(sum / \text{size}(L));$

Algorithm 4: PRI (Process to Rotates Image - Processo para Rotacionar Imagem)

Input: $I_i, Angle$
Output: I_r

- 1 $Rad \leftarrow \text{converte_radianos}(Angle);$
- 2 $W_i \leftarrow \text{calcula_nova_largura}(Rad);$
- 3 $H_i \leftarrow \text{calcula_nova_altura}(Rad);$
- 4 $C_i \leftarrow \text{obtem_centro_imagem}(I_i);$
- 5 $M_i \leftarrow \text{obtem_matrix_rotacao}(C_i, Angle);$
- 6 $M_i \leftarrow \text{ajusta_centro}(M_i, W_i, H_i, C_i);$
- 7 $I_r \leftarrow \text{aplica_transformacao_afim}(M_i, I_i);$
- 8 **return** $I_r;$

O algoritmo 3 é responsável por identificar as linhas de texto e calcular o ângulo médio de inclinação. Como entrada, o algoritmo recebe uma imagem e o comprimento mínimo das linhas de texto que serão consideradas. Como saída, o algoritmo retorna o ângulo para rotacionar a imagem. Na linha 1, realiza-se a binarização adaptativa da imagem de entrada. A binarização é semelhante à realizada no algoritmo 1, exceto pelo parâmetro *THRESH_BINARY_INV*, que neste procedimento converte o texto para a cor branca e o fundo da imagem para a cor preta. Na linha 2 aplica-se uma dilatação para conectar no sentido horizontal os *pixels* que compõem as palavras que foram o texto. Na linha 3, realiza-se a contração parcial dos píxeis dilatados na linha 2. O objetivo do tratamento realizado nas linhas 2 e 3 é conectar todas as palavras de uma determinada linha.

Na linha 4, identifica-se as retas formadas dentro de cada linha de texto, respeitando o limiar de comprimento definido. Na linha 5, inicializa-se uma variável para armazenar o somatório dos ângulos da reta. No laço 6-8, realiza-se o somatório dos ângulos das retas. Por fim, calcula-se o ângulo médio das retas.

O algoritmo 4 é responsável por rotacionar as imagens. Como entrada, recebe-se uma imagem para rotacionar e o ângulo da rotação desejada. O algoritmo retorna a imagem devidamente rotacionada. Na linha 1 converte-se o ângulo de rotação para radianos. Nas linhas 2 e 3 calcula-se, respectivamente, a nova largura e a nova altura que a imagem terá após a rotação. Na linha 4 obtém-se o ponto central da imagem de entrada. Na linha 5 obtém-se a *matriz de rotação 2D*³ da imagem a ser rotacionada. Na linha 6 ajusta-se o centro da matriz de rotação. Por fim (linha 7), submete-se a imagem para realizar a *transformação afim*⁴, conforme a matriz de rotação especificada.

³ Documentação do OpenCv: matriz de rotação 2D. Disponível em: <https://docs.opencv.org/3.1.0/da/d54/group__imgproc__transform.html#gafbbc470ce83812914a70abfb604f4326> Acessado em: 25/08/2018.

⁴ Documentação do OpenCv: transformação afim. Disponível em: <https://docs.opencv.org/3.1.0/da/d54/group__imgproc__transform.html#ga0203d9ee5fcd28d40dbc4a1ea4451983> Acessado em: 25/08/2018.

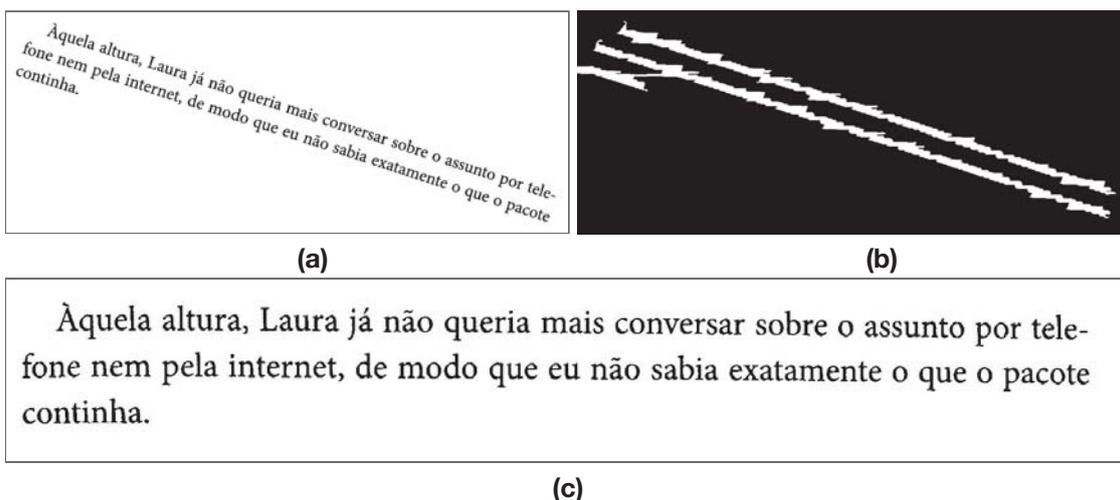


Figura 5. Eficácia da extração em relação às características de interesse.

A Figura 5.(a) apresenta um exemplo de imagem submetida ao algoritmo 2. A Figura 5.(b) apresenta a imagem resultante do processamento para identificar as linhas de texto do algoritmo 3. A Figura 5.(c) apresenta a imagem resultante do algoritmo 2 sobre a imagem de entrada (Figura 5.a). Observa-se que a imagem foi devidamente rotacionada, alinhando as linhas de texto no sentido horizontal.

3.4. Extração

Na etapa de *Extração*, os elementos textuais contidos nas imagens são extraídos. A abordagem *Toth* faz uso do motor *OCR Tesseract*. Este motor de *OCR* é regularmente citado entre os *softwares* de *OCR Open Source* com maior acurácia. Além disso, é possível aplicar processos de treinamento para identificar novos símbolos.

4. Avaliação experimental

Esta seção descreve os experimentos realizados com o intuito de avaliar a eficácia e a eficiência da extração da abordagem *Toth*. Esta seção está organizada da seguinte forma. A subseção 4.1 apresenta as bases de dados utilizadas. A seção 4.2 define as métricas utilizadas e o baseline. A subseção 4.3 apresenta as configurações do ambiente onde os experimentos foram conduzidos. A subseção 4.4 descreve os experimentos realizados e os resultados obtidos. Finalmente, a Seção 4.5 discute os casos de falha da abordagem *Toth*.

4.1. Bases de dados

Para os experimentos foram utilizadas três bases de dados, as quais foram criadas exclusivamente para este trabalho. A base de dados ERBD é uma base real, constituída de 160 imagens de trechos de livros capturadas por *smartphones*. Desse total, 82 imagens (51,25%) foram coletadas manualmente de publicações de redes sociais. As 78 amostras restantes (48,75%) foram fornecidas por um grupo de voluntários (leigos).

As imagens contidas nesta base possuem características variadas, entre elas: (i) linhas de texto inclinadas; (ii) linhas de texto com aspecto curvilíneo; (iii) variação de iluminação; (iv) caracteres pouco nítidos. Ressalta-se que as imagens dessa base não foram segmentadas corretamente, ou seja, na fase de recorte da imagem realizada pelos voluntários, manteve-se fragmentos de parágrafos vizinhos. A base de dados ERBD foi

utilizada para: (i) identificar características nas imagens que potencialmente afetam a eficácia da extração; (ii) avaliar a relação da eficácia da extração e as características identificadas.

A base de dados BD_ILUMINA contém 50 imagens de trechos de livros capturadas por *smartphones* com variação de iluminação. Criou-se essa base de dados para avaliar a etapa de correção de iluminação (seção 3.2). As imagens que compõem essa base de dados foram capturadas com aproximadamente 25 centímetros de distância em relação às páginas dos livros. Ressalta-se que o processo de captura das imagens, bem como a definição da área de interesse foram realizadas por um usuário especialista. Dessa forma, a área de interesse das imagens foram definidas corretamente, não havendo presença de termos ou caracteres pertencentes aos parágrafos vizinhos.

A base de dados BD_ROTACIONA possui 100 imagens com trechos de livros capturadas por *smartphones* com linhas de texto inclinadas. Essa base de dados foi criada para avaliar a etapa de correção de inclinação (seção 3.3). Para compor a base de dados, foram capturadas imagens há aproximadamente 25 centímetros de distância em relação às páginas dos livros. Após, a área de interesse das imagens foram definidas por um usuário especialista, não havendo presença de termos ou caracteres pertencentes aos parágrafos vizinhos. As imagens foram então automaticamente rotacionadas por um algoritmo que atribuiu a cada uma das imagens uma rotação aleatória entre $[-30, 30]$ graus. Portanto, essa é uma base sintética que simula a ocorrência de linhas inclinadas.

As três bases de dados possuem seus respectivos gabaritos que descrevem a extração ideal para cada imagem. Os gabaritos foram criados por um usuário especialista que transcreveu manualmente o conteúdo textual de cada imagem. O texto em formato digital foi então armazenado, mantendo a devida relação entre o arquivo da imagem e seu respectivo conteúdo textual.

4.2. Métricas e baseline

Para avaliar a eficácia da extração, foram adotadas as métricas tradicionais de recuperação de informação: precisão (*precision*), revocação (*recall*) e F1 (*F-measure*). Para este trabalho, contextualizando a definição de Precisão e Revocação apresentada em [Baeza-Yates and Ribeiro-Neto 2013], assume-se que a precisão mede a fração dos termos extraídos que é relevante e a revocação mensura a fração dos termos relevantes que foi extraída. Portanto:

$$precisão = \frac{|Termos\ relevantes \cap Termos\ extraídos|}{|Termos\ extraídos|}$$

$$revocação = \frac{|Termos\ relevantes \cap Termos\ extraídos|}{|Termos\ relevantes|}$$

Um termo relevante é aquele que foi extraído de uma imagem corretamente e, portanto, está presente no gabarito. Por fim, o F1 consiste na média harmônica entre os índices de precisão e revocação, com o objetivo de fornecer um só índice de medida. Quanto maior a revocação, precisão e a F1, maior a eficácia da abordagem. A F1 é definida como segue:

$$F1 = 2 \cdot \frac{precisão \cdot revocação}{precisão + revocação}$$

Ressalta-se que acentuações, distinção de letras maiúsculas e minúsculas, bem como, pontuações foram desconsideradas na análise.

Utilizou-se o tempo de extração como métrica de eficiência. O tempo de extração indica o tempo total gasto para extrair os elementos textuais de uma determinada imagem. No que se refere à abordagem *Toth*, o tempo de extração consiste no tempo de pré-processamento das imagens, acrescido do tempo demandado pelo extrator.

A significância estatística foi verificada através do teste T (Student's t-test) [FINN 1996]. O limiar de significância utilizado foi $\alpha = 0,05$ (padrão). Portanto, quando o valor de p_bicaudal calculado pelo Teste T for menor que α , confirma-se com uma confiança de 95% que há diferença estatisticamente significativa entre as abordagens comparadas.

A abordagem *Toth* foi comparada com o *Tesseract*. Os fatores determinantes para a adoção deste motor OCR como baseline foram: (i) sua acurácia; (ii) o fato de ser um software Open Source; (iii) a possibilidade de aplicar processos de treinamento.

4.3. Ambiente experimental

O extrator adotado no experimento foi o *Tesseract* na versão 3.05. Para este trabalho, utilizou-se uma versão pré-compilada do *Tesseract* para *smartphones* com o sistema operacional *Android* (*Tess-two*)⁵.

A abordagem *Toth* faz uso das funções de processamento de imagem da biblioteca *OpenCv* [OpenCv 2018], portanto, utilizou-se uma versão pré compilada (Versão 3.1.0) para para *smartphones* com o sistema operacional *Android*.

O ambiente de execução do experimento consiste em um *Smartphone Galaxy S4 - I9505, Android 6.0 Lollipop*, com 2 GB de RAM e 16 GB de armazenamento.

O modo de extração definido no *Tesseract* foi *OEM_DEFAULT*, utilizando os arquivos padrão de dados do idioma Português Brasil e Inglês. Após realizada a configuração, as imagens foram isoladamente submetidas ao extrator. Foram registrados os tempos necessários para a extração. Registrou-se também, os tempos de processamento demandados para as etapas de *Correção de iluminação* e *Correção de inclinação*.

O parâmetro do algoritmo 3 para comprimento mínimo entre o ponto de início e fim da linha, foi definido empiricamente para 70% a largura da imagem a ser rotacionada.

4.4. Experimentos e resultados

Esta subseção apresenta os experimentos realizados, bem como seus respectivos resultados. A subseção 4.4.1 apresenta o experimento no qual foram identificadas as características que potencialmente afetam a eficácia da extração. O experimento apresentado na subseção 4.4.2 aborda a relação entre as características e a eficácia da extração. Na subseção 4.4.3, verifica-se a melhor configuração de parâmetros (modo de segmentação) para o *Tesseract* e *Toth*. A subseção 4.4.4 analisa qual o método mais eficaz para imagens rotacionadas. A subseção 4.4.5 analisa qual o método mais eficaz para corrigir adversidades referentes à variação de iluminação. A subseção 4.4.6 analisa

⁵ Repositório oficial do projeto Tess-two. Disponível em: <<https://github.com/rmtheis/tess-two>> Acessado em: 08/07/2018.

os tempos de processamento para cada um dos métodos propostos. Por fim, na subseção 4.5.7 são analisados os casos de falha.

4.4.1. Características que afetam a eficácia da extração

O objetivo desse experimento foi verificar a ocorrência de características que potencialmente afetam a extração. Para responder essa questão, classificou-se a ocorrência de características presentes nas imagens da base ERBD.

A identificação de características nas imagens foi dividida em processos manuais e automatizados. Como processos manuais, com possibilidade binária de resposta (Afirmativo ou Negativo) as seguintes questões foram respondidas por um usuário especialista: (i) as linhas de texto apresentam aspectos curvilíneos?; (ii) é possível identificar variações de iluminação ou presença parcial ou total de sombra sobre a imagem?; (iii) os caracteres do conteúdo textual estão nítidos? Esse processo foi repetido para cada uma das imagens da base de dados.

Para a identificação da inclinação das linhas de texto, utilizou-se o algoritmo de rotação de imagens (algoritmo 4). Foram consideradas imagens rotacionadas, aquelas que apresentaram graus de inclinação fora do intervalo $[-4, 4]$ graus.

Tabela 1. Ocorrência das características de interesse na base de dados.

Característica	Ocorrência da característica	Ocorrência exclusiva das característica
Linhas de texto inclinadas	23 (14,38%)	15 (9,38%)
Linhas de texto curvilíneas	30 (18,75%)	10 (6,25%)
Variação de iluminação	65 (40,63%)	42 (26,25%)
Caracteres pouco nítidos	28 (17,5%)	16 (10%)
Ausência das características de interesse	48 (30%)	48 (30%)

A Tabela 1 apresenta a ocorrência de cada característica na base de dados. Observa-se que o maior percentual (40,63%) refere-se a imagens com variação de iluminação. As imagens que não apresentam nenhuma das características de interesse, representam 30% da base.

Ressalta-se que há imagens que apresentam mais de uma característica, visto que a base de dados é constituída de imagens reais. Os números entre parênteses na segunda coluna da Tabela 1 referem-se ao percentual de imagens que apenas possui *aquela* característica.

4.4.2. Relação entre as características e a eficácia da extração

O objetivo do experimento foi responder a seguinte questão: **quais as características que mais afetam a extração?** Para isso foi analisada a F1 da extração em imagens com as características de interesse. Esse experimento foi realizado sobre o *Tesseract* com o modo de segmentação de página padrão para a *API C++* (PSM_SINGLE_BLOCK).

Ressalta-se que para realizar o experimento, considerou-se apenas as imagens que possuíam uma única característica de interesse, as quais estão representadas na segunda coluna da Tabela 1.

A Figura 6 apresenta os resultados obtidos, onde as colunas representam o percentual de F1.

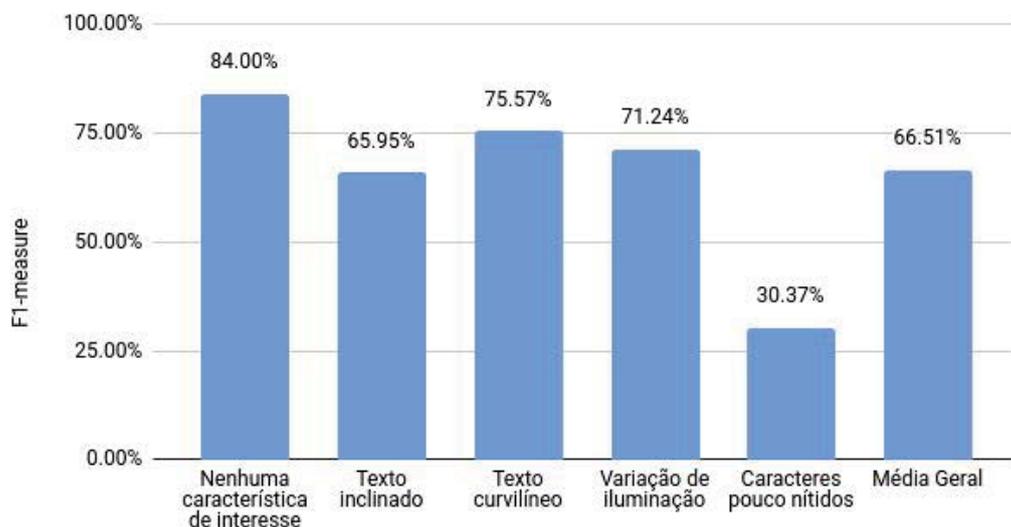


Figura 6. Eficácia da extração em relação às características de interesse.

Observa-se na Figura 6 que as imagens que não possuem nenhuma das características de interesse obtiveram uma média de F1 de 84%. As imagens que possuem texto inclinado obtiveram uma eficácia de 65,95%. Imagens com texto curvilíneo e imagens com variação de iluminação obtiveram respectivamente 75,57% e 71,24% de eficácia. A média geral da eficácia da extração alcançou 62,51%, sugerindo que outras características não consideradas neste trabalho podem estar comprometendo a eficácia de extração.

Este trabalho propõe métodos para duas classes de características, quais sejam, imagens com variação de iluminação e imagens com linhas de texto inclinadas. O critério adotado para selecionar essas características foi o seguinte: (i) selecionar do experimento 1 a característica com maior ocorrência na base de dados (imagens com variação de iluminação que ocorreu em 40,63% das imagens); e (ii) selecionar do experimento 2 a segunda característica que obteve o menor percentual de eficácia (imagens com linhas de texto inclinadas, com eficácia de 65,95%). Embora as imagens que contêm caracteres pouco nítidos tenham obtido a menor eficácia, desconsiderou-se uma implementação para corrigir essa característica, porque ela está relacionada a múltiplos fatores, entre eles: (i) baixa resolução da câmera; (ii) movimentos indesejados no instante da captura; (iii) baixa iluminação.

4.4.3. Modo de segmentação de página

O objetivo desse experimento foi responder a seguinte questão: **qual o modo de segmentação de página com maior eficácia?** Para responder essa questão, foram comparados diferentes modos de segmentação de página do *Tesseract*. Para este experimento, foram utilizadas as bases de dados *BD_ILUMINA* e *BD_ROTACIONA*.

Foram comparados quatro modos de segmentações de páginas disponíveis no *Tesseract*: *PSM_AUTO_OSD* - Segmentação automática de páginas com orientação e detecção de linguagem(*OSD*); (ii) *PSM_AUTO_ONLY* - Segmentação de página

automática, mas sem *OSD*; (iii) *PSM_AUTO* - Segmentação de página automática mas sem *OSD*; (iv) *PSM_SINGLE_BLOCK* - Assume um único bloco uniforme de texto.

Foram desconsiderados, por estarem fora do escopo deste trabalho, os demais modos de segmentação de páginas disponíveis no *Tesseract*, por se tratarem de configurações específicas para (i) extração de linhas, termos ou caracteres isolados; (ii) extrações para texto com orientação vertical e extração de elementos textuais dispostos em círculos; e (iii) extrações de textos esparsos ou que não levam em consideração a ordem dos elementos textuais da imagem.

Na sequência, a Figura 7 apresenta comparativos entre os quatro modos de segmentação de página.

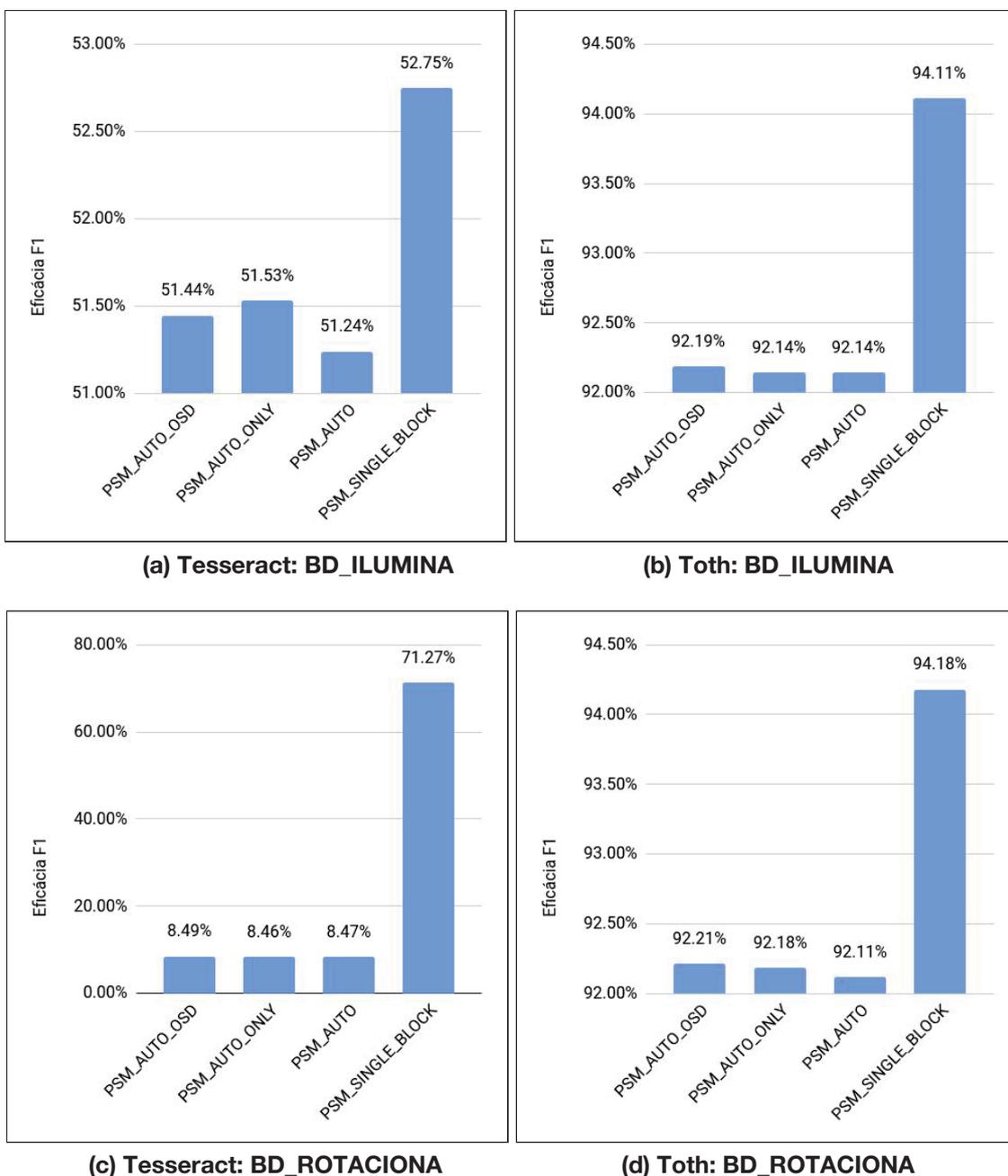


Figura 7. Eficácia da extração em relação às características de interesse.

A Figura 7.(a) apresenta a F1 para o *Tesseract* sobre a base de dados BD_ILUMINA. A Figura 7.(b) apresenta a F1 para o *Toth* também sobre a base de dados BD_ILUMINA. A Figura 7.(c) apresenta a F1 para o *Tesseract* sobre a base de dados BD_ROTACIONA. Por fim, a Figura 7.(d) apresenta a F1 para o *Toth* sobre a base de dados BD_ROTACIONA.

Conforme os resultados obtidos, observa-se que o modo de segmentação de página com maior eficácia é o PSM_SINGLE_BLOCK. Isto ocorre nas duas bases de dados e para os dois métodos. Os próximos experimentos utilizam esse modo de segmentação de página, tanto para o *Tesseract* quanto para o *Toth*, referindo-se a eles apenas como *Tesseract* e *Toth*, respectivamente.

4.4.4. Método para imagens rotacionadas

O objetivo desse experimento foi responder a seguinte questão: **qual o método com maior eficácia para imagens rotacionadas?** Para responder essa questão, comparou-se o *Tesseract* com a abordagem *Toth*. Para este experimento, utilizou-se a base de dados BD_ROTACIONA.

A Figura 8 apresenta os resultados de precisão, revocação e F1.

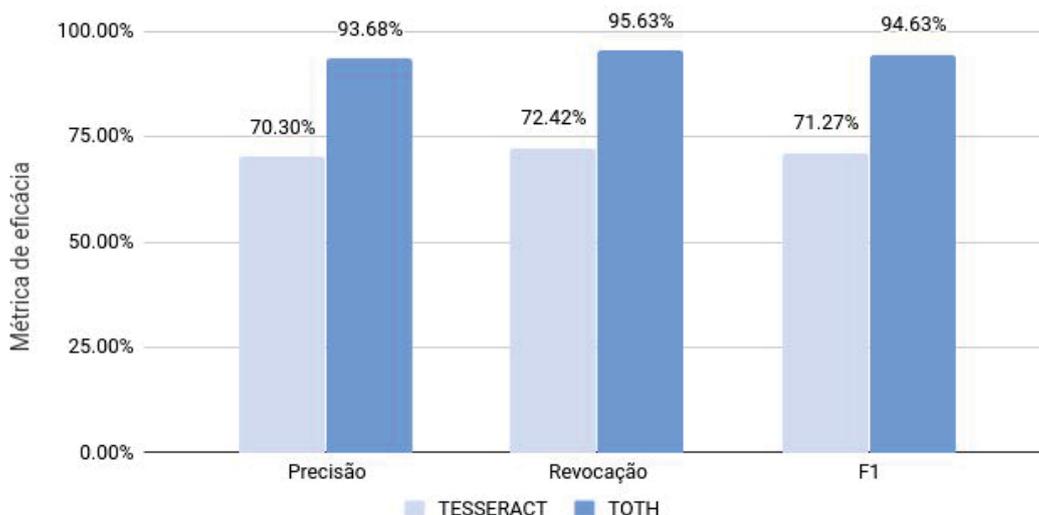


Figura 8. Eficácia da extração para imagens com linhas de texto inclinadas.

Conforme pode ser observado na Figura 8, a abordagem *Toth* obteve maior eficácia, apresentando maior precisão (93,68%) em relação ao *Tesseract* (70,30%) e maior revocação (95,63%), contra 72,42% obtido pelo *Tesseract*. Dessa forma, a abordagem *Toth* obteve uma F1 de 94,63%, contra 71,27% obtida pelo *Tesseract*. Isso se deve pelo fato de que *Toth* rotaciona as imagens antes de submeter ao motor de extração. O Teste T mostra que esse ganho é estatisticamente significativo visto que o valor do p_bicaudal (1,27E-12) é menor que o coeficiente de significância adotado ($\alpha = 0,05$).

4.4.5. Método para imagens com variação de iluminação

Esse experimento tem por objetivo responder a seguinte questão: **Qual o método com maior eficácia para imagens com variação de iluminação?** Para responder essa questão, comparou-se o *Tesseract* com a abordagem *Toth*. Utilizou-se a base de dados BD_ILUMINA para realização do experimento.

A Figura 9 apresenta os resultados obtidos de precisão, revocação e F1.

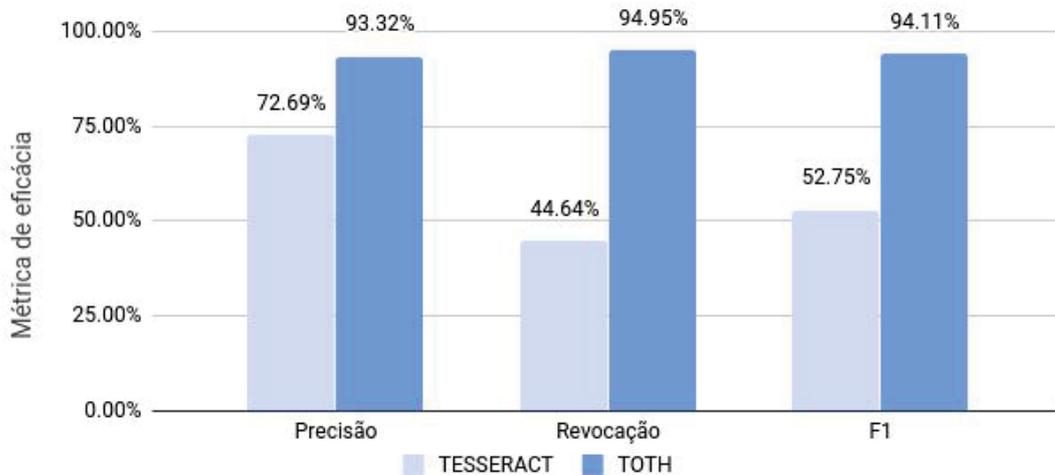


Figura 9. Eficácia da extração para imagens com variação de iluminação.

Conforme apresentado na Figura 9, a abordagem *Toth* obteve 93,32% de precisão, contra 72,69% obtida pelo *Tesseract*. A abordagem *Toth* também obteve maior revocação, apresentando eficácia de 94,95% contra 44,64% do *Tesseract*. Assim, a abordagem *Toth* obteve 94,11% de eficácia, contra 52,75% de eficácia obtida pelo *Tesseract*. O ganho de F1 obtido pela abordagem *Toth* é estatisticamente significativo, visto que o valor do p -bicaudal ($2,13E-23$) é menor que o coeficiente de significância adotado ($\alpha = 0,05$). Os resultados obtidos demonstram que a baixa eficácia do *Tesseract* (52,75%) é em grande parte ocasionada pela baixa revocação (44,64%).

Em outras palavras, significa que, embora 72,69% dos termos extraídos pelo *Tesseract*, tenham sido extraídos corretamente, apenas foram extraídos 44,64% do total de termos contidos nas imagens. A Figura 10 apresenta um exemplo de caso que aborda esse comportamento.

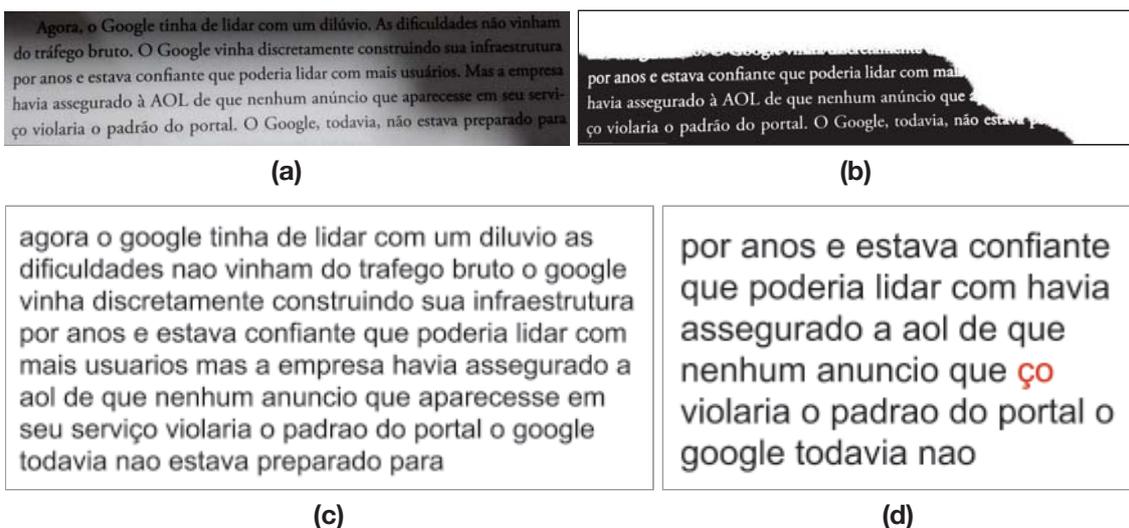


Figura 10: Respectivamente, um exemplo de imagem (a), a imagem resultante do pré-processamento realizado pelo Tesseract (b), o gabarito ideal de extração (c) e os termos extraídos pelo Tesseract (d).

A Figura 10.(a) apresenta uma das imagens submetidas ao *Tesseract*. Destaca-se que essa imagem possui variação de iluminação, uma vez que o canto superior direito é mais escuro que o canto inferior esquerdo. A Figura 10.(b) apresenta a imagem resultante do pré-processamento realizado pelo *Tesseract*. Observa-se que o texto da parte mais escura da imagem é apagada no pré-processamento do *Tesseract*. A Figura 10.(c) apresenta o gabarito com os termos da extração ideal para esta imagem. O gabarito possui 62 termos. A Figura 10.(d) apresenta os termos extraídos pelo *Tesseract*. Foram extraídos 28 termos, sendo destes, 1 termo extraído incorretamente. Observa-se que a etapa de pré-processamento não é capaz de lidar corretamente com a variação de iluminação, fazendo com que apenas uma parcela dos elementos textuais permaneçam visíveis. Observa-se também, que dos termos extraídos (Figura 10.d), apenas o fragmento “ço”, não corresponde a um termo válido no gabarito. Portanto, essa imagem obteve uma precisão de 96,42%. Entretanto, 35 termos pertencentes ao gabarito não foram extraídos, pois estavam localizados no espaço em branco da Figura 10.(b), obtendo assim, uma revocação de apenas 43,54%.

A Figura 11.(a) apresenta a imagem resultante da etapa de correção de iluminação da abordagem *Toth* para a mesma imagem da Figura 10.(a).

Agora, o Google tinha de lidar com um dilúvio. As dificuldades não vinham do tráfego bruto. O Google vinha discretamente construindo sua infraestrutura por anos e estava confiante que poderia lidar com mais usuários. Mas a empresa havia assegurado à AOL de que nenhum anúncio que aparecesse em seu serviço violaria o padrão do portal. O Google, todavia, não estava preparado para

(a)

agora o google tinha de lidar com um diluvio as dificuldades nao vinham do trafego bruto o google vinha discretamente construindo sua infraestrutura por anos e estava confiante que poderia lidar com mais usuarios mas a empresa havia assegurado a aol de que nenhum anuncio que aparecesse em seu servi ço violaria o padrao do portal o google todavia nao estava preparado para

(b)

Figura 11. Respectivamente, a imagem resultante do pré-processamento realizado pelo Toth (a) e os termos extraídos (b).

Observa-se que a etapa é capaz de realizar o processo de binarização sem interferências visíveis ocasionadas pela variação de iluminação. A Figura 11.(b) apresenta a extração obtida para a imagem da Figura 11.(a). Observa-se que do total de termos extraídos (63), apenas dois termos não foram extraídos corretamente: (1) *servi*; e (2) *ço*. Portanto, a precisão obtida foi de 96,83 %. Ainda, dos 62 termos presentes no gabarito, 1 (um) termo não foi extraído - o termo *serviço* -, obtendo assim, uma revocação de 98,38%. Ressalta-se que uma etapa de pós-processamento léxico - fazendo uso de dicionários - poderia corrigir o erro na extração desta imagem.

Portanto, conforme os resultados obtidos, é possível constatar que, tanto para imagens rotacionadas (Experimento 4.4.4), quanto para imagens com variações de iluminação (Experimento 4.4.5), a abordagem *Toth* obtém eficácia significativamente superior ao *Tesseract*.

4.4.6. Tempo

Esse experimento tem por objetivo responder a seguinte questão: **Qual o tempo de processamento para os métodos propostos?** Para responder essa questão, foram registrados os tempos demandados para realizar as extrações.

A Figura 12.(a) apresenta os tempos (em milissegundos) demandados pelo *Tesseract* e *Toth*, respectivamente, para realizar as extrações das imagens da base BD_ILUMINA.

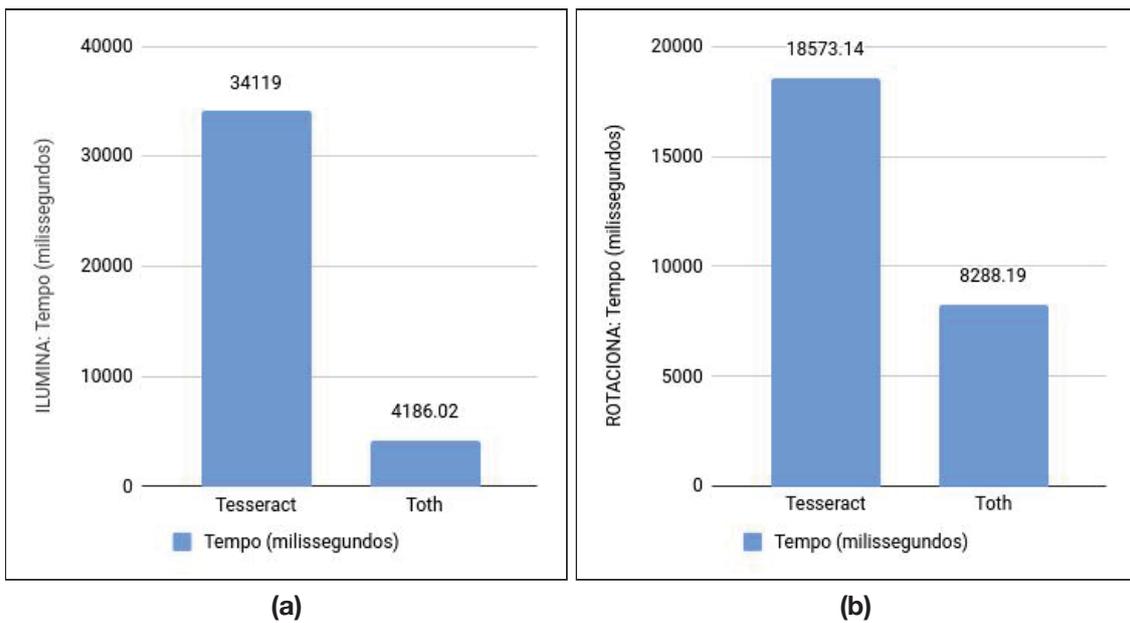


Figura 12. Tempos necessários para a extração demandado pelo Tesseract e Toth.

A Figura 12.(b) apresenta os tempos demandados pelo *Tesseract* e *Toth* respectivamente, para realizar as extrações das imagens da base BD_ROTACIONA. Ressalta-se que para os tempos da abordagem *Toth*, considerou-se o tempo de pré-processamento, acrescido do tempo de extração. Dessa forma, os tempos para a abordagem *Toth* para a base BD_ILUMINA soma-se o tempo demandado pelo Algoritmo 1 + o tempo demandado pela extração. Para os tempos demandados pela abordagem *Toth* na base BD_ROTACIONA, soma-se o tempo demandado pelo Algoritmo 2 + o tempo demandado pela extração. Conforme os resultados obtidos, observa-se que a abordagem *Toth* realiza a extração em menos tempo nos dois casos. Para a base BD_ILUMINA, *Toth* foi 87,73% mais eficiente. Para a base de dados BD_ROTACIONA foi 55,37% mais eficiente. Isso ocorre porque as etapas de pré-processamento do *Toth* reduzem o esforço demandado pelo extrator.

O ganho de eficiência obtido pela abordagem *Toth* é estatisticamente significativo, visto que tanto para as extrações das imagens da base BD_ILUMINA ($p_{\text{bicaudal}} 4,48E-05$), quanto para as extrações das imagens da base BD_ROTACIONA ($p_{\text{bicaudal}} 1,0E-09$) os valores dos $p_{\text{bicaudais}}$ são menores que o coeficiente de significância adotado ($\alpha = 0,05$).

4.5. Casos de falha

Embora no contexto geral, analisando o resultado do processamento sobre as bases de dados com um grande número de imagens, a abordagem *Toth* tenha obtido eficácia significativamente superior ao *Tesseract*, quando analisados isoladamente, existem extrações em que a abordagem *Toth* obteve menor eficácia. Esta seção abordará casos de falha, em que a abordagem *Toth* não obteve eficácia superior quando comparado com *Tesseract*.

Para a base de dados BD_ILUMINA, das 50 imagens, nenhuma extração realizada pela abordagem *Toth* obteve menor eficácia. Analisou-se as 4 (quatro) imagens com menor F1 - as quais juntas, possuem uma média de 85.53% de eficácia - obtido pelo *Toth* e não foram encontrados indícios visíveis que apontassem um motivo para a menor eficácia. Trabalhos futuros poderiam aprofundar esta análise com o objetivo de encontrar aspectos relacionados à menor eficácia.

Para a base de dados BD_ROTACIONA, das 100 imagens, 14 extrações realizadas pela abordagem *Toth* obtiveram menor eficácia e, ainda, 17 extrações obtiveram a mesma eficácia apresentada pelo *Tesseract*. Ao analisar empiricamente estas imagens, observou-se que tanto nos 14 casos que obtiveram menor eficácia, quanto nas 17 extrações que obtiveram a mesma eficácia, as imagens apresentavam pouca rotação. Para trabalhos futuros, sugere-se uma análise mais detalhada, analisando a relação da eficácia com diferentes faixas de graus de rotação.

5. Considerações finais

Este trabalho apresentou uma abordagem para extração de imagens com variação de iluminação e linhas de texto inclinadas, denominada *Toth*. A abordagem supera o baseline em termos de revocação, precisão e F1, realizando as extrações com maior eficiência. Dentre as contribuições realizadas durante o desenvolvimento desse trabalho, destacam-se:

1. Elaboração de três bases de dados com gabaritos para avaliar a eficácia da extração de softwares de OCR: (i) dataset ERBD - uma base de dados com imagens genéricas de trechos de livros, constituída de 160 imagens, das quais, 82 imagens (51.25%) foram coletadas manualmente de publicações de redes sociais e as 78 amostras restantes (48.75%) foram fornecidas por um grupo de voluntários; (ii) dataset BD_ROTACIONA - uma base de dados com 100 imagens rotacionadas para avaliar mecanismos de rotação de imagens com linhas de texto inclinadas; (iii) e dataset BD_ILUMINA - uma base de dados com 50 imagens com iluminação não homogênea, para avaliar mecanismos de correção de adversidades referente à variação de iluminação.
2. Análise da relação entre as características das imagens e eficácia da extração. Essa avaliação analisou a relação entre diferentes características presentes nas imagens e o quanto essas características afetam a eficácia da extração do software de OCR *Tesseract*. Esta contribuição está publicada no seguinte artigo:

KUHN, Daniel M.; CERVI, Cristiano R.; MANICA, Edimar. Extração de elementos textuais em imagens capturadas por smartphones : análise da relação entre as características das imagens e a eficácia da extração . **Escola Regional de Banco de Dados (ERBD)**, [S.l.], v. 14, n. 1/2018, apr. 2018.

3. Implementação da abordagem *Toth* com etapas para correção de imagens com variação de iluminação, bem como, imagens com linhas de texto inclinadas. A abordagem obteve um ganho de eficácia de 23,36 pontos percentuais em relação ao *Tesseract* em imagens com linhas de texto inclinadas. Para imagens com variação de iluminação, *Toth* obteve aumento de eficácia de 41,36 pontos percentuais em relação ao *Tesseract*. Em relação aos tempos de extração, *Toth* viabilizou extrações em média 55,38,% mais eficientes para imagens rotacionadas e 87,73% mais eficientes para imagens com variação de iluminação.

A abordagem *Toth* pode ser utilizada para aumentar significativamente a eficácia de aplicações que dependem de softwares de *OCR* para realizar extração de elementos textuais. Dentre os possíveis softwares que podem beneficiar-se da abordagem *Toth*, destacam-se:

1. Sistemas de recuperação de informação (RI): buscadores como Google e Bing⁶, que fazem uso de softwares de *OCR* para extrair elementos textuais em imagens presentes nas páginas web com o objetivo de indexá-los e aumentar a qualidade dos resultados apresentados aos usuários.
2. Sistemas de recomendação (SR): diversas aplicações utilizam algoritmos de recomendação para sugerir ao usuário itens que podem ser de seu interesse. Tais sugestões são baseadas no perfil do usuário, onde dados em relação ao seu comportamento, seu contexto e sua interação são coletados a todo o instante. O próprio Google possui um serviço de apresentar imagens para o usuário solicitando que ele participe do processo de identificação destas imagens. O propósito deste serviço é, no futuro, gerar recomendações para os usuários.

No decorrer das análises dos resultados obtidos nos experimentos, foram identificadas possíveis ampliações dos experimentos, as quais podem ser desenvolvidas em trabalhos futuros: (i) ampliar o número de características a serem analisadas; (ii) verificar a relação da eficácia da extração por intervalos de graus de inclinação; e (iii) analisar os casos específicos em que *Toth* obteve menor eficácia.

Referências

- E. Manica; C. F. Dorneles; R. Galante. (2017). R-Extractor: a method for data extraction from template-based entity-pages. In *Computer Software and Applications Conference (COMPSAC), IEEE 41st Annual*. IEEE. p. 778-787.
- A. Labrinidis, H. V. Jagadish. (2012). Challenges and opportunities with big data, *Proceedings of VLDB Endowment*, v. 5, n.12, pp. 2032-2033.
- D. Agrawal, P. Bernstein, E. Bertino, et. al. (2012). *Challenges and Opportunities with Big Data - A community white paper developed by leading researchers across the United States*.
- Statistic Brain. (2017). Instagram Company Statistics. Disponível em: brain <https://www.statisticbrain.com/instagram-company-statistics>. Acessado em: 15 de Janeiro de 2018.

⁶ Disponível em: <<https://bing.com>>. Acessado em: 23/10/2018.

- D. Berchmans; S. S. Kumar. (2014). Optical character recognition: An overview and an insight. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, Kanyakumari, pp. 1361-1365.
- N. Islam; Z. Islam; N. Noor. (2016). A Survey on Optical Character Recognition System. *Journal of Information & Communication Technology-JICT* Vol. 10 Issue.2.
- F. Asad et al. (2016) High Performance OCR for Camera-Captured Blurred Documents with LSTM Networks. In *Document Analysis Systems (DAS)*, 2016 12th IAPR Workshop on. IEEE. p. 7-12.
- J. Liang, D. Doermann, and H. Li. (2005). Camera-based analysis of text and documents: a survey. *International Journal on Document Analysis and Recognition (IJ DAR)*, v. 7, n. 2-3, pp. 84–104.
- D. M. Kuhn; C. R. Cervi; E. Manica (2018). Extração de elementos textuais em imagens capturadas por smartphones: análise da relação entre as características das imagens e a eficácia da extração. *Escola Regional de Banco de Dados (ERBD)*, [S.l.], v. 14, n. 1/2018.
- R. C. Gonzalez; R. E. Woods. (2000). *Processamento de imagens digitais*. Edgard Blucher.
- T. W. Anderson; J. D. Finn. (1996). *The New Statistical Analysis of Data*. New York, USA: Springer-Verlag.
- C. Olah. (2015). Understanding LSTM. Disponível em: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Acesso em: novembro de 2017.
- T. Kil; W. Seo; H. I. Koo and N. I. Cho. (2017). Robust Document Image Dewarping Method Using Text-Lines and Line Segments. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, 2017, pp. 865-870.
- R. W. Smith. (2017). *The Extraction and Recognition of Text from Multimedia Document Images*, PhD Thesis, University of Bristol, November 1987.
- Tesseract (2015). Tesseract. Disponível em: <https://github.com/tesseract-ocr/tesseract>. Acesso em: novembro de 2017.
- A. Gabasio. (2013). *Comparation of Optical Character Recognition (OCR) Software*. Department of Computer Science, Faculty of Engineering, LTH, Lund University, 2013.
- S. Dhiman; A. Singh. Tesseract vs gocr a comparative study. *International Journal of Recent Technology and Engineering*, v. 2, n. 4, p. 80, 2013.
- OpenCv. (2018). OpenCv Disponível em: <https://opencv.org>. Acessado em: novembro de 2017.
- R. Baeza-Yates; B. Ribeiro Neto. (2013). *Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca*. Porto Alegre: Bookman Editora.