



Alinhamento Automático de Imagens de Documentos

Thales Rogerio Sales Almeida, Hélio Pedrini

Resumo

Neste trabalho, realizamos uma extensa avaliação de diferentes técnicas de alinhamento de imagens, procurando maximizar a resposta de um posterior algoritmo de reconhecimento ótico de caracteres sobre uma imagem de documento. Vários experimentos foram conduzidos para validar nossos resultados.

Palavras chaves:

Alinhamento de imagens, OCR, pré-processamento, Transformadas de Fourier e Hough.

Motivação

Em inúmeras aplicações, algoritmos capazes de extrair o texto presente em uma imagem tornam-se necessários. Para realizar tal tarefa, algoritmos denominados OCR (*Optical Character Recognition*) são aplicados, os quais são baseados em técnicas de aprendizado de máquina e são treinados para reconhecer caracteres. Alguns exemplos de aplicações de OCR são apresentados em [1,2,3]. Atualmente, há técnicas robustas o bastante para reconhecer caracteres em inúmeras fontes diferentes e até mesmo manuscritos.

Para que os algoritmos de OCR obtenham o melhor resultado possível, um pré-processamento da imagem normalmente é necessário, em que a imagem deve ser binarizada apropriadamente, ou seja, deve conter apenas pixels pretos e brancos, e o texto deve estar adequadamente alinhado. Disto decorre a importância do alinhamento de texto, já que este é um fator crítico para que o algoritmo de OCR apresente um bom desempenho.

Bases de Dados

As seguintes técnicas de alinhamento foram implementadas e avaliadas:

- Projeção Horizontal
- Transformada de Hough
- Transformada de Fourier
- Projeção Horizontal em partições da imagem
- Transformada de Hough em partições da imagem
- Transformada de Fourier em partições da imagem
- Combinação estatística das técnicas já citadas.

Metodologia

A fim de testar a eficácia e a precisão das técnicas de alinhamento, foi utilizado um subconjunto de 719 imagens do repositório RVL-CDIP [4], o qual foi investigado detalhadamente em [5]. O conjunto selecionado é composto de documentos de inúmeras naturezas: documentos de texto apenas, documentos com tabelas, documentos com gráficos e partituras de música foram utilizadas para avaliar as técnicas de alinhamento de texto.

Uma rotação aleatória entre -180 e 180 graus foi aplicada a cada uma das imagens. O ângulo de rotação foi anotado para cada uma das imagens, possibilitando posteriormente a validação dos resultados das técnicas aplicadas.

Para avaliar os resultados, aplicamos uma técnica a cada uma das 719 imagens e, utilizando a resposta do método e o ângulo anotado previamente, calculamos o erro absoluto da técnica sobre uma determinada imagem utilizando a seguinte fórmula:

$$Er = \text{Erro absoluto} = |\theta_{\text{anotado}} - \theta_{\text{técnica}}|$$

Para alguns contextos, um alinhamento da imagem para alguma orientação múltipla de 90 graus pode ser bom o suficiente, uma vez que se a imagem estiver alinhada em tal orientação, seu ajuste é uma tarefa mais simples. Portanto, torna-se interessante julgar as técnicas não somente pelo seu erro absoluto, mas também o quão longe o resultado apresentado está de um múltiplo de 90. Podemos medir esse erro da seguinte forma:

$$\text{Erro relativo} = \min(\text{Er mod } 90, 90 - \text{Er mod } 90)$$

Note que o valor máximo para o erro relativo é 45 graus, que é a maior distância que qualquer ângulo tem de um múltiplo de 90.

Resultados

Categorizamos os erros relativos das técnicas nos seguintes grupos:

- Aceitável = erro \leq 2 graus.
- Erro Pequeno = 2 graus < erro \leq 5 graus.
- Erro Médio = 5 graus < erro \leq 10 graus.
- Erro Grande = erro > 10 graus.

Os três melhores resultados são mostrados nas Figuras 1, 2 e 3.

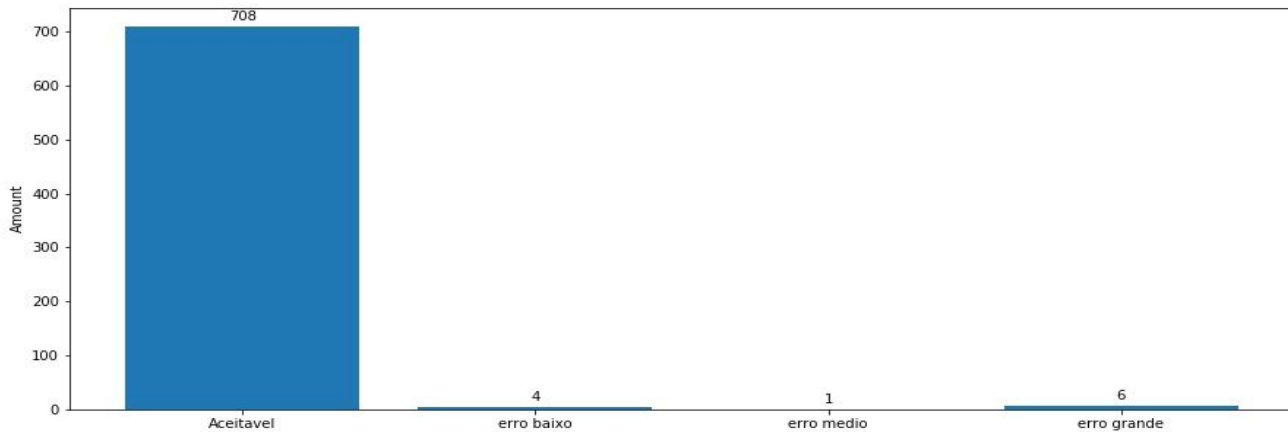


Figura 1. Categorização dos resultados da Projeção Horizontal.

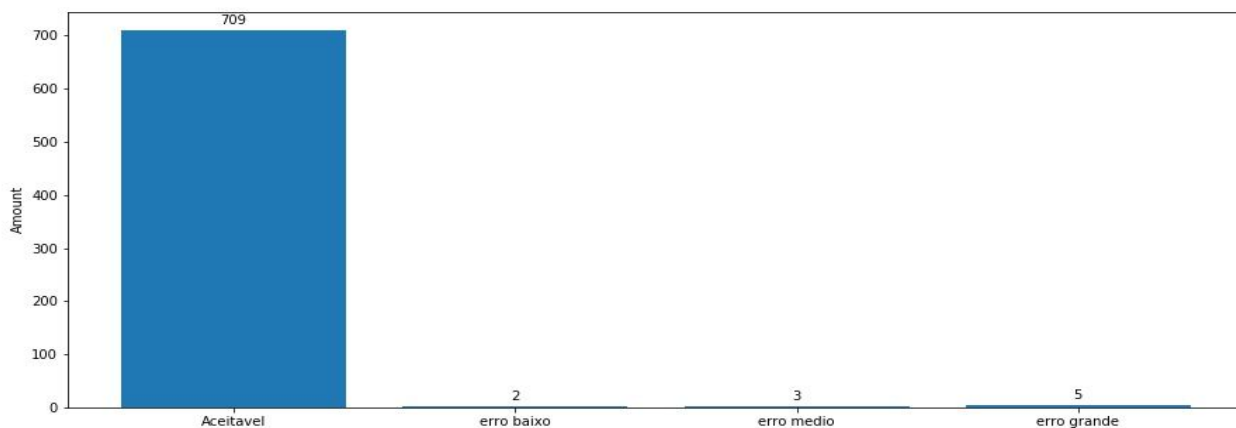


Figura 2. Categorização dos resultados da Transformada de Hough.

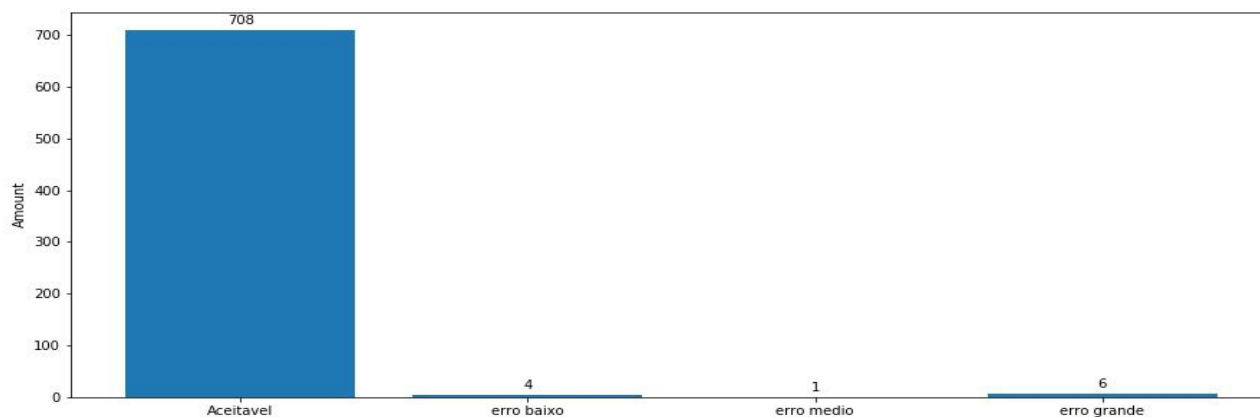


Figura 3. Categorização dos resultados da moda das técnicas.

Podemos verificar que as técnicas mostradas nas Figuras 1 a 3 obtiveram resultados excelentes nas imagens de teste, alcançando erros aceitáveis para mais do que 98% das imagens. Devemos, no entanto, lembrar que o conjunto de imagens utilizados nos testes foi limitado. Testes em uma escala muito maior podem apresentar resultados diferentes.

Vale destacar que das 3 técnicas apresentadas, a mais interessante para aplicação seria a transformada de Hough, uma vez que possui um custo computacional menor comparado às outras.

Interface e Código Fonte

Uma interface Web foi desenvolvida neste projeto, capaz de aplicar as técnicas apresentadas em uma imagem, sem a necessidade de qualquer configuração do usuário. A interface pode ser encontrada em [6] e pode ser vista na Figura 4.

A interface gráfica é recomendada para o alinhamento de pequenos grupos de imagens ou testes de pequenas proporções. O código e as imagens utilizadas nesta pesquisa estão disponíveis em [7].

You can use the forms below to align your image
Only white background documents are supported in the current version.
Using the method "BEST" will compute the result of all the other techniques and combine them, therefore the page might take a while to load.

Technique:

Image: No file selected.

This webpage is a interface for the results of the research "Alinhamento Automático de Imagens de Documentos" supported by the PIBIC/CNPQ
Student: Thales Rogerio Sales Almeida
Orientation: Prof. Hélio Pedrini
For more informations about the research and the algorithms used, please see the source code.

Figura 4. Interface Web de alinhamento de imagens.

Conclusão

As técnicas de alinhamento de texto apresentadas neste trabalho conseguem alinhar, de forma satisfatória, a maioria das imagens testadas (mais que 98% das imagens alinhadas com erros menores do que 2 graus). Os melhores resultados de técnicas individuais nas imagens testadas (719 imagens de documentos de naturezas diversas) foram obtidos pela transformada de Hough e pela projeção horizontal.

A melhor técnica a ser aplicada é fortemente dependente da aplicação em questão. A quantidade e a natureza das imagens de entrada, além das condições necessárias impostas na imagem de saída, são fatores fundamentais para a escolha da melhor técnica para cada situação. Por exemplo, caso um alinhamento a algum múltiplo de 90 graus seja suficiente (como no caso de alguns OCR), a projeção horizontal é recomendada.

Bibliografia

- [1] S.R. Aher, N.D. Kapale. "Automatic number plate recognition system for vehicle identification using optical character recognition". International Research Journal of Engineering and Technology (IRJET), pp. 2516-2519, Junho de 2017.
- [2] Amarjot Singh, Ketan Bacchuwar, Akshay Bhasin. "A Survey of OCR Applications". International Journal of Machine Learning and Computing, vol. 2, n. 3, pp. 314-318, Junho de 2012.
- [3] Sarfraz, M., Nawaz, S.N., Al-Khuraidly, A. "Offline Arabic text recognition system". International Conference on Geometric Modeling and Graphics. London, England, pp. 30-36, Julho de 2003.
- [4] RVL-CDIP Dataset, acessado em Outubro de 2019.
- [5] A.W. Harley, A. Ufkes, K.G. Derpanis. "Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval". International Conference on Document Analysis Recognition, pp. 1-5, 2015.
- [6] Web Interface for document image alignment, criada em Agosto de 2020.
- [7] Repositório com as imagens utilizadas e código fonte, criado em Agosto de 2020.