

Análise de Métodos para Correção de Orientação e Perspectiva em Imagens de Documentos

Gabriel Sichelero¹

¹Universidade Tecnológica Federal do Paraná UTFPR - Departamento Acadêmico de Informática
Av. Sete de Setembro 3165, Rebouças, Curitiba - PR

`gabriel.sichelero@hotmail.com`

Abstract. *A digitalização de documentos físicos frequentemente resulta em imagens desalinhadas ou distorcidas, o que prejudica a extração automática de informações. Este trabalho apresenta uma revisão bibliográfica sobre métodos para correção de distorções de orientação e perspectiva em imagens, além da análise de dois métodos para correção da orientação de imagens de páginas de documentos: um baseado em redes neurais convolucionais e outro utilizando a Transformada de Hough. O método com redes neurais convolucionais identifica os cantos da página e calcula a matriz de homografia para corrigir a orientação e distorções de perspectiva. Por outro lado, o método com a Transformada de Hough determina o ângulo predominante das linhas detectadas na imagem, corrigindo apenas a rotação.*

1. Introdução

Atualmente existe uma crescente necessidade de digitalizar documentos físicos para facilitar o armazenamento e a distribuição de informações. Este processo de digitalização normalmente envolve a captura de uma fotografia ou escaneamento do documento, resultando em uma imagem digital, a partir dessa imagem são extraídas as informações contidas no documento. No entanto, para melhorar a eficiência da extração dessas informações, é preferível que as páginas contidas nas imagens estejam alinhadas horizontalmente e verticalmente, na orientação que o documento foi escrito originalmente. Além disso, a presença de distorções de perspectiva nas imagens também pode dificultar a extração de informações.

Mais especificamente o problema de alinhamento das páginas de um documento consiste em encontrar a ângulação da imagem digitalizada em relação à orientação original do documento e aplicar uma rotação para corrigir essa ângulação, caso a imagem não contenha distorções de perspectiva.

Já no caso de distorções de perspectiva, o problema consiste em encontrar a matriz de homografia que mapeia os pontos da imagem para os pontos da página do documento com a orientação correta e aplicar essa matriz para corrigir a distorção.

Este estudo analisa os métodos existentes na bibliografia para abordar estes problemas. Além disso, foram implementados e analisados em profundidade dois métodos para a correção da orientação de imagens de páginas de documentos.

2. Revisão Bibliográfica

O alinhamento de imagens é um problema que vai além do escopo de documentos digitalizados, existem métodos para alinhamento de imagens em geral, como o RANSAC (Random Sample Consensus) [Shen et al. 2020], que faz o alinhamento de duas imagens através da identificação de características em comum entre elas e a aplicação de uma transformação geométrica para alinhá-las. A desvantagem deste método para o alinhamento das imagens de documentos é que nem sempre teremos a imagem do documento com a orientação correta para fazer a comparação.

Para o problema mais específico da correção da orientação de imagens de documentos existem abordagens mais tradicionais que utilizam métodos de processamento de imagens para extrair características das imagens, como a Transformada de Hough [Ahmad et al. 2021] ou a Transformada de Fourier [Fabrizio 2014] para a detecção de linhas paralelas e com isso determinar a orientação da imagem.

Além disso, existem métodos que utilizam redes neurais convolucionais treinadas com imagens de documentos com diversas orientações para tentar prever a diferença entre a orientação da imagem e a orientação correta do documento [Akhter and Rege 2020] [Aggarwal et al. 2021]. Esses métodos são treinados usando bases de dados que contêm imagens de documentos com a orientação distorcida, como o ICDAR 2013 Document Image Skew Estimation Contest (DISEC 2013) [Papandreou et al. 2013] e não necessitam de uma imagem de referência do documento com a orientação certa para fazer a correção.

Alguns dos métodos baseados em redes neurais convolucionais também são capazes de corrigir distorções de perspectiva nas imagens, como o método proposto por Abbas e Hussain [Abbas and ul Hussain 2017], que utiliza uma rede neural convolucional para detectar os cantos de uma página de um documento e a partir disso calcular a matriz de homografia que corrige a orientação da imagem.

3. Métodos

Para definir quais métodos irão ser analisados com mais profundidade e desenvolvidos vamos considerar os seguintes critérios: primeiramente, vamos considerar que não teremos a imagem do documento com a orientação correta para usar de referência e fazer a comparação, portanto métodos que necessitam de uma imagem de referência não serão considerados. Além disso, como existem bases de dados disponíveis com imagens de documentos com a orientação distorcida, podemos considerar métodos baseados em aprendizagem profunda que precisam ser treinadas com essas bases de dados.

Com base nesses critérios, foi implementado um método que utiliza redes neurais convolucionais [Körber 2019] para detectar os cantos de uma página de um documento e a partir disso calcular a matriz de homografia que corrige a orientação e perspectiva da imagem e também um método tradicional baseado na Transformada de Hough [Hinds et al. 1990] que calcula o ângulo das linhas detectadas na imagem para determinar a orientação correta da imagem.

4. Análise dos Métodos

Nesta seção, serão apresentados os métodos analisados para a correção da orientação de imagens de páginas de documentos.

4.1. Método Baseado em Redes Neurais Convolucionais

Primeiramente, iremos analisar o método baseado em redes neurais convolucionais [Körber 2019] que utiliza uma rede neural convolucional para detectar os quatro cantos de uma página de um documento e a partir disso calcular a matriz de homografia que corrige a orientação da imagem.

A precisão desse método foi medida com base no cálculo da diferença entre os valores das coordenadas horizontais e verticais dos cantos das páginas geradas após a correção da orientação e as coordenadas dos cantos das páginas originais.

As imagens usadas para treinamento e teste foram geradas sinteticamente, com páginas de documentos rotacionadas em diferentes ângulos e com diferentes mudanças de perspectiva.

Após a detecção dos cantos das páginas, a matriz de homografia é calculada utilizando os pontos dos cantos detectados e os pontos dos cantos do destino, normalizados para o tamanho das imagens utilizadas no treinamento do modelo.

Um exemplo do processo de detecção dos cantos e cálculo da matriz de homografia é mostrado a seguir considerando a seguinte imagem de entrada:



Figura 1. Imagem do documento de entrada

A rede neural convolucional detectou os seguintes cantos na imagem de entrada:

$$\text{Cantos Detectados: } \mathbf{C} = \begin{bmatrix} 46.391613 & 65.016945 \\ 216.54828 & 66.119965 \\ 241.43393 & 380.79517 \\ 8.454815 & 375.54947 \end{bmatrix}$$

Os cantos do destino são fixos e representam os cantos de uma página de um documento com a orientação correta:

$$\text{Cantos do Destino: } \mathbf{D} = \begin{bmatrix} 0.0 & 0.0 \\ 256.0 & 0.0 \\ 256.0 & 384.0 \\ 0.0 & 384.0 \end{bmatrix}$$

A matriz de homografia é calculada a partir dos cantos detectados e dos cantos do destino. Para isso, é necessário resolver o sistema de equações lineares $H \cdot \mathbf{C} = \mathbf{D}$, onde H é a matriz de homografia.

O resultado é a matriz de homografia que corrige a orientação da imagem:

$$\text{Matriz de Homografia: } H = \begin{bmatrix} 1.64712001 & 0.201223557 & -89.4954948 \\ -0.0118965475 & 1.83521367 & -118.768086 \\ 0.0000475566663 & 0.00129109745 & 1.00000000 \end{bmatrix}$$

O resultado pode ser observado comparando a imagem de entrada com a imagem corrigida:

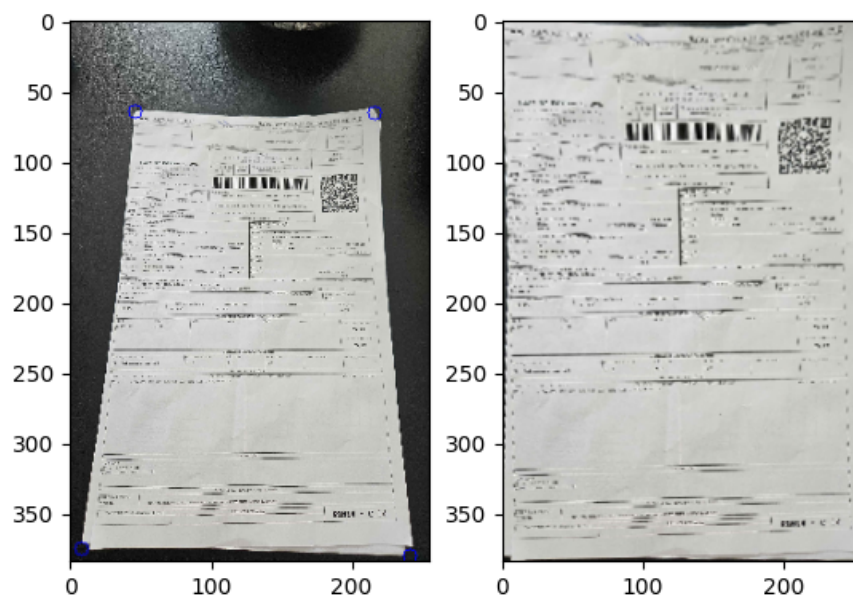


Figura 2. Comparação entre a imagem de entrada e a imagem corrigida

4.2. Método Baseado na Transformada de Hough

O outro método analisado foi baseado na Transformada de Hough [Hinds et al. 1990], que detecta linhas na imagem e calcula a inclinação média dessas linhas para determinar a orientação da imagem.

Um exemplo do processo de detecção de linhas e cálculo da inclinação média é mostrado a seguir. Primeiramente a imagem de entrada é convertida para escala de cinza e aplicado um filtro de Canny para detecção de bordas:

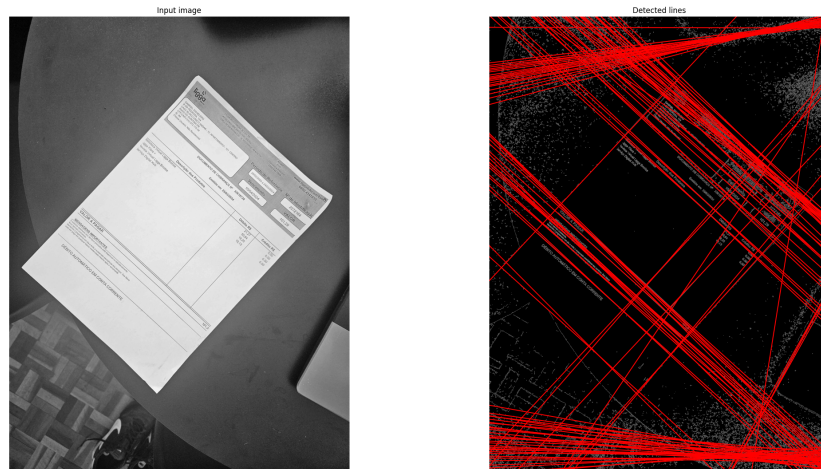


Figura 3. Imagem de entrada e linhas detectadas após a aplicação do filtro de Canny

Após a detecção das linhas, a orientação predominante da imagem é determinada a partir da moda dos ângulos das linhas detectadas, ou seja, o valor mais frequente entre os ângulos. Esse ângulo predominante é então ajustado subtraindo 90 graus, já que a Transformada de Hough mede os ângulos em relação ao eixo horizontal, resultando na inclinação da imagem.

$$\text{Inclinação Calculada } \theta = 42.35$$

A imagem é rotacionada em $-\theta$ graus para corrigir a orientação:

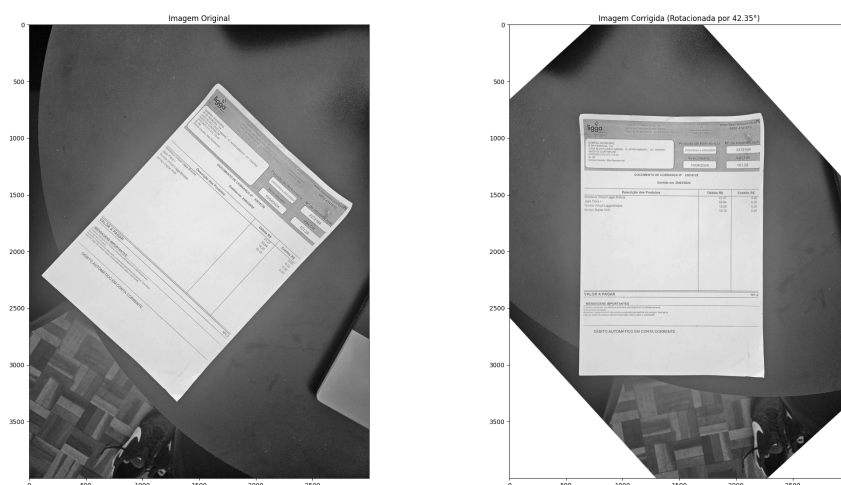


Figura 4. Imagem corrigida

4.3. Considerações sobre os Métodos

Ambos os métodos apresentaram resultados satisfatórios na correção da orientação das imagens de documentos. A grande vantagem do método baseado em redes neurais convolucionais é a capacidade de correção de imagens com distorções de perspectiva, enquanto o método baseado na Transformada de Hough apenas corrige a orientação da imagem, assumindo que a imagem não possui distorções de perspectiva.

Referências

- Abbas, S. A. and ul Hussain, S. (2017). Recovering homography from camera captured documents using convolutional neural networks. *ArXiv*, abs/1709.03524.
- Aggarwal, S., Gaur, S. S., and Manju (2021). Text document orientation detection using convolutional neural networks. In Sharma, H., Saraswat, M., Kumar, S., and Bansal, J. C., editors, *Intelligent Learning for Computer Vision*, pages 153–164, Singapore. Springer Singapore.
- Ahmad, R., Naz, S., and Razzak, I. (2021). Efficient skew detection and correction in scanned document images through clustering of probabilistic hough transforms. *Pattern Recognition Letters*, 152:93–99.
- Akhter, S. S. M. N. and Rege, P. P. (2020). Improving skew detection and correction in different document images using a deep learning approach. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6.
- Fabrizio, J. (2014). A precise skew estimation algorithm for document images using knn clustering and fourier transform. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2585–2588.
- Hinds, S., Fisher, J., and D’Amato, D. (1990). A document skew detection method using run-length encoding and the hough transform. In *[1990] Proceedings. 10th International Conference on Pattern Recognition*, volume i, pages 464–468 vol.1.
- Körber, N. (2019). Improving camera-based document analysis with deep learning. In *International Conference on Applied Informatics*.
- Papandreou, A., Gatos, B., Louloudis, G., and Stamatopoulos, N. (2013). Icdar 2013 document image skew estimation contest (disec 2013). In *2013 12th International Conference on Document Analysis and Recognition*, pages 1444–1448.
- Shen, X., Darmon, F., Efros, A. A., and Aubry, M. (2020). Ransac-flow: Generic two-stage image alignment. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, pages 618–637, Cham. Springer International Publishing.