UM MÉTODO PARA IDENTIFICAÇÃO E SEGMENTAÇÃO DE LOGOTIPOS E TEXTO EM REVISTA DO INPI

Braian Rodrigues Barbosa de Sousa *- braiansousa6@gmail.com*
Graduando em Análise e Desenvolvimento de Sistemas - Instituto Federal do Piauí
Igor Bezerra Reis *- igor.bezerra@ifpi.edu.br*
Instituto Federal do Piauí
Cícero Eduardo de Sousa Walter *- eduardowalter@ifpi.edu.br*
Instituto Federal do Piauí
Rafael Angelo dos Santos Leite *- rafaelangelo@ifpi.edu.br*
Instituto Federal do Piauí

Resumo—A segmentação de imagens em documentos desempenha um papel fundamental em várias aplicações relacionadas ao processamento de documentos. Essa técnica permite separar diferentes elementos presentes em uma imagem de documento, como texto, imagens e outros componentes estruturais com o objetivo de facilitar a busca de informações. Neste artigo, apresentamos uma metodologia para segmentação de imagens em documentos publicados pelo INPI, combinando diversas técnicas de processamento de imagens para extrair logotipos. O desempenho do método demonstrou uma acurácia de 96,86%.
Palavras-chave—Extração de logotipos, INPI, Processamento de imagens, Segmentação de imagens..

Abstract—Document image segmentation plays a key role in various applications related to document processing. This technique allows separating different elements present in a document image, such as text, images and other structural components with the aim of facilitating information search. In this paper, we present a methodology for image segmentation in documents published by INPI, combining several image processing techniques to extract logotypes. The performance of the method demonstrated an accuracy of 96.86%.
Keywords—Document image segmentation, Image processing, INPI, Logo extraction.

1 INTRODUÇÃO

A determinação da estrutura lógica de uma imagem de documento, como a classificação das regiões de texto e gráficos e a localização das regiões do título e do autor de um artigo, é uma etapa importante nos sistemas que analisam grandes quantidades de informações. O principal objetivo é segmentar regiões semanticamente semelhantes, como texto, gráficos, comentários, planos de fundo etc. (Jobin & Silva, 2018). As técnicas de análise de documentos vêm sendo estudadas há muitos anos, propondo métodos baseados na transformada de Hough (Fletcher and Kasturi, 1988), em segmentação baseada em região (Wong et al., 1982) ou baseados em texturas (Lin et al., 2006). Recentemente, algumas abordagens usando redes neurais também foram propostas para análise de documentos (Chiron et al., 2021; Busch et al., 2023).

A estrutura de um documento é geralmente representada por rótulos de componentes físicos baseados em um conjunto de regras (Mao et al., 2003). Aplicações potenciais de análise de estrutura de documento incluem pré-processamento para reconhecimento óptico de caracteres e imagens, por exemplo (Mao et al., 2003; Kim and Kim, 2001; Nagy,. 2000).

A publicação semanal das revistas de marcas pelo Instituto Nacional de Propriedade Intelectual (INPI) tem como finalidade disponibilizar marcas que solicitaram registro. Essas publicações são de extrema importância para escritórios de advocacia e outras organizações, pois permitem realizar análises das marcas a fim de identificar possíveis violações em relação às marcas monitoradas por cada escritório. Neste artigo, apresentamos um novo algoritmo de segmentação de páginas para revistas de marcas publicadas pelo INPI. O método proposto possui potencial para ser utilizado como entrada em algoritmos de classificação de imagens e medição de similaridade, facilitando a identificação de marcas similares ou com potencial de infringir direitos.

2 REVISÃO BIBLIOGRÁFICA

Várias metodologias que aproveitam as vantagens das Redes Neurais Profundas sobre imagens de documentos foram propostas na última década para o problema de classificação de imagens de documentos. Em Kang et al. (2014), uma arquitetura CNN básica é proposta para aprender recursos de pixels de imagem brutos, em vez de depender de recursos artesanais. Em Harley et al. (2015), foi mostrado que uma CNN pode extrair características robustas de diferentes partes da imagem (cabeçalho, rodapé esquerdo e corpo direito) e, treinando essas redes diferentes e usando-as em um esquema de conjunto, melhorias significativas na classificação ou na precisão de recuperação são alcançadas. Além disso, a redução do espaço de recursos usando a Análise de Componentes Principais (PCA) é importante, enquanto o desempenho não é afetado significativamente.

Na literatura, as abordagens de segmentação de texto/não-texto geralmente podem ser classificadas em três grupos: (i) segmentação baseada em região (ou bloco ou zona) (Wong et al., 1982; Okun et al., 1999), (ii) segmentação baseada em pixel (Moll and Baird, 2008b,a), e (iii) segmentação baseada em componentes conectados (Fletcher and Kasturi, 1988; Tombre et al., 2002; Bukhari et al., 2010). Em abordagens baseadas em regiões, uma etapa de segmentação de página na imagem do documento é realizada primeiro para obter as zonas do documento e, em seguida, um classificador é aplicado para classificar essas zonas (Shafait et al., 2008). Métodos desse tipo de abordagem dependem muito da precisão da etapa de segmentação, o que pode ser desafiador em documentos menos estruturados e de layout complexo. Nas abordagens baseadas em pixels, a classificação é aplicada em cada pixel em um documento. Os métodos que seguem esta abordagem tendem a ser sensíveis ao ruído e demorados. As abordagens baseadas em componentes conectados, por outro lado, são independentes da etapa de segmentação de bloco (zona) e são mais robustas devido a considerar o nível de componente conectado para discriminar entre texto e não texto.

A revisão feita por Okun et al. (1999) detalha as abordagens de segmentação de páginas e também de classificação de zonas, resumindo as principais abordagens utilizadas para segmentação de documentos e classificação de regiões na década de 1990. O Run Length Smearing Algorithm (RLSA), apresentado por Wong et al. (1982), é uma das primeiras abordagens usando segmentação baseada em região. O RLSA analisa os espaços entre os pixels pretos para mesclar os caracteres em blocos e, em seguida, usa uma técnica de análise para classificar cada bloco. Lin et al. (2006) propôs uma abordagem baseada na região. Eles usaram diferentes recursos GLCM (Matriz de Coocorrência de Nível de Cinza) baseados em textura

para dividir um documento em blocos de gráficos, texto e zonas de espaço, e usaram K-means para agrupar os blocos em zonas. Eles então usaram regras heurísticas pré-aprendidas para classificação de zona.

Na segmentação baseada em componentes conectados, existem algumas abordagens notáveis, como Fletcher e Kasturi (1988), que propuseram um método baseado na transformada de Hough para agrupar componentes conectados em uma string de texto e depois classificá-los usando o método de análise. Tombre et al. (2002) apresentam uma consolidação do método proposto por Fletcher e Kasturi, com algumas melhorias no método de análise.

3 METODOLOGIA

Este artigo descreve uma metodologia baseada em um fluxo de algoritmo para a análise de imagens. O método proposto utiliza uma abordagem de agrupamento de pixels brancos, onde as linhas de texto são identificadas como linhas verticais de pixels brancos, após a aplicação das etapas de binarização e dilatação. O fluxo consiste em quatro etapas principais, descrito na Figura 1:

Limiarização global: A imagem é convertida em uma forma binária, distinguindo o objeto de interesse do fundo.

Segmentação de bordas usando Canny: São detectadas as bordas na imagem, destacando as transições de intensidade.

Dilatação das bordas encontradas: As bordas são expandidas para melhorar a continuidade e preencher lacunas entre segmentos.

Criação de caixas de contorno: São delimitadas regiões de interesse com caixas que envolvem os objetos detectados.

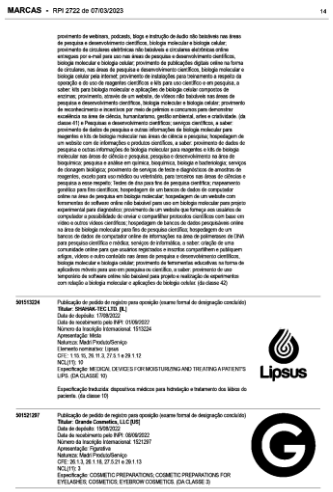
Figura 1. Fluxo do algoritmo

Fonte: Autores(2025)

3.1 Passo 1: Limiarização global

Inicia-se carregando as imagens em escala de cinza. Em seguida, é aplicado um processo de binarização para separar as regiões de interesse das demais. O método utilizado é o Global Thresholding (Sahoo et al., 1988), que consiste em definir um valor limite para separar os pixels em duas classes. No caso específico, o valor de Thresholding utilizado é de 200. Esse valor foi determinado experimentalmente e gera uma separação precisa das classes desejadas nas imagens do INPI.

Figura 2. Imagem após Threshold.



Fonte: Autores(2025)

3.2 Passo 2: Segmentação de bordas usando Canny

Para segmentar a imagem, utilizamos o método de Canny, um algoritmo clássico proposto por John F. Canny em 1986. O algoritmo de Canny utiliza um processo em várias etapas para detectar as bordas de uma imagem: suavização da imagem, cálculo dos gradientes, supressão de não-máximos, thresholding e histerese.

Figura 3. Imagem após segmentação com algoritmo Canny.

Fonte: Autores(2025)

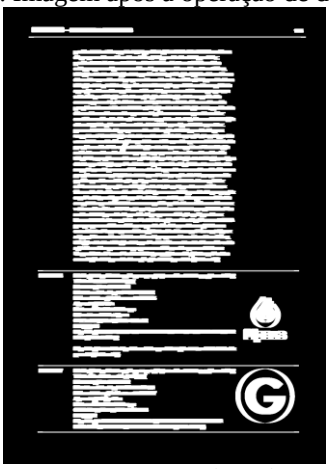
3.3 Passo 3: Dilatação das bordas encontradas

Na terceira etapa, aplica-se uma operação de dilatação na imagem. A dilatação é uma operação morfológica realizada através da convolução da imagem de entrada com um elemento estruturante (kernel). A formulação matemática da dilatação é dada por:

$$(A \oplus B)(x,y) = \bigvee_{(i,j) \in B} A(x+i, y+j)$$

Onde A é a imagem de entrada, B é o elemento estruturante, $\oplus$ denota a operação de dilatação e $\vee$ representa a operação lógica OR. Neste trabalho, utiliza-se um kernel de tamanho 11x11, expandindo as regiões brancas ao longo do eixo X para preencher lacunas.

Figura 4. Imagem após a operação de dilatação.



Fonte: Autores(2025)

3.4 Passo 4: Criação de caixas de contorno

A criação das caixas de contorno é essencial para definir os objetos que serão segmentados. A lógica proposta segue os seguintes passos:

Definição dos parâmetros: Define-se a largura e altura mínimas e máximas dos retângulos. Para logotipos, o tamanho mínimo é de 20x50.

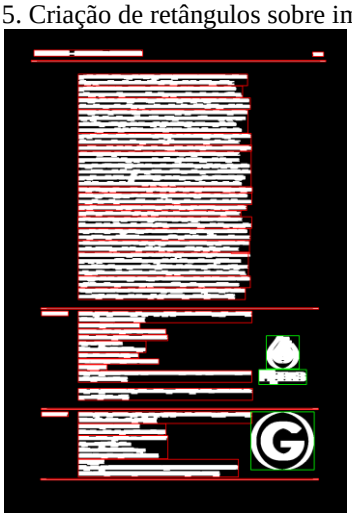
Identificação dos retângulos: O algoritmo percorre a imagem e identifica áreas de pixels brancos que correspondem às dimensões especificadas.

Expansão e mesclagem: Os retângulos identificados são expandidos e unidos para garantir a integridade e evitar sobreposição.

Pintura dos retângulos: Os retângulos que segmentam logotipos são marcados em verde, enquanto os que segmentam áreas de texto são marcados em vermelho.

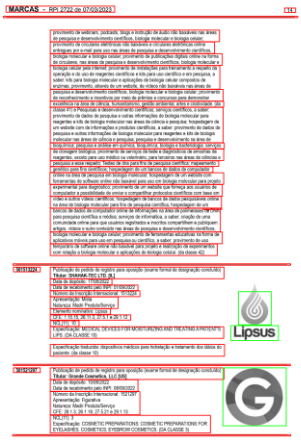
O resultado da marcação pode ser visto na Figura 5, e a aplicação das máscaras na imagem original, na Figura 6.

Figura 5. Criação de retângulos sobre imagem.



Fonte: Autores(2025)

Figura 6. Marcação na imagem original.

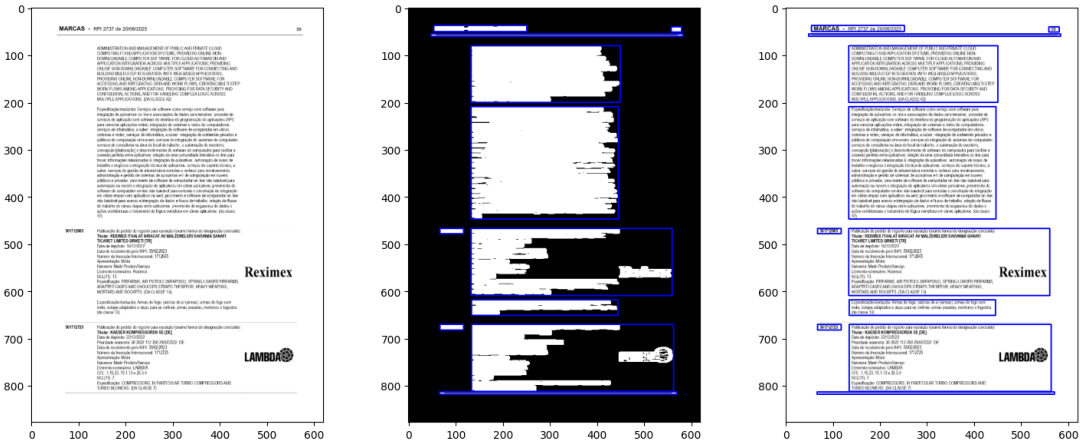


Fonte: Autores(2025)

4 RESULTADOS E DISCUSSÕES

Para testar o desempenho do algoritmo, foram selecionadas aleatoriamente 100 páginas da edição 2737 de uma revista publicada em 20 de junho de 2023. Dentro dessa seleção, havia um total de 223 marcas. O algoritmo foi capaz de classificar corretamente 216 dessas marcas, resultando em uma taxa de acurácia de aproximadamente 96,86%. A Figura 7 apresenta exemplos dessas classificações bem-sucedidas.

Figura 7. Exemplos de segmentação das páginas.



Fonte: Autores(2025)

Algumas classificações foram identificadas como incorretas. Acredita-se que esses erros tenham ocorrido devido a um possível contato entre as máscaras das regiões de texto e dos logotipos durante o processo de dilatação. Esse contato resultou em uma sobreposição indesejada, levando o algoritmo a uma classificação equivocada ao criar e mesclar os retângulos, como podemos observar na Figura 8.

5 CONCLUSÕES

Os resultados obtidos forneceram uma acurácia satisfatória de 96,86%. No entanto, foi identificado que alguns casos foram classificados de forma incorreta devido ao contato entre as máscaras de texto e de logos durante a etapa de dilatação.

Para futuras melhorias, sugere-se investigar técnicas de segmentação mais avançadas e refinadas, que possam lidar de maneira mais precisa com a complexidade dos elementos presentes nas páginas do INPI. Apesar dos desafios, os resultados alcançados são encorajadores e fornecem uma base sólida para o aprimoramento contínuo desta abordagem, que tem o potencial de otimizar o processo de pesquisa e análise de registros de marcas.

REFERÊNCIAS

Jobin, K.V., Jawahar, C.V. Segmentação de imagens de documentos usando recursos profundos. In: Rameshan, R., Arora, C., Dutta Roy, S. (eds) Visão Computacional, Reconhecimento de Padrões, Processamento de Imagens e Gráficos. NCVPRIPG 2017. Comunicações em Ciência da Computação e Informação, vol 841. Springer, Cingapura (2018).

Wahl, F.M., Wong, K.Y., Casey, R.G. Block segmentation and text extraction in mixed text/image documents. Comput. Graphics Image Process. 20 (1982).

Wang, D., Srihari, S.N. Classification of newspaper image blocks using texture analysis. Comput. Vision Graph. Image Process., 47 (1989), pp. 327-352.

Jain, A.K., Zhone, Y. Page segmentation using texture discrimination masks. Pattern Recognition, 29 (5) (1996), pp. 747-757.

Kang, L., Kumar, J., Ye, P., Li, Y., Doermann, D. Convolutional neural networks for document image classification. In: 22nd International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, pp. 3168–3172 (2014).

Harley, A.W., Ufkes, A., Derpanis, K.G. Evaluation of deep convolutional nets for document image classification and retrieval. In: 13th International Conference on Document Analysis and Recognition (ICDAR), Nancy, France, pp. 991–995 (2015).

Mao, S., Rosenfeld, A., Kanungo, T. Document structure analysis algorithms: a literature survey. Proceedings of SPIE Conference on Electrical Image, January, vol. 5010, SPIE (2003), pp. 197–207.

Kim, J., Le, D.X., Thoma, G.R. Automated labeling in document images. Proceedings of SPIE Conference on Document Recognition and Retrieval VIII, January (2001), pp. 111-122.

Nagy, G. Twenty years of document image analysis in PAMI. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (1) (2000), pp. 38-62.

Wong, K. Y., Casey, R. G., and Wahl, F. M. Document analysis system. IBM Journal of Research and Development, vol. 26, no. 6, pp. 647-656, 1982.

Okun, O., Dærmann, D., Pietikainen, M. Page segmentation and zone classification: the state of the art. DTIC Document Tech. Rep., 1999.

Moll, M., Baird, H., An, C. Truthing for pixel-accurate segmentation. In Document Analysis Systems 2008. DAS '08. The Eighth IAPR International Workshop on (pp. 379-385). Sept 2008.

Moll, M. A., Baird, H. S. Segmentation-based retrieval of document images from diverse collections. In Electronic Imaging 2008. International Society for Optics and Photonics (pp. 68150L-68150L). 2008.

Fletcher, L. A., Kasturi, R. A robust algorithm for text string separation from mixed text/graphics images. Pattern Analysis and Machine Intelligence IEEE Transactions on, vol. 10, no. 6, pp. 910-918, 1988.

Tombre, K., Tabbone, S., Pélissier, L., Lamiroy, B., Dosch, P. Text/graphics separation revisited. In Document Analysis Systems V, Springer (pp. 200-211), 2002.

Bukhari, S. S., Azawi, A., Ali, M. I., Shafait, F., Breuel, T. M. Document image segmentation using discriminative learning over connected components. In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems (pp. 183-190), 2010.

Shafait, F., Keysers, D., Breuel, T. Performance evaluation and benchmarking of six-page segmentation algorithms. Pattern Analysis and Machine Intelligence IEEE Transactions on, vol. 30, no. 6, pp. 941-954, June 2008.

Lin, M.-W., Tapamo, J.-R., and Ndovie, B. "A texture-based method for document segmentation and classification", South African Computer Journal, vol. 36, pp. 49-56, 2006.

Chiron, G., Arrestier, F., and Awal, A. M. "Fast End-to-End Deep Learning Identity Document Detection, Classification and Cropping," in Document Analysis and Recognition – ICDAR 2021, J. Lladós, D. Lopresti, and S. Uchida, Eds., 2021.

Busch, L., van Heusden, R., and Marx, M. "Using Deep-Learned Vector Representations for Page Stream Segmentation," Algorithms, vol. 16, no. 6, p. 259, 2023.

Sahoo, P.K., Soltani, S., and Wong, A.K.C. "A survey of thresholding techniques," Computer Vision, Graphics, and Image Processing, vol. 41, issue 2, pp. 233-260, 1988.