

Aplicação de análise morfológica para segmentação de páginas em imagens de documentos

Ricardo de Cillo (aluno) e Nina S. T. Hirata (orientadora)

Departamento de Ciência da Computação

Instituto de Matemática e Estatística

Universidade de São Paulo

I. INTRODUÇÃO

Uma das aplicações da teoria de visão computacional é a análise de imagens de documentos. A análise de imagens de documentos, ou apenas análise de documentos, é um campo de pesquisa atualmente ainda bastante ativo apesar de estar sendo explorado desde algumas décadas atrás [1]. Isto se deve a sua importância prática e a complexidade dos problemas abordados.

O objetivo em análise de documentos é extrair informações sobre o conteúdo e estrutura de um documento digitalizado. O diagrama 1, adaptado de [2], mostra os diferentes problemas abordados nesse campo de pesquisa.

Uma das etapas envolvidas na análise de layout é a segmentação de página que consiste na separação e classificação de regiões da imagem. Diferentes técnicas de processamento e análise de imagens têm sido propostos e aplicados para tratar esse problema. O objetivo deste trabalho é investigar a aplicação de operadores morfológicos [3] na segmentação de páginas. A qualidade da solução obtida será medida e comparada, segundo os mesmos critérios aplicados à resultados considerados estado da arte por pesquisadores da área [1].

II. METODOLOGIA

Uma grande diversidade de métodos já foram explorados na solução do problema de segmentação de páginas. Em [4] os autores propõem um método que observa a distribuição dos espaços em branco em um documento para classificar a região em texto ou não. Já em [5] extrai-se características dos pixels e sua vizinhança, classificando-os e posteriormente agrupando-os em regiões.

A. Treinamento de operadores morfológicos

Os operadores morfológicos [3] são bastante utilizados na área de visão computacional, para diferentes tipos de processamento de imagens. A construção de operadores morfológicos eficazes consiste em geral na combinação sequencial de operadores simples e pode ser uma tarefa difícil, além de demandar muita experiência. Portanto, neste trabalho propomos a construção de tais operadores, para a tarefa de segmentação de páginas, de forma automática a partir de imagens de treinamento, como descrito em [6].

A figura 2 ilustra o esquema geral de construção de um operador morfológico baseado em treinamento.

As imagens binárias definidas em um certo domínio E (geralmente $E = \mathbb{Z}^2$) podem ser modeladas por uma função $f : E \rightarrow \{0, 1\}$ tal que $f(x) = 1$ se e somente se x é um pixel correspondente a um objeto na imagem (portanto, $f(x) = 0$ se x é um pixel do fundo (*background*) da imagem).

O conjunto de todas as imagens binárias definidas em E é denotado por $\{0, 1\}^E$. Desta forma, um operador de imagens binárias é um mapeamento do tipo $\Psi : \{0, 1\}^E \rightarrow \{0, 1\}^E$.

Seja W uma janela de observação. Imagens podem ser processadas pixel a pixel, considerando-se a vizinhança de cada pixel definida pela janela W . Tais processamentos podem ser caracterizados por uma função do tipo $\psi : \{0, 1\}^W \rightarrow \{0, 1\}$, da seguinte forma

$$[\Psi(f)](x) = \psi(f_{-x}|_W), \quad (1)$$

na qual $f_{-x}|_W$ representa a imagem binária f restrita a W em torno de x .

Dada uma imagem f a ser processada e a respectiva imagem ideal I esperada como resultado do processamento, o erro de um operador Ψ é caracterizado por

$$MAE\langle\Psi\rangle = E[\Psi(f)(z) - I(z)], \quad (2)$$

na qual E denota o valor esperado. Supondo estacionaridade, o ponto z é arbitrário. Na prática o erro é calculado tomando-se a média do erro absoluto computado sobre todos os pixels da imagem.

Portanto, o problema de projetar operadores morfológicos localmente caracterizados reduz-se ao problema de projetar funções binárias do tipo $\psi : \{0, 1\}^W \rightarrow \{0, 1\}$. Dado que P é a distribuição conjunta do processo (\mathbf{X}, \mathbf{y}) (padrões observados pela janela W e respectivo valor da imagem de saída para o pixel considerado), pode-se mostrar que o operador ótimo em relação ao erro MAE é dado por

$$\psi(X) = \begin{cases} 1, & \text{se } P(X, 0) < P(X, 1), \\ 0, & \text{se } P(X, 0) > P(X, 1), \\ 1 \text{ ou } 0, & \text{if } P(X, 0) = P(X, 1). \end{cases} \quad (3)$$

Na prática, essas probabilidades não são conhecidas. Portanto, no processo de aprendizado de operadores, as mesmas são estimadas a partir de imagens de treinamento (pares de imagens entrada-saída, sendo que as imagens de saída em geral são geradas editando-se a imagem de entrada). A partir das probabilidades estimadas, pode-se obter a decisão ótima de acordo com a equação 3. No entanto, nem todos os padrões X

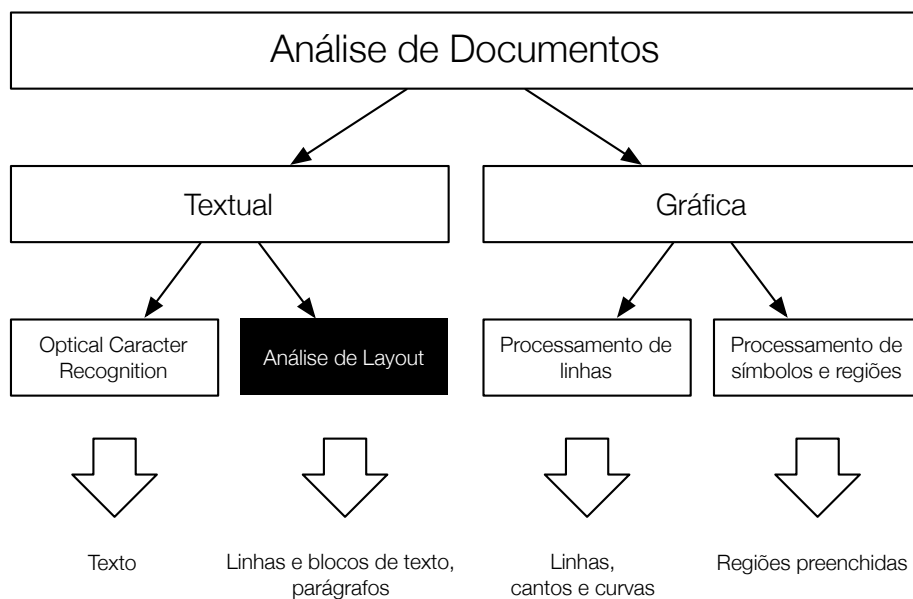


Figura 1. Contextualização do tema do trabalho entre as áreas da análise de documentos.

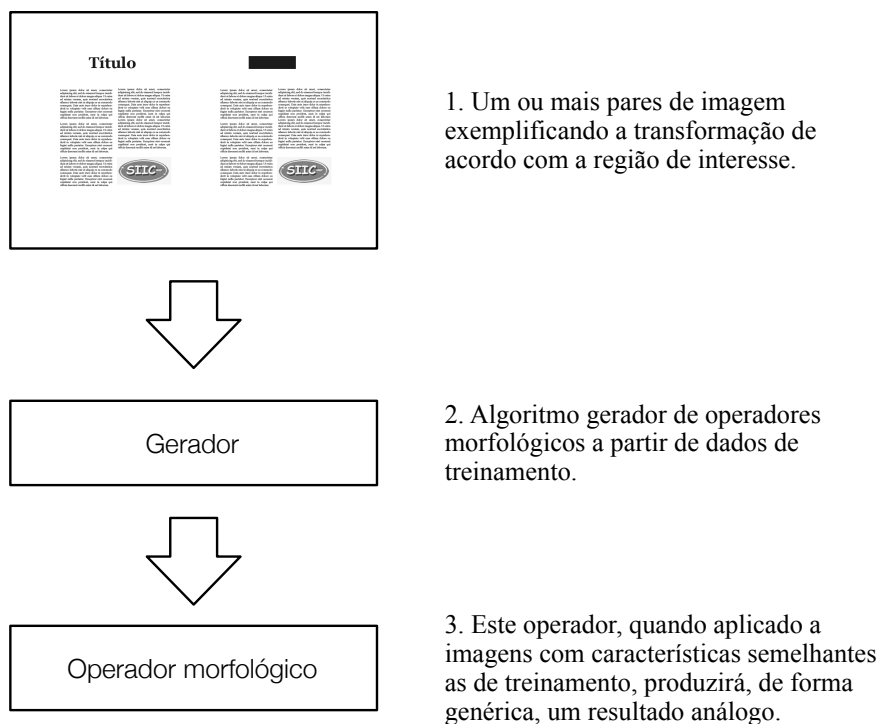


Figura 2. Visão global do funcionamento.

são observados nas imagens de treinamento. Portanto, utiliza-se um algoritmo de aprendizado para que a função característica do operador resultante fique completamente definida. Frequentemente, a esse processo de atribuir uma classificação para os padrões não pertencentes ao conjunto de treinamento

é denominado de generalização.

Diferentes algoritmos de aprendizado podem ser utilizados. Em [6], utiliza-se a minimização de funções booleanas não especificadas completamente. Maiores detalhes do método de treinamento podem ser encontrados em [7].

B. Avaliação da segmentação

A fim de comparar de forma objetiva o desempenho do método desenvolvido, aplicaremos aos resultados o avaliador automático utilizado na competição internacional de segmentação de página ICDAR2009 [8].

Este avaliador compara os resultados gerados pelo algoritmo a um resultado manualmente construído. A representação das regiões é feita com polígonos isotéticos, onde todos os ângulos possuem 90 graus. Esta geometria é flexível o suficiente para representar regiões complexas e não apenas áreas retangulares.

O processo de medição da acurácia leva em consideração as seguintes relações entre as áreas computadas e esperadas:

- intersecção nula,
- cobertura total da região esperada,
- duas regiões computadas cobrindo apenas uma região esperada,
- uma região computada cobrindo duas regiões esperadas, e
- região esperada esquecida.

Alguns dos erros descritos acima podem receber um peso maior dependendo do contexto. Agrupar regiões distintas cuja ordem de leitura seja sequencial, por exemplo, não é considerado um erro.

III. RESULTADOS ESPERADOS

Alguns resultados publicados atestam a viabilidade do treinamento de operadores morfológicos na segmentação de páginas [9], [7]. No entanto, nesses trabalhos a segmentação de páginas é apresentada apenas como um exemplo de possível aplicação, e em geral considerando páginas com estruturas bem controladas.

Neste trabalho, os experimentos estão sendo realizados com um conjunto de dados diversificados [10], que simula de forma mais realista uma situação de aplicação real e não apenas documentos estruturados de forma mais previsível e controlada. Para o treinamento do operador morfológico está sendo utilizado o pacote TRIOS (*TR*aining *I*mage *O*perators *f*rom *S*amples [11]), que é uma versão para Python do software utilizado em trabalhos anteriores.

Entre os principais desafios enfrentados encontra-se o treinamento de um operador morfológico que seja capaz de lidar com fontes de diferentes tipos e tamanhos. Está em investigação também a forma adequada para as imagens de saída para o treinamento: imagens consistindo apenas de textos da página de entrada, ou imagens consistindo do polígono isotético preenchido das regiões de texto.

Finalmente, os resultados obtidos serão comparados com os resultados esperados (*ground-truth*), segundo a métrica descrita acima.

REFERÊNCIAS

- [1] A. Antonacopoulos and D. Bridson. Performance analysis framework for layout analysis methods. *Document Analysis and Recognition, International Conference on*, 2:1258–1262, 2007.
- [2] Rangachar Kasturi, Lawrence O’Gorman, and Venu Govindaraju. Document image analysis: A primer. *Sadhana*, 27(1):3–22, 2002.
- [3] Jean Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.
- [4] A. Antonacopoulos and R. T. Ritchings. Representation and classification of complex-shaped printed regions using white tiles. In *Proceedings of the 3rd ICDAR*, pages 1132–1135. IEEE-CS Press, 1995.
- [5] Michael A. Moll and Henry S. Baird. Document content inventory and retrieval. In *In Proc., IAPR 9th Int’l Conf. on Document Analysis and Recognition (ICDAR2007)*, 2007.
- [6] Nina Sumiko Tomita. Programação automática de máquinas morfológicas binárias baseada em aprendizado pac. Master’s thesis, Departamento de Ciência da Computação, Instituto de Matemática e Estatística, April 1996. This work has been supported by ProTeM-CC/CNPq through the AnImoMat project, contract 680067/94-9.
- [7] N. S. T. Hirata. Morphological operator design from training data: A state of the art overview. In H. Kwasnicka and L. C. Jain, editors, *Innovations in Intelligent Image Analysis*, pages 31–58. Springer, 2011.
- [8] *10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009*. IEEE Computer Society, 2009.
- [9] N. S. T. Hirata, J. Barrera, and R. Terada. Text Segmentation by Automatically Designed Morphological Operators. In *Proc. of SIBGRAPI*, pages 284–291, 2000.
- [10] A. Antonacopoulos, D. Bridson, C. Papadopoulos, and S. Pletschacher. A realistic dataset for performance evaluation of document layout analysis, 2009.
- [11] Trios - training image operators from samples. June 2012.