

Aplicação de Operadores Morfológicos para Segmentação de Páginas em Imagens de Documentos

Aluno: Ricardo de Cillo Supervisora: Nina S. T. Hirata

Trabalho de Conclusão de Curso, Bach. Ciência da Computação, IME-USP, 2012

Introdução

Neste trabalho analisamos, experimentalmente, a aplicação de operadores morfológicos à segmentação de páginas, etapa importante na análise de documentos que busca extrair informações sobre a sua estrutura: regiões com títulos, figuras e parágrafos. A qualidade da solução obtida foi medida e comparada à solução ideal.

Segmentação de páginas de documentos

Segmentação de página refere-se à tarefa de separar e rotular os diferentes componentes que fazem parte da estrutura das páginas de um documento. Um exemplo de segmentação é apresentado na figura 1.

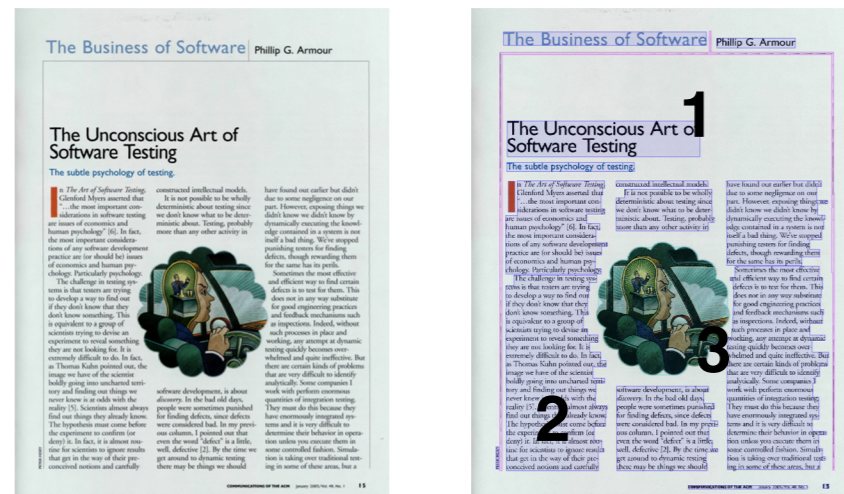


Figura 1: Exemplo de rotulação de componentes de uma página. Título (1), parágrafo (2) e figura (3).

Em geral, a segmentação de página é um dos primeiros passos no processo de entendimento de um documento. Uma vez identificados os blocos estruturais, processamentos específicos para cada tipo de bloco podem ser aplicados. Por exemplo, no caso de blocos de textos é conveniente fazer o reconhecimento de texto para que o mesmo possa ser armazenado em formato

texto (e não imagem). Documentos digitalizados podem ser processados eficientemente em processos que envolvem armazenamento, edição, transmissão, ou busca, por exemplo.

Operadores morfológicos como classificadores

Operadores morfológicos binários constituem uma classe de transformações sobre imagens que tem por objetivo analisar características geométricas da mesma. Neste trabalho aplicamos operadores morfológicos como classificadores que para cada pixel da imagem decidem se ele faz parte de uma dada região, apagando-o ou não. Como os operadores são binários, ou seja, permitem representar apenas duas classes, usamos um operador para cada região de interesse. Pós-processamentos consolidariam os resultados posteriormente.

Geração automática de operadores morfológicos

Ao trabalhar com problemas complexos como o de segmentação de página, a construção de um operador morfológico através da combinação de diferentes operadores mais simples torna-se uma tarefa muito custosa requerindo conhecimento específico e experiência. Desta forma utilizamos uma técnica de aprendizado computacional PAC para construção automática de operadores [1] implementada pela biblioteca TRIOS [2]. A seguir resumimos as etapas deste processo.

1. Entrada: pares de imagens exemplificando a aplicação do operador que se deseja obter. No caso deste trabalho utilizamos um dataset do PRImA [3]. A figura 2 contém um exemplo de entrada.

2. Coleta de exemplos: passa-se a janela pelas imagens original e ideal contando a frequência com que cada padrão é transformado em zero ou um.

3. Decisão: com base na contagem feita no passo anterior, minimiza-se o erro definindo como correto o padrão de maior ocorrência.

4. Minimização: A fim de otimizar a aplicação dos operadores gerados, utilizamos o algoritmo ISI [1] de minimização de funções booleanas.

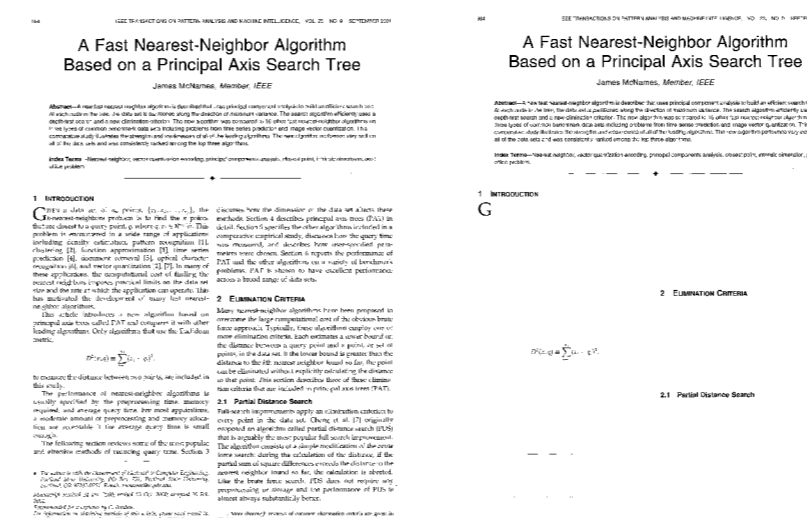


Figura 2: Par de imagens de entrada. À esquerda, imagem original e à direita imagem com texto dos parágrafos apagado.

Experimentos

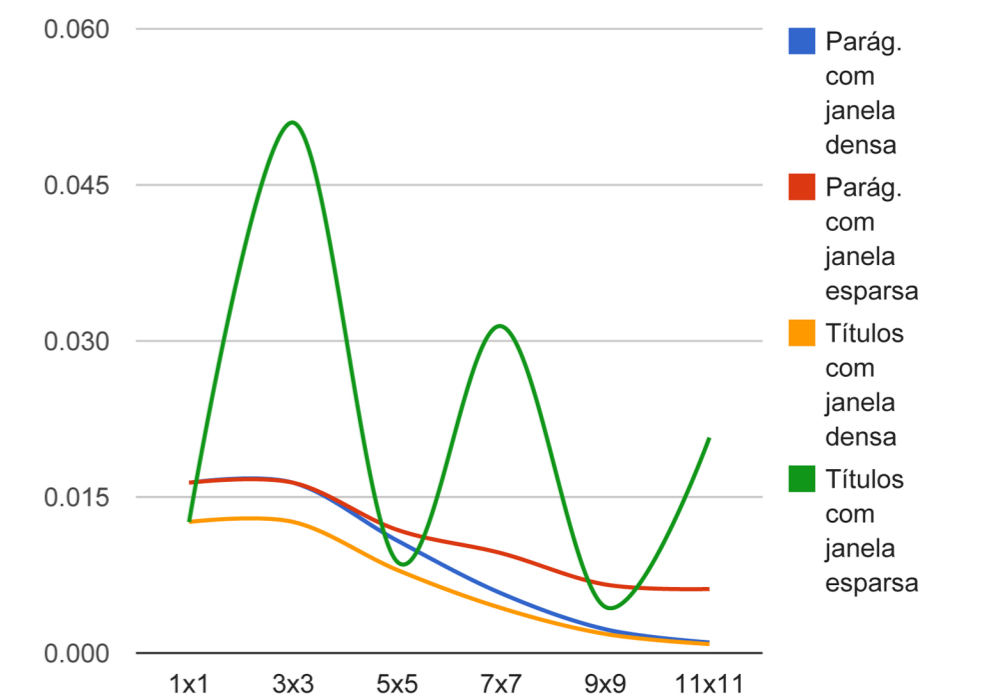
Realizamos experimentos utilizando um subconjunto do dataset da competição ICDAR2009 [4]. Analisamos a sensibilidade à mudança das seguintes variáveis na qualidade do resultado final: tamanho da janela, forma da janela e quantidade de exemplos.



Figura 3: Exemplo de aplicação de operador automaticamente gerado para identificação de títulos. Da esquerda para a direita: imagem original, aplicação de operador com janela esparsa 5x5 e 7x7.

Resultados

Na tabela abaixo apresentamos a média de erros normalizada por milhares de pixels, ou seja, quantos pixels em média foram modificados para valores errados.



Conclusão

Ajustar o tamanho da janela é computacionalmente custoso porém muito importante para conseguir bons resultados. Janelas muito pequenas costumam apresentar taxas de erro mais altas pois ocorrem muitos conflitos na etapa de contagem de frequência dos padrões. Já as janelas muito grandes sofrem com as lacunas de padrões que nunca ocorrem nos exemplos. Janelas esparsas são de computação mais rápida porém apresentam resultados instáveis, como podemos observar no gráfico oscilante.

Referências

[1] 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.

[2] Trios - training image operators from samples. <https://sourceforge.net/projects/trioslib/>, Junho 2012.

[3] Pattern Recognition & Image Analysis Research - <http://www.primaresearch.org>.

[4] 10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009. IEEE Computer Society, 2009.