

**Introdução**

A quantidade de documentos em papel é, atualmente, muito grande e uma tarefa comum é a conversão manual das informações contidas neles para formatos eletrônicos convenientes. Existe um esforço de pesquisa em busca da automatização destas conversões. Este trabalho apresentará um método de análise e conversão de um tipo específico de documento: tabloides promocionais distribuídos em varejo.



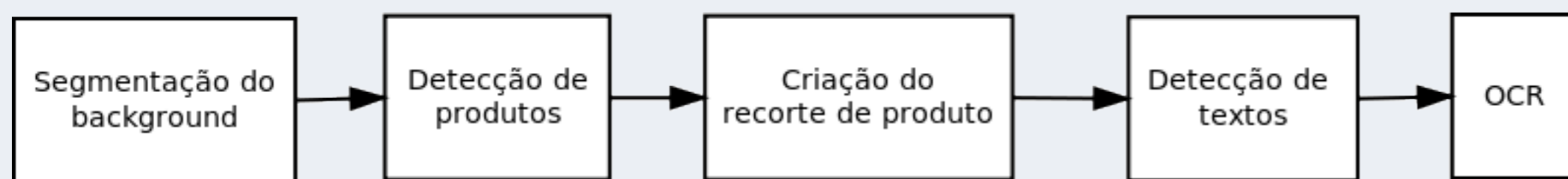
Figura: Exemplo de tabloide promocional analisado

**Objetivos**

Os objetivos do trabalhos são construir um sistema que:

- produza uma listagem com os preços de cada produto encontrado nas páginas de um tabloide;
- torne viável a análise de tabloides de diferentes estabelecimentos comerciais.

Para poder lidar com a constante modificação de *layout* que ocorre nos tabloides, o trabalho de análise foi dividido nas seguintes etapas:

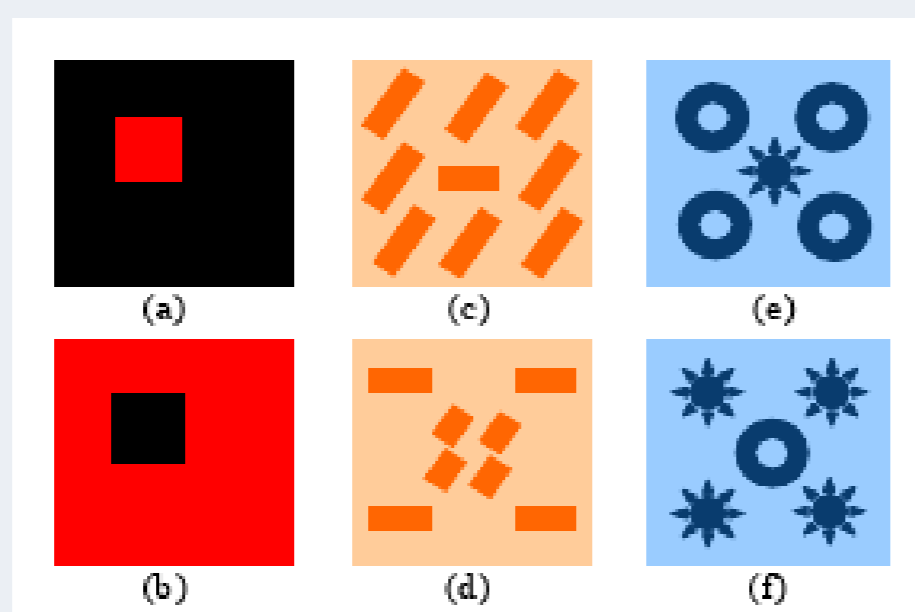


Esta divisão permite adicionar suporte a um novo tipo de tabloide sobrescrevendo as etapas que falham se usado um analisador para outro tipo.

**Atenção visual e mapa de saliências**

É natural que um leitor, ao observar uma imagem, examine algumas partes primeiramente e com maior atenção que outras. Em tabloides promocionais este fato é mais relevante, já que não existe uma ordem natural de leitura. Este fenômeno, chamado de *atenção seletiva*, é descrito em [1], juntamente com sistemas de computação que o utilizam para analisar imagens.

No trabalho [2], o contraste é usado como medida principal de atenção. A figura abaixo, retirada deste trabalho, procura evidenciar a importância do contraste na atenção visual.



Na primeira coluna, em ambas as imagens o elemento de maior atenção é o quadrado pequeno, o que indica que a cor não é um elemento central na atenção visual. Na segunda, a orientação das barrinhas não é o fator mais decisivo, mas sim a diferença nas orientações. O mesmo ocorre na terceira coluna, mas com o formato do elementos. Em todas as imagens o contraste entre cores, formas e orientações parece mais relevante do a cor, forma ou orientação por si só.

Um mapa de saliências é usado para medir a atenção visual. Cada ponto  $C_{ij}$  deste mapa é definido segundo a equação (1), sendo que cada pixel da imagem original corresponde a um ponto no mapa de saliências.

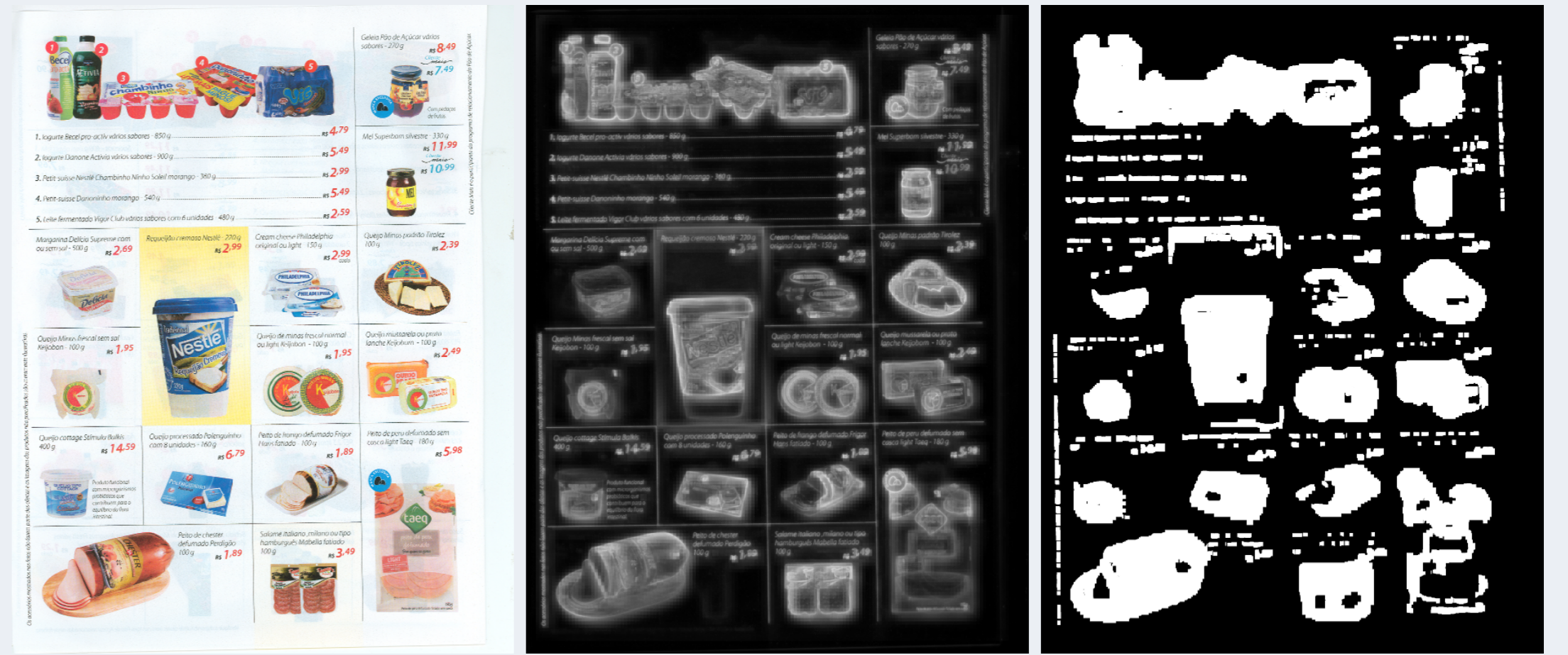
$$C_{ij} = \sum_{q \in \theta} d(p_{ij}, q) \tag{1}$$

onde  $p_{ij}$  e  $q$  são a quantidade de estímulo recebida,  $\theta$  é o tamanho da vizinhança analisada e  $d$  é a Distância Gaussiana [3] entre  $p_{ij}$  e  $q$  (definida na equação (2), onde  $d_e$  é a distância euclidiana entre duas cores no espaço RGB e  $\sigma$  é o desvio padrão da gaussiana usada na distância):

$$d(p, q) = 1 - e^{-d_e(p,q)/2\sigma^2} \tag{2}$$

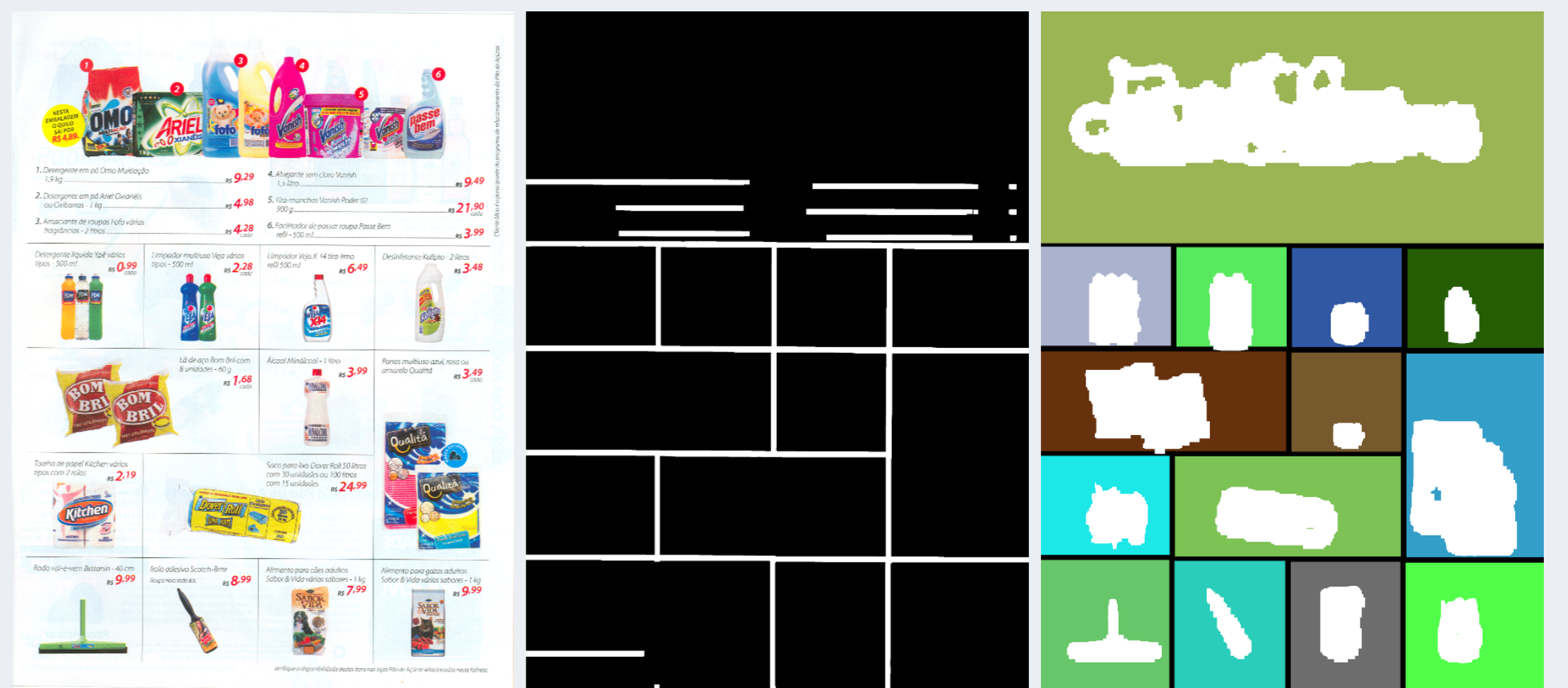
**Segmentação do background**

O mapa de saliências da imagem é calculado e tudo o que não for foco de atenção na imagem será considerado como *background*. Veja a imagem original, seu mapa de saliências e a segmentação final.



**Detecção de produtos e definição do recorte**

A partir do fundo segmentado, é possível usar o tamanho de cada região para detectar produtos. Para realizar o recorte de cada produto, aplica-se um filtro sobre o mapa de saliências para obter as linhas divisórias e toma-se a célula que contém cada produto como seu recorte. Veja abaixo.



**Detecção de textos e OCR**

A detecção de textos dentro de um recorte foi feita usando *classificação supervisionada* com o classificador *k-Vizinhos mais próximos*. Foram extraídas 5 características e usadas 3 classes: *nome*, *preço* e *qualquer outra coisa*. Abaixo está uma figura que ilustra o processo de classificação.

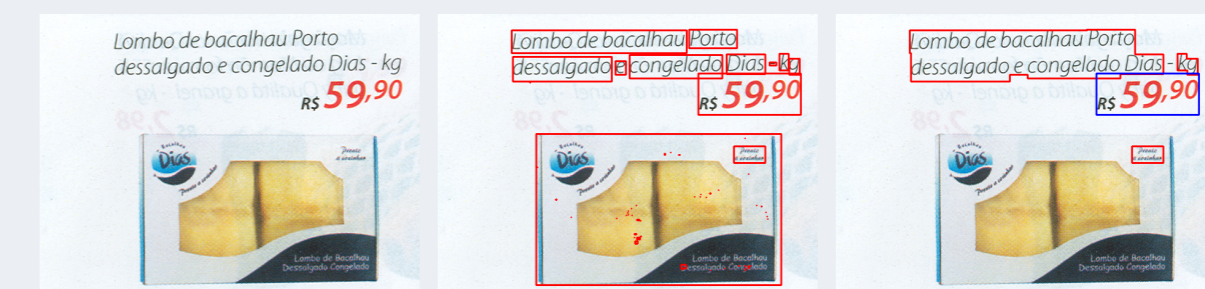
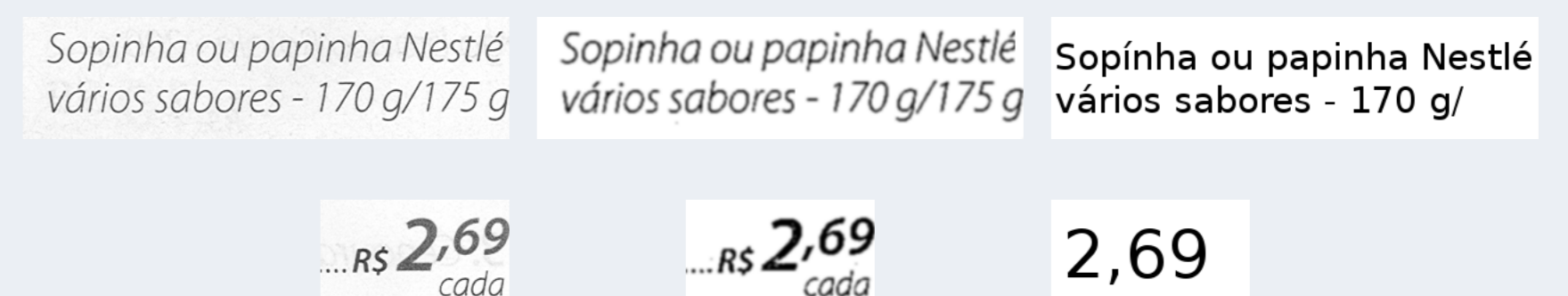


Figura: Recorte de um produto, as regiões a serem classificadas e o resultado final da detecção de nomes de produtos (em vermelho) e preços (em azul)

A leitura dos caracteres foi feita usando o *Tesseract OCR*. Cada região de texto é pré-processada antes da leitura dos caracteres. Abaixo um exemplo da imagem original, a imagem tratada e o resultado da leitura.



**Conclusões**

O trabalho mostrou a viabilidade da construção de um sistema analisador de tabloides promocionais e, com as etapas definidas, é possível expandir o programa para outros *layouts*. A teoria de *atenção visual* foi essencial para a realização do trabalho e o mapa de saliências baseado em contrastes foi usado em todas as etapas do trabalho.

**Agradecimentos**

Agradeço à Diretoria de Inovação do Ibope Media pelo apoio ao trabalho e à Celina Takemura pela ajuda nas soluções do trabalho e na elaboração do texto.

**Referências**

[1] S. Frintrop, E. Rome, H. I. Christensen. *Computational visual attention systems and their cognitive foundations: A survey*. ACM Trans. Appl. Percept. 7, 1 (Jan. 2010), 1-39. 2010.  
[2] Y. Ma, H. Zhang. *Contrast-based image attention analysis by using fuzzy growing*. Proceedings of the Eleventh ACM international Conference on Multimedia Berkeley, CA, USA. 2003.  
[3] H. Liu, S. Jiang, Q. Huang, C. Xu, W. Gao *Region-based visual attention analysis with its application in image browsing on small displays*. In Proceedings of the 15th international Conference on Multimedia (Augsburg, Germany, September 25 - 29, 2007). 2007.