

Análise e conversão de tablóides de promoções

Igor dos Santos Montagner
Orientador: Roberto Marcondes Cesar Junior

20/09/2010

1 Introdução

Com os avanços na área de computação e a demanda crescente por informação, é evidente a necessidade de buscar, indexar e armazenar documentos e informações de forma eficiente. Documentos em formato eletrônico podem suprir melhor estas necessidades, pois possuem vantagens significativas em relação ao papel. Mesmo assim, a quantidade de documentos em papel usadas atualmente é muito grande e uma tarefa muito comum em algumas áreas, como análise de mercado ou pesquisas de opinião, é a conversão manual das informações contidas em papel para formatos eletrônicos que facilitem a tarefa de análise. Sendo assim, existe um esforço de pesquisa em busca de uma técnica capaz de automatizar esta conversão de forma robusta. Neste trabalho trataremos da conversão e análise de um tipo específico de documento: tablóides de promoções distribuídos em varejo.

1.1 Motivação

A leitura de tablóides de promoções possui uma característica peculiar: não existe uma ordem explícita de leitura de uma página, diferentemente de outros documentos semiestruturados ou que se assemelha com outros documentos do mesmo tipo, como artigos científicos, revistas e jornais (trabalhos como [5] apresentam modelos para a determinação da ordem de leitura de uma página). Em um documento que não possui ordem de leitura, é natural que algumas partes sejam examinadas primeiro, enquanto outras recebem atenção mais periférica. Esta seleção de quais partes serão melhor examinadas é chamada de *atenção seletiva* e é descrita em [4] juntamente com diversos sistemas de computação que usam *atenção visual* para analisar imagens e selecionar regiões de interesse. Outros trabalhos, como [1, 2, 3], também descrevem abordagens para análise de imagens baseada em *atenção visual*. Neste trabalho os conceitos apresentados em [4] utilizados para a construção de um sistema para a análise de tablóides.

Para o consumidor, o trabalho ajudaria a encontrar os produtos em promoção em cada estabelecimento e a comparar estabelecimentos para a decisão do local com melhores preços.

1.2 Objetivos

O trabalho de leitura dos tablóides foi dividido em duas partes: *segmentação e detecção de textos e OCR (leitura ótica de caracteres)*. A etapa de *segmentação* corresponde a, dada uma página de tablóide, identificar os produtos nela presentes e fazer um recorte que contenha a imagem do produto, seu

preço e nome. A *etapa de detecção de textos e OCR* compreende a localização do nome do produto e do preço dentro de um recorte obtido na segmentação e o uso de um programa de OCR para a obtenção da informação textual contida na imagem. Os objetivos do trabalho são:

1. Produzir uma listagem com os preços de cada produto encontrado nas páginas de um tablóide;
2. Construir um sistema que torne viável a análise de tablóides de diferentes comércios.

1.3 Organização do trabalho

A seção 2 descreve a etapa de segmentação de uma página de um tablóide. A seção 3 descreve a etapa de detecção de textos e uso de um programa de OCR. A seção 4 trata de uma arquitetura que torne o sistema robusto e expansível para leitura de tablóides em diversos formatos. Na seção 5 são apresentados os resultados obtidos e na seção 6 são explicitadas as conclusões finais do trabalho, assim como uma avaliação do cumprimento dos objetivos propostos. Na seção 7 é apresentada uma análise subjetiva do curso do Bacharelado em Ciência da Computação.

2 Segmentação dos tablóides

Para uma segmentação da página estar correta, é necessário que todos os produtos, juntamente com seu nome e preço, apareçam em pelo menos uma divisão. As divisões identificadas não precisam formar uma partição da imagem. Esta definição inclui os *Grupos de produtos*, que são um conjunto de vários produtos muito próximos um do outro e cujos nomes e preços estão associados ao grupo todo e não a um produto em específico. Veja a figura 1.

As seguintes etapas são realizadas para obter uma segmentação correta: segmentação do background, detecção dos produtos da página e identificação de divisões entre produtos.

2.1 Segmentação do background e identificação de produtos

A primeira etapa para a segmentação do tablóide é a segmentação do background e identificação dos produtos presentes em cada página. A abordagem descrita nesta seção tem uma restrição importante: *em qualquer imagem analisada, um leitor humano deveria considerar os produtos da imagem como*



Figure 1: 1(a) e 1(b) representam divisões válidas e 1(c) é um *Grupo de produtos*

maiores focos de atenção. Em outras palavras: quando uma pessoa comum lê o tablóide pela primeira vez, seu olhar se dirige primeiramente para os produtos e depois para os outros elementos da página, como preços, nomes de produtos e detalhes do plano de fundo.

Em alguns trabalhos, por exemplo [1, 2, 3, 4], o contraste, seja entre pixels individuais ou entre regiões da imagem, é citado como um componente importante para identificar quais áreas da imagem são possíveis focos de atenção do leitor e em [1] é usado como único parâmetro para este fim.

Em [1], uma imagem de tamanho $M \times N$ pode ser tratada como um *campo perceptivo* composto por $M \times N$ *unidades de percepção*, se cada *unidade de percepção* contiver um pixel. O contraste C_{ij} ($i \in [0, M], j \in [0, N]$) da *unidade de percepção* (i, j) é definido a seguir:

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

Aonde p_{ij} e q são a quantidade de estímulo recebida pelas *unidades de percepção*, θ é o tamanho da vizinhança analisada e d é distância entre p_{ij} e q e pode variar conforme a aplicação. Neste trabalho usamos a Distância

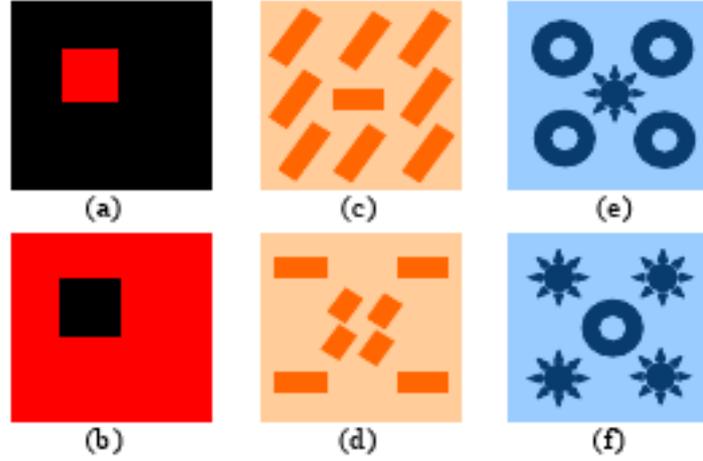


Figure 2: Esta figura, retirada de [1], demonstra que o contraste parece ter mais importância sobre outras características como formato, orientação ou cor para a identificação de regiões de interesse

Gaussiana [3], que é definida pela seguinte fórmula, aonde d_e é a distância entre duas cores no espaço RGB e σ é o desvio padrão da gaussiana usada na distância:

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

Em [1] a mesma distância é usada, porém com cores no espaço LUV. Outros tipos de estímulo podem ser usados, dependendo do resultado desejado.

Normalizando para $[0, 255]$, os contrastes $C_{i,j}$ formam um *mapa de saliências*. É por meio da análise deste mapa que é feita a detecção dos produtos, que segundo a restrição apresentada no começo da seção, devem possuir os maiores valores no *mapa de saliências*.

Para diminuir o ruído e fazer uma análise multiescala na imagem usamos uma sequência de imagens em que, dado um elemento i , o elemento $i + 1$ possui a metade das dimensões da imagem anterior. O nível 0 é formado pela imagem nas dimensões originais. Neste trabalho a sequência de imagens contém 3 elementos. Para cada elemento $i \in [0, 2]$, um mapa de saliências MS_i usando o método acima com $\theta = 2$ e $\sigma = 10$ é construído e o mapa de saliências final MS_f é dado pela seguinte fórmula:

$$MS_{f_{ij}} = \frac{MS_{0_{ij}} + 2\left(\frac{MS_{1_{kl}} + 6MS_{2_{mn}}}{7}\right)}{3}, k = i/2, l = j/2, m = i/2, n = j/2 \quad (3)$$

- 3 Detecção de textos e OCR
- 4 Arquitetura de um sistema de leitura de tablóides
- 5 Análise de resultados
- 6 Conclusões
- 7 Análise subjetiva

References

- [1] Ma, Y. and Zhang, H. *Contrast-based image attention analysis by using fuzzy growing*. Proceedings of the Eleventh ACM international Conference on Multimedia Berkeley, CA, USA. 2003.
- [2] Laurent Itti, Christof Koch. *A saliency-based search mechanism for overt and covert shifts of visual attention* Vision Research, Volume 40, Issues 10-12, Pages 1489-1506, ISSN 0042-6989, DOI: 10.1016/S0042-6989(99)00163-7. 2000.
- [3] Liu, H., Jiang, S., Huang, Q., Xu, C., and Gao, W. *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.
- [4] Frintrop, S., Rome, E., and Christensen, H. I. *Computational visual attention systems and their cognitive foundations: A survey*. ACM Trans. Appl. Percept. 7, 1 (Jan. 2010), 1-39. 2010.
- [5] T. M. Breuel. *High performance document layout analysis*. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.13032003>