

## Uma Técnica Eficiente para Segmentação Baseada no Algoritmo de Médias-K Localmente Aplicado nas Estatísticas da Imagem

Alexandre P. Mendonça  
*Departamento de Engenharia Elétrica  
Instituto Militar de Engenharia &  
PEE/COPPE/DEL/UF RJ  
Rio de Janeiro, RJ, Brasil  
alexmend@aquarius.ime.eb.br*

Eduardo A. B. da Silva  
*Programa de Engenharia Elétrica  
COPPE/DEL/UF RJ  
Rio de Janeiro, RJ, Brasil  
Cx. Postal 68504, CEP 21945-970  
eduardo@lps.ufrj.br*

### RESUMO

Neste artigo, descrevemos uma técnica eficiente de segmentação baseada na aplicação local do algoritmo de Médias-K sobre um espaço de vetores de características estatísticas da imagem. Tal técnica requer um pequeno esforço computacional e pode ser utilizada para diferentes classes de imagens, oferecendo resultados favoráveis se comparado com outros algoritmos da literatura.

### 1. INTRODUÇÃO

A importância das técnicas de segmentação não pode ser subenfatizada. Como exemplos de aplicações, citam-se a visão computacional e a compressão de imagens. Também o padrão MPEG-4 [1], que permite a divisão da imagem em planos de objetos, vem incentivando o desenvolvimento de trabalhos associados à segmentação de imagens e vídeo.

Os mais recentes e populares algoritmos abordam de forma diferente o problema da segmentação. Muitos deles concentram-se em detetores de bordas, contornos e retas. Ji e Haralick [2] propuseram um eficiente detetor de retas utilizando a transformada de Hough. Chan [3] preocupou-se com a extração de lábios humanos. Lai e Figueiredo [4], Ji e Haralick [5], Abdou e Pratt [6] e Song e Udpa [7] analisaram estatísticas locais para construir detetores de bordas. Já Arie e Rao [8,9,10,11] criaram detetores a partir da projeção da imagem em um conjunto adequado de funções-bases e proporcionaram grande contribuição na detecção não só de bordas, mas também de junções e degradês ("ramp"). Já Bellon, Direne e Silva [12] optaram primeiramente por segmentar a imagem para posteriormente extrair as bordas. Motivados pelas aplicações em vídeo-conferência e MPEG-4, Li, Hatzinakos e Venetsanopoulos [13], Darrell e Pentland [14] e Black e Jepson [15] desenvolveram algoritmos de segmentação baseados no movimento.

Neste artigo, como no de Pappas [16], abordamos com ênfase a segmentação de imagens naturais em poucas regiões, independentemente do índice de incidência de bordas ou degradês.

Para ter-se um parâmetro de comparação de técnicas e resultados, foram tomados os trabalhos de Boukerroui, Basset e Baskurt [17], Lim e Ho [18] e Won [19].

Em [17], os autores conseguiram ótimos resultados ao aplicarem o problema de estimação via máxima probabilidade a posteriori com o objetivo de segmentar imagens com texturas bem caracterizadas, como aquelas geradas por processos gaussianos. Nele, o método foi aplicado com imagens sintetizadas computacionalmente.

Em [18], utilizou-se uma técnica interessante que consiste em transformar a região de uma imagem principal, sobre um fundo secundário, num conjunto formado pela união de figuras geométricas bem conhecidas, como retângulos e elipses, de diferentes tamanhos, de forma a separar a imagem principal. Esta técnica mostrou-se ser de custo computacional muito baixo e bastante atraente para aplicações em vídeo-conferência.

Em [19], Won analisou a DCT<sub>2x2</sub> calculada ao longo de uma imagem e segmentou imagens baseado no grau de similariedade entre os coeficientes das DCT<sub>2x2</sub> calculados em janelas 2x2 vizinhas. O grande benefício deste trabalho foi oferecer facilidades para diferenciarem-se regiões com texturas diferentes.

Neste artigo, propomos um algoritmo que processa uma imagem qualquer e a segmenta num conjunto reduzido de regiões, baseado na análise local da média e variância entre pontos vizinhos da imagem. O algoritmo sugerido pode ser aplicado tanto em imagens de tons de cinza ou coloridas. Entre as suas vantagens, pode-se destacar a possibilidade de identificar regiões com grandes áreas.

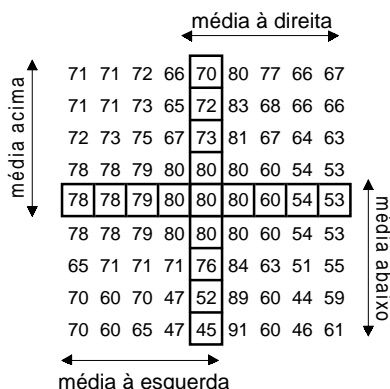
Para apresentar a técnica proposta, inicialmente descreveremos as estatísticas locais utilizadas na aplicação do algoritmo de Médias-K. Em seguida, abordaremos o processo de varredura da imagem e alguns métodos de análise e refinamento dos resultados. Finalmente, mostraremos simulações que comprovaram a eficiência da técnica de segmentação aqui descrita.

### 2. ESCOLHA DAS ESTATÍSTICAS LOCAIS

A escolha das estatísticas locais exerce grande influência no desempenho de uma técnica de segmentação.

Antes de escolher-se a forma de varrer uma imagem em busca de regiões, foi primeiramente necessário escolher que características da imagem são consideradas relevantes no sentido de obter-se uma classificação eficiente. Won [19] defendeu o uso dos coeficientes de uma DCT 2x2, que, na verdade, representam uma média de 4 pontos e 2 médias de 2 derivadas direcionais. Este método é particularmente eficiente onde não exista uma grande variação entre as intensidades de "pixels" vizinhos.

Neste trabalho, investigamos o uso de duas novas estatísticas, que chamamos "Média Direcional Mais Próxima", ou CDM (*Closest Directional Mean*), e "Menor Variância Direcional", ou SDV (*Smallest Directional Variance*). Tanto a CDM quanto a SDV são calculados para cada ponto da imagem. Um exemplo de cálculo da CDM é mostrado na Figura 1. Nela, a CDM associada ao ponto central vale 79, que é a média calculada à esquerda, já que todas as outras médias direcionais (65,4, 75 e 66,6) são mais distantes de 80 (valor da intensidade do pixel). Simulações comprovaram que a CDM é muito eficiente para determinar com precisão as fronteiras dos segmentos. Já a SDV se mostra adequada para representar regiões com texturas regulares. Com a intenção de evitar a separação de uma textura bem caracterizada em diferentes regiões, a SDV foi limitada a um valor de saturação igual a "LimSat".



**Fig. 1:** Exemplo de cálculo da mais próxima média direcional. Em (a), têm-se as intensidades da imagem, destacando-se o valor 80 para o ponto central, de onde se deseja obter as médias direcionais. De (b), chega-se ao valor 79 para a média à esquerda. De (c), chega-se ao valor 65,4 para média à direita. De (d), o valor 75 para média acima. De (e), o valor 66,6 para média abaixo. Portanto, a CDM vale 79.

Uma justificativa para o uso da CDM e da SDV pode ser obtida ao serem considerados os pontos numa fronteira entre regiões. Se a região for razoavelmente suave, em geral pelo menos um dos 4 conjuntos de pontos indicados na fig.1 vai pertencer à mesma região do ponto central. Desta forma, a média referente àquele conjunto tenderá a ser próxima à intensidade do pixel central. Isto quer dizer

que a CDM tem o atributo de caracterizar pontos pertencentes a uma região. De forma semelhante, para regiões com texturas ou grande variação de intensidade, a SDV provê uma caracterização efetiva. Experimentos indicaram que o uso de mais direções no cálculo da CDM e da SDV não causou qualquer ganho de desempenho no algoritmo.

Desta forma, um vetor de 2 dimensões (CDM e SDV) foi associado a cada ponto de uma imagem representada por tons de cinza. Para imagens coloridas, o número de dimensões é 6, ou seja, foram calculadas uma CDM e uma SDV para cada uma das componentes de cor.

### 3. VARREDURA DA IMAGEM

Para buscar as regiões que apresentem semelhança nas estatísticas escolhidas, dividiu-se uma imagem de  $K_1 N_p \times K_2 N_p$  pontos,  $K_1$  e  $K_2$  inteiros quaisquer, em  $K_1 \times K_2$  blocos de  $N_p \times N_p$  pontos. No restante desta discussão, chamaremos cada conjunto de  $2N_p \times 2N_p$  pixels de superbloco. A cada superbloco, aplicou-se o algoritmo de Médias-K [20], com 3 iterações e usando 4 centros iniciais. Ao final da aplicação do Médias-K, chega-se a um total de 4 regiões, onde cada uma delas é representada por um centro cujo vetor de estatísticas possua a menor distância euclidiana para o vetor médio de estatísticas daquela região. Caso dois centros representantes de regiões tenham vetores de estatísticas com distância euclidiana entre si menor que um limiar pré-estabelecido, chamado a partir deste ponto de "LimCen", as duas regiões são agrupadas.

Como mostrado na Figura 2, que ilustra o caso de uma imagem de 16x16 com  $N_p=4$ , os blocos  $N_p \times N_p$  são varridos da seguinte forma:

- 1) a varredura da imagem é feita em linhas de superblocos, da esquerda para a direita; os centros iniciais do primeiro superbloco são arbitrados como sendo os vetores de estatísticas dos seus 4 cantos; aplica-se o Médias-K;
- 2) para a primeira linha de superblocos utiliza-se uma superposição de 2 blocos  $N_p \times N_p$  (fig.2-c e 2-d) e continua-se aplicando o Médias-K; agora, os centros são inicializados da seguinte forma: 2 centros são escolhidos das 2 regiões mais populosas na fronteira à direita dos dois blocos  $N_p \times N_p$  superpostos (analisado no passo anterior); um terceiro centro é aquele com a maior média harmônica das distâncias aos 2 centros iniciais, no espaço de vetores de estatísticas; o quarto centro é escolhido de forma análoga, porém levando-se em conta a média harmônica aos 3 centros anteriores; apenas os pontos pertencentes às regiões correspondendo aos 2 centros

iniciais e os pontos dos novos blocos poderão ser processados neste passo;

3) ao iniciar-se a varredura de uma nova linha de superblocos, faz-se uma superposição de 2 blocos  $N_p \times N_p$  da linha anterior de blocos (fig.2-e); neste caso, as 2 regiões mais populosas que cruzam a fronteira superior dos blocos superpostos participam do Médias-K;

4) a partir deste ponto (fig.2-f e fig.2-g), utiliza-se uma superposição de 3 blocos  $N_p \times N_p$ , onde os 3 centros iniciais são os centros das 3 regiões mais populosas na fronteira (à esquerda e abaixo) do novo bloco e insere-se no processamento mais um centro, o mais distante (via média harmônica) dos 3 centros pré-inicializados;

5) chega-se a uma segmentação como a da fig.2-h, que, no exemplo, possui 5 regiões.

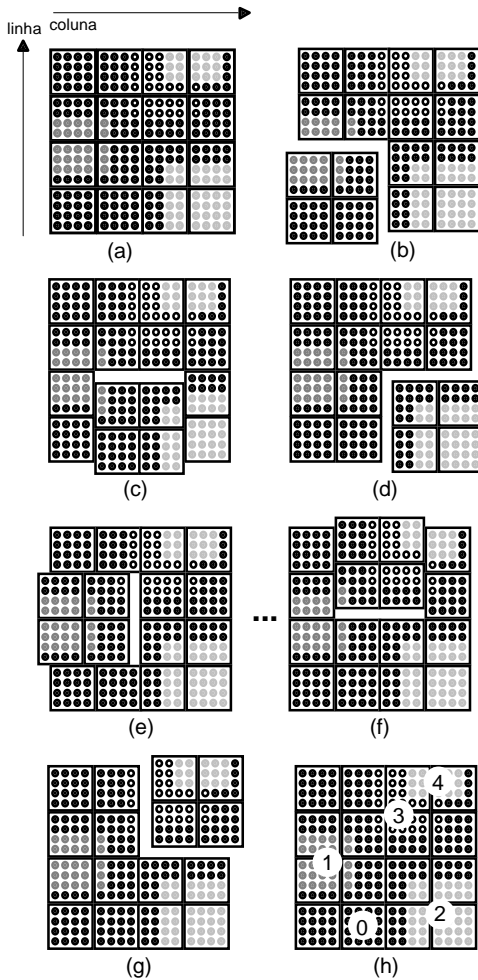


Fig. 2: Sequência para varrer blocos 4x4 e superblocos 8x8.

#### 4. REFINAMENTO DA SEGMENTAÇÃO

Após ser realizada a varredura da imagem, calcula-se um vetor médio de estatísticas que representa uma região " $\alpha$ ", a partir de agora chamado de "CodeBook[  $\alpha$  ]".

Faz-se uma nova varredura dos pontos  $X_{ij}$  da imagem, com o objetivo de refinar o processo de segmentação. Sejam  $\alpha$  e  $\beta$  regiões distintas. Primeiramente, definimos a sua fronteira pelas equações

$$X_{ij} \in \alpha, \quad X_{mn} \in \beta, \quad (1)$$

$$|i - m| \leq 1, \quad |j - n| \leq 1. \quad (2)$$

Considerando  $v(X_{ij})$  o vetor de estatísticas de um pixel  $X_{ij}$ , se as condições

$$\|v(X_{ij}) - v(X_{mn})\| \leq \text{LimCen}, \quad (3)$$

$$\| \text{CodeBook}[\alpha] - \text{CodeBook}[\beta] \| \leq \text{LimCen}, \quad (4)$$

forem satisfeitas para qualquer dos pixels da fronteira (eqs. 1 e 2), então as regiões  $\alpha$  e  $\beta$  deverão ser agrupadas numa única região  $\alpha \cup \beta$ . Note que a equação 3 significa que os dois pixels têm estatísticas muito parecidas e que a equação 4 diz que suas regiões  $\alpha$  e  $\beta$  também.

Em imagens com variações de intensidade (degradês) progressivos ao longo de toda a imagem, como nos ombros da Lena, é possível que regiões afastadas, com estatísticas muito diferentes, mas separadas por um degradê, sejam segmentadas numa mesma região. Para evitar este tipo de problema, cada região passou por um processo de análise. Neste processo, extraem-se os pontos  $X_{\max}$  e  $X_{\min}$ , que são aqueles com as estatísticas mais distantes na região. Se  $\|v(X_{\max}) - v(X_{\min})\| > T$ , definimos duas novas regiões associadas a  $X_{\max}$  e  $X_{\min}$ , através das equações 5 e 6, onde "T" é um fator de tolerância:

$$\|v(X_{ij}) - v(X_{\min})\| \leq T, \quad (5)$$

$$\|v(X_{ij}) - v(X_{\max})\| \leq T. \quad (6)$$

Os pontos  $X_{ij}$  que satisfizerem à equação 5 e forem conectados definem uma outra região. O mesmo procedimento é adotado com relação à equação 6. Se  $\|v(X_{\min}) - v(X_{\max})\| \leq 2T$ , o termo "T" das equações 5 e 6 é substituído por  $1/2 \|v(X_{\min}) - v(X_{\max})\|$ , ou seja, metade da distância entre  $v(X_{\min})$  e  $v(X_{\max})$ . Este procedimento é repetido até alcançar-se a estabilidade das regiões.

Como último procedimento, proíbe-se que determinadas regiões sejam segmentadas com menos de  $P_{\min}$  pontos. Caso, após os procedimentos anteriores, isto tenha acontecido, a região com menos de  $P_{\min}$  pontos é eliminada e cada um de seus pontos é arbitrado como pertencente à região vizinha mais próxima.

## 5. SIMULAÇÕES

Para simular o algoritmo de segmentação proposto neste artigo, foram feitos testes com as imagens "Lena" e "Hardware" construídas a partir de 256 tons de cinza. Para as duas imagens, foram adotados os seguintes parâmetros:  $N_p = 8$ ,  $LimCen = 0,07$ ,  $LimSat = 3000$ ,  $LimVar = 0,14$ ,  $P_{\min} = 20$ ,  $T = 0,25$ . As figuras 3 e 4 mostram os resultados obtidos com este conjunto de parâmetros. A imagem "Lena" foi segmentada em 88 regiões e a "Hardware" em 44. O número de 44 foi considerado alto para a imagem "Hardware". Então, os valores de  $LimCen$  e  $T$  foram alterados para 0,14 e 0,4, respectivamente. Com isto, obteve-se uma nova segmentação com 22 regiões. As figuras 3 e 4 mostram também resultados obtidos ao adicionar-se ruído branco uniforme de média zero e variância 8,3. Nota-se que o algoritmo é bastante robusto à adição de ruído. Numa das simulações para a imagem "Lena", optou-se por pré-equalizar os histogramas de intensidade, o que resultou numa segmentação muito bem definida e com apenas 26 regiões.

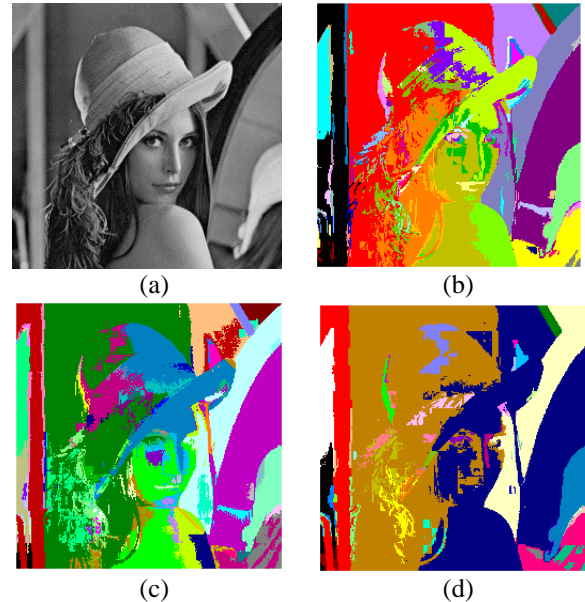
O esforço computacional exigido pelo algoritmo proposto é muito pequeno e pode ser facilmente implementado com um Processador Digital de Sinais para ser utilizado em aplicações em tempo real. O único procedimento que demanda um esforço maior é o que separa regiões unidas por degradês progressivos, mas este esforço não chega a afetar significativamente o desempenho em termos de tempo de processamento.

## 6. CONCLUSÕES

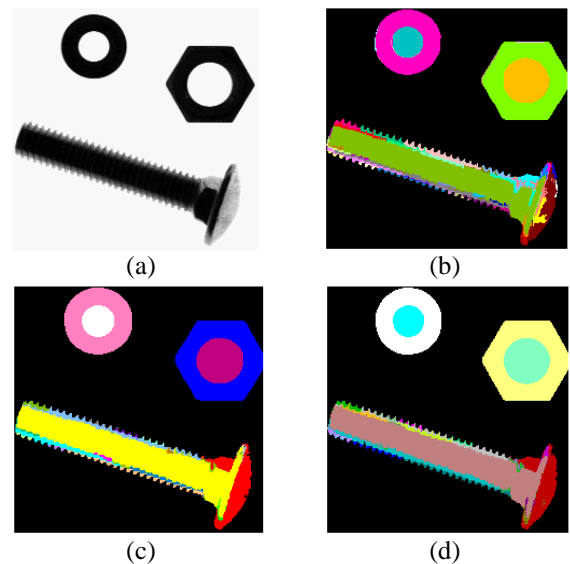
Neste artigo, propusemos uma técnica de segmentação de imagens em diferentes regiões através da aplicação do método de Médias-K num conjunto de vetores associados a pontos da imagem. Estes vetores foram construídos a partir de estatísticas locais, que denominamos "Média Direcional mais Próxima" (ou CDM) e "Menor Variância Direcional" (ou SDV).

As imagens obtidas após as simulações apresentaram-se, sob o critério de qualidade subjetiva, bastante satisfatórias, mesmo quando foi adicionado ruído à imagem original. Demais técnicas da literatura também apresentam resultados semelhantes, contudo, estabelecem algumas restrições não impostas pelo trabalho descrito neste artigo, como o conhecimento prévio de estatísticas de textura [17] ou do fundo da imagem [18]. Já o trabalho

[19] não impõe qualquer restrição, mas fornece resultados de segmentação com um número elevado de regiões se comparado com o trabalho que propomos.



**Fig. 3:** Simulação com a imagem "Lena". (a) Imagem original. (b) Imagem obtida após a aplicação do algoritmo dividida em 88 regiões, cada uma representada por uma cor diferente. (c) Imagem com ruído branco uniforme de média zero e variância 8,3 segmentada em 75 regiões. (d) Imagem segmentada em 36 regiões após a equalização do histograma de intensidades.



**Fig. 4:** Simulação com a imagem "Hardware". (a) Imagem original. (b) Resultado da segmentação em 44 regiões. (c) Resultado da segmentação em 22 regiões. (d) Imagem com ruído branco uniforme de média zero e variância 8,3 segmentada em 21 regiões.

## 7. REFERÊNCIAS

- [1] {{ISO{/}MPEG 90{/}176}}, *Coding of Moving Pictures and Associated Audio*, Committee Draft of Standard {ISO} 11172, 1990.
- [2] Ji, Q. e Haralick, R.M., *An Optimal Bayesian Hough Transform for Line Detection*, IEEE Inter. Conf. on Image Processing, 1999.
- [3] Chan, M.T., *Automatic Lip Model Extraction for Constrained Countour-Based Tracking*, IEEE Inter. Conf. on Image Processing, 1999.
- [4] Lai, G.C. e Figueiredo, R.J.P., *Robust Extraction of Low Contrast Edges Using Clustering-Based Segmentation and Refinement*, IEEE Inter. Conf. on Image Processing, 1999.
- [5] Ji, Q. e Haralick, R.M., *Quantitative Evaluation of Edge Detectors Using the Minimum Kernel Variance Criterion*, IEEE Inter. Conf. on Image Processing, 1999.
- [6] Abdou, K.E. e Pratt, W.K., *Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors*, Proc. of IEEE, 67(5)753-763, 1979.
- [7] Song, Y.W. e Udpa, S.S., *A New Edge Detection Algorithm Using Data Fusion Approaches*, IEEE Inter. Conf. on Image Processing, 1999.
- [8] Arie, J.B. e Rao, R., *A Novel Approach for Template Matching by Nonorthogonal Image Expansion*, IEEE Trans. on Circuits and Systems for Video Technology, Vol.3, N° 1, 1993.
- [9] Rao, R e Arie, J.B., *Optimal Edge Detection Using Expansion Matching and Restoration*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.16, N° 12, 1994.
- [10] Wang, Z., Rao, R e Arie, J.B., *Optimal Ramp Edge Detection Using Expansion Matching*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.18, N° 11, 1996.
- [11] Nandy, D. e Arie, J.B., *EXM Eigen Templates for Detecting and Classifying Arbitrary Junctions*, IEEE Inter. Conf. on Image Processing, 1998.
- [12] Bellon, O.R.P., Direne, A.I. e Silva, L., *Edge Detection to Guide Range Image Segmentation by Clustering Techniques*, IEEE Inter. Conf. on Image Processing, 1999.
- [13] Li, Y., Hatzinakos, D. e Venetsanopoulos, A.N., *A Multi-Frame Region-Feature Based Technique for Motion Segmentation*, IEEE Inter. Conf. on Image Processing, 1999.
- [14] Darrell, T. e Pentland, A., *Robust Estimation of a Multi-Layered Motion Representation*, Proc. IEEE Workshop on Visual Motion, pp 173-178, 1991.
- [15] Black, M.J. e Jepson, A., *Estimation Multiple Independent Motions in Segmented Images Using Parametric Models with Local Deformations*, Proc. IEEE Conference on Computational Vision Pattern Recog., 1994.
- [16] Pappas, T.N., *An Adaptive Clustering Algorithm for Image Segmentation*, IEEE Trans. on Signal Processing, SP-40(4):901-914, 1992.
- [17] Boukerroui, D., Basset, O. e Baskurt, A., *Multiresolution Adaptive Image Segmentation Based on Global and Local Statistics*, IEEE Inter. Conf. on Image Processing, 1999.
- [18] Lim, D.K. e Ho, Y.S., *Image Segmentation Using Hierarchical Meshes*, IEEE Inter. Conf. on Image Processing, 1999.
- [19] Won, C. S., *Improved Block-Based Image Segmentation*, IEEE Inter. Conf. on Image Processing, 1999.
- [20] Tou, J.T. e Gonzalez, R.C., *Pattern Recognition Principles*, Addison Wesley, 1974.