

# Uso de Saliências do Contorno via Esqueletização para Caracterização de Formas

Glauco V. Pedrosa<sup>1</sup> (Bolsista PIBIC/CNPq), Marcos A. Batista<sup>2</sup>, Henrique C. Fernandes<sup>1</sup>,  
Celia A. Z. Barcelos<sup>1</sup> (Orientadora)

<sup>1</sup> Universidade Federal de Uberlândia      <sup>2</sup> Universidade Federal de Goiás

E-mail: glaucovitor@comp.ufu.br, marcos@catalao.ufg.br, celiazb@ufu.br

## Resumo

*Neste trabalho, mostraremos o descritor saliências do contorno para a caracterização de formas juntamente com o algoritmo de casamento proposto para medir a distância entre duas formas caracterizadas pelos seus pontos de saliências. Alguns experimentos são mostrados e discutidos para demonstrar o potencial do descritor na recuperação de formas.*

## 1. Introdução

Inúmeras áreas do conhecimento humano vêm se apoiando cada vez mais no uso de imagens e, como conseqüência, surge a necessidade de armazenar, manter e consultar grandes bases de dados de imagens eficientemente. Assim, nasceram os CBIR's (CBIR - Context Based Image Retrieval), que são sistemas que realizam a recuperação de imagens.

A recuperação de imagens pode ser feita baseada principalmente em três aspectos: cor, forma e textura, sendo que, em um sistema de recuperação, se pode utilizar apenas um ou mesmo uma combinação desses três elementos. Em algumas situações, a forma é o aspecto mais importante, sendo que invariância quanto a transformações é imprescindível.

Uma maneira eficiente de representar uma forma é através de suas saliências. Elas são definidas como os pontos de maior curvatura ao longo do contorno, por isso capturam a essência de uma forma, uma vez que qualquer parte sobressalente ou mesmo reentrância são consideradas saliências, e são essas as características que diferenciam duas formas.

Neste artigo abordaremos a recuperação de imagens baseada em forma utilizando as saliências de contorno como descritor. As saliências serão obtidas a partir do esqueleto como em [2] utilizando uma técnica denominada Transformada Imagem Floresta. Juntamente apresentaremos um algoritmo de casamento

originalmente proposto em [1]. Por fim, mostraremos alguns resultados experimentais da implementação de um CBIR baseado em formas usando o descritor.

## 2. IFT – Image Foresting Transform

Uma forma eficiente de se encontrar as saliências é através da técnica denominada Transformada Imagem Floresta (IFT- Image Foresting Transform), bastante explorada em [1; 2; 3; 4], que utiliza grafos para a representação de uma imagem através de seus relacionamentos de conectividade: os nodos são pixels e os arcos são relações de adjacência entre os pixels. A IFT é uma transformada projetada para operadores de processamento de imagem que gera uma representação em multiescala.

Cada ponto do contorno é tomado como uma semente e a partir desta um caminho de custo mínimo é determinado. Cada caminho constitui uma árvore e a união delas constitui uma floresta cobrindo a imagem inteira. A IFT gera duas diferentes representações da forma: contornos multiescala por dilatação exata e esqueletos multiescala por propagação de rótulos.

O algoritmo da IFT utiliza um conjunto de sementes  $S$  formado pelos pixels do contorno. Cada elemento de  $S$  possui um rótulo diferente e o objetivo é propagá-los por toda a imagem. Para isso, deve-se percorrer todos os pixels da imagem e encontrar o pixel semente correspondente, em  $S$ , com menor custo. O custo é calculado de acordo com a distância euclidiana e caso seja mínimo entre um pixel semente (pertencente ao contorno) e um pixel qualquer, o rótulo da semente é atribuído a este pixel. Se o pixel já pertencer ao conjunto  $S$  o custo dele é 0.

No final do algoritmo, teremos um mapa de raízes  $R$  e um mapa de custo  $C$  para todos os pixels da imagem.

A figura 1-c mostra o resultado da IFT aplicada à imagem 1-a. A propagação dos rótulos das sementes de  $S$  resulta em uma imagem totalmente rotulada.

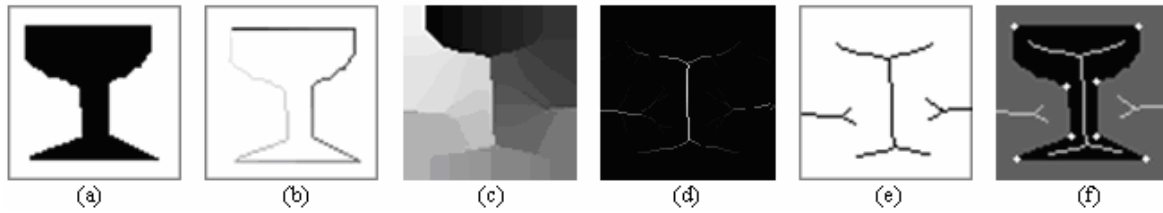


Figura 1: Saliências através da IFT. (a) Forma original. (b) Contorno rotulado. (c) Imagem rotulada. (d) Imagem diferença. (e) Threshold aplicado à imagem diferença. (f) Pontos claros são saliências obtidas da relação entre a extremidade do esqueleto e o contorno.

### 3. O processo de rotulação

Para a obtenção do conjunto de sementes  $S$ , necessários para a aplicação da IFT, é aplicado o algoritmo de rotulação, o qual encontra as sementes e atribui rótulos incrementais às mesmas começando de 1. Assim, os rótulos variam no intervalo de 1, 2, ...,  $N$  onde  $N$  é o número de pixels do contorno. A estratégia usada para o algoritmo de rotulação é visitar cada pixel do contorno uma única vez caminhando no sentido anti-horário (ou horário) atribuindo um rótulo diferente a cada um.

Na figura 1-b pode ser visto o resultado da rotulação dos pixels do contorno da figura 1-a.

### 4. Esqueletos

Inúmeras são as aplicações de esqueletos devido à sua capacidade de representar formas de maneira compacta, dentre essas aplicações tem-se a obtenção das saliências visto que, toda extremidade de esqueleto interno, corresponde a uma saliência convexa, e extremidades do esqueleto externo correspondem a saliências côncavas.

Esqueletos multiescala são facilmente obtidos da IFT, já que desta resulta um mapa de rótulos  $L$  que possibilita separar os pixels em regiões discretas de Voronoi. O esqueleto é extraído a partir das divisões destas regiões, sendo que a obtenção da imagem diferença  $D$  é a parte responsável por tal etapa, seu cálculo é feito pela seguinte equação:

$$D(p) = \max_{\forall q \in A_4(p)} \{\min(\delta(p, q), N - \delta(p, q))\} \quad (1)$$

onde  $A_4(p)$  são os *pixels* na vizinhança-de-quatro relativa ao pixel  $p$ ,  $\delta(p, q) = L(q) - L(p)$ , e  $L(q)$  é o rótulo do pixel  $q$ . Dessa equação pode se observar que valores altos são gerados quando os valores dos pixels em  $L$  são muito diferentes, caso contrário valores baixos serão obtidos. Ao representar a imagem diferença,

figura 1-d, pode-se perceber linhas mais claras próximas ao centro da forma em que se tem uma forte diferença entre os pixels, ou seja, trata-se de linhas significativas do esqueleto, tais linhas escurecem à medida que se aproximam das extremidades pois nesses locais a diferença entre os pixels não é tão acentuada como no centro da forma. Portanto, quanto maior a diferença de um pixel para seus vizinhos, maior o valor atribuído a esse pixel e maiores são as chances de que ele pertença ao esqueleto.

A imagem diferença possui inúmeros detalhes sendo que a maioria não pode ser vista através do sistema visual humano, estes são os detalhes menos relevantes que geram linhas muito escuras e imperceptíveis e muitas vezes não devem estar presentes no esqueleto. A fim de obter um esqueleto que represente a forma convenientemente, aplica-se um threshold na imagem diferença, ou seja, um limiar que separa os pontos com tons mais claros (os mais relevantes) dos mais escuros (os menos relevantes).

Uma estimativa para tal valor, é encontrada em [3], que sugere 5% de  $N$ , sendo que  $N$  é o número de pixels do contorno e por consequência o número de regiões discretas de Voronoi [5]. Esta estratégia se fundamenta no fato de que a máxima diferença que pode ser obtida é  $N/2$  e que, no entanto, não se pode obter diferenças tão grandes já que uma vizinhança de quatro está sendo analisada e os pixels vizinhos tendem a ter valores próximos devido à rotulação incremental.

A figura 1-e mostra o resultado do threshold aplicado à imagem 1-d.

### 5. Geração do conjunto de características

Os valores contidos na imagem diferença, obtida pela equação 1, podem ser interpretados como sendo o comprimento máximo do segmento mais curto entre duas raízes equidistantes [3], segmento  $\overline{dab}$ , figura 2. Com isso, é necessário apenas fazer a relação entre a extremidade do esqueleto e o meio deste segmento, já que esta é a saliência. Desta forma, partindo da extremidade do esqueleto  $c$ , deve-se encontrar a raiz

deste pixel através do mapa raiz  $R$  (obtido pela IFT) e em seguida percorrer a distância  $\overline{dab}/2$  no sentido horário (ou anti-horário). Com isso a posição da saliência torna-se conhecida.

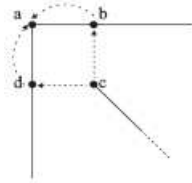


Figura 2: Relação entre o esqueleto e a saliência.

Obtida a posição da saliência, é necessário atribuir um valor para a mesma. Aqui novamente a IFT é de grande importância: através do mapa de raízes  $R$  e o mapa de custo  $C$ , obtidos pela IFT, é possível obter o tamanho da área de influência de cada ponto do contorno através da simples construção de um histograma. O valor da saliência é tomado como sendo a área de influência interna mais a área de influência externa da saliência.

Falta apenas determinar se a saliência é côncava ou convexa. No primeiro caso o valor da saliência torna-se negativo e no segundo permanece positivo. Se a extremidade é de esqueleto interno, a saliência é convexa, caso contrário é côncava.

Para tornar o conjunto de características mais eficiente a posição a ser utilizada deve ser a posição relativa da saliência com relação ao contorno: escolha-se um ponto arbitrário como sendo o ponto inicial, com isso cada ponto tem uma posição. Em seguida, divide-se o valor da posição pelo total de pontos no contorno. Essa última medida visa garantir que haja normalização quando mudanças de escala são realizadas.

Portanto, uma saliência  $b$  de uma forma é representada pelo par  $(u_b, s_b)$ , onde  $u_b \in [0,1]$  é a posição relativa e  $s_b \in [0,1]$  é o tamanho da área de influência.

## 6. Algoritmo de Casamento

O objetivo do algoritmo de casamento é medir a distância entre duas formas A e B. Essas formas são, respectivamente, representadas pelos vetores:

$$F_A = \{(u_{A1}, s_{A1}), \dots, (u_{An}, s_{An})\}$$

$$F_B = \{(u_{B1}, s_{B1}), \dots, (u_{Bm}, s_{Bm})\}$$

onde  $(u_{Ai}, s_{Ai})$ , representa a  $i$ -ésima saliência da forma A,  $u_{Ai}$  é a posição relativa e  $s_{Ai}$  é o valor que representa o tamanho da área de influência.

O número de saliências  $n$  de A pode ser diferente do número de saliências  $m$  de B. O casamento entre as formas A e B é definido pelo algoritmo:

1. Crie  $F'_A = \{(u'_{A1}, s'_{A1}), \dots, (u'_{An}, s'_{An})\}$  e  $F'_B = \{(u'_{B1}, s'_{B1}), \dots, (u'_{Bm}, s'_{Bm})\}$ , ordenando  $F'_A$  e  $F'_B$  por ordem decrescente da área de influência.

2. Cria a lista L contendo pares de pontos candidatos a casamento de  $F'_A$  e  $F'_B$ . Um par  $((u'_{Ai}, s'_{Ai}), (u'_{Bj}, s'_{Bj}))$  está em L se  $|s'_{Ai} - s'_{Bj}| \leq 0.2 * s'_{A1}$ . Um par  $((u'_{Bj}, s'_{Bj}), (u'_{Ai}, s'_{Ai}))$  está em L se  $|s'_{Bj} - s'_{Ai}| \leq 0.2 * s'_{B1}$ .

3. Para cada par de candidatos a casamento na forma  $P_{ij} = ((u'_{Ai}, s'_{Ai}), (u'_{Bj}, s'_{Bj}))$  em L, encontre o parâmetro de deslocamento  $\alpha$  tal que  $\alpha = u'_{Ai} - u'_{Bj}$ . Desloque os pontos de saliências de  $F'_A$  por  $\alpha$ , obtendo  $F''_A = \{(u''_{A1}, s''_{A1}), \dots, (u''_{An}, s''_{An})\}$

4. A distância  $d$  entre  $F''_A$  e  $F'_B$  é dada por:

$$d = \sum_{k=1}^{\min\{n,m\}} d_k$$

onde

$$d_k = \begin{cases} \sqrt{(u''_{Ak} - u_{Bk})^2 + (s''_{Ak} - s_{Bk})^2}, & \text{se } |u''_{Ak} - u_{Bk}| \leq 0.2 \\ s''_{Ak} + s_{Bk} & \text{caso contrário} \end{cases}$$

Finalmente, se  $n \neq m$ , uma altura  $s$  dos pontos não casados é adicionada à soma  $d$ .

5. Repita os passos 3 e 4 considerando outros pares de candidatos em L

6. Selecione a menor distância  $d$  como sendo a distância entre  $F_A$  e  $F_B$ .


































Query		Ranking									
		1	2	3	4	5	6	7	8	9	10
A											
B											
C											

Tabela 1. Resultado das 10 primeiras posições usando as queries A, B, C

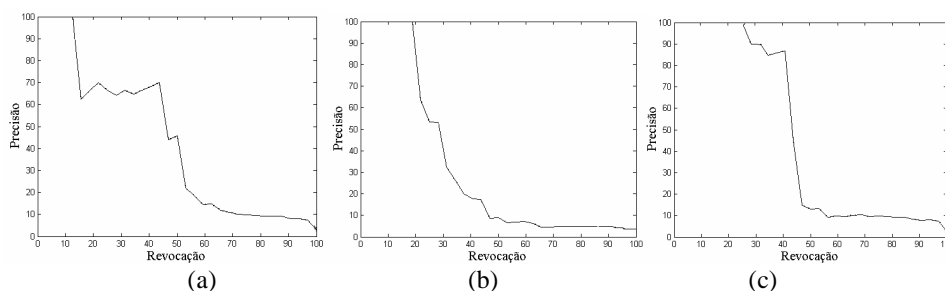


Figura 3 - Gráfico Precisão x Revocação da queries da tabela 1: (a) Query A (b) Query B (c) Query C

## 7. Resultados Experimentais

Implementamos um CBIR baseado em formas, a fim de testar a construção e o comportamento do descritor num ambiente de recuperação de imagens.

A base de dados é formada por 40 tipos diferentes de formas, e cada uma possui 32 variações quanto à escala e rotação. A imagem da forma juntamente com suas variações é considerada uma classe e as imagens relevantes de uma imagem consulta é a classe a qual essa imagem consulta pertence.

Nas consultas uma imagem exemplo é fornecida, suas características são extraídas, a busca é efetuada e um ranking é construído. A tabela 1 mostra o ranking das 10 primeiras posições de 3 imagens, usadas como consulta. Na primeira consulta, podemos observar que as 6 primeiras posições são relevantes. Na segunda, apenas as posições 1 a 4, 8, 9, 10 são relevantes. Na terceira, apenas a posição 9 não é relevante.

## 8. Conclusão

Este trabalho apresentou o descritor saliências do contorno, usado para caracterizar uma forma. Implementamos um CBIR baseado em formas e utilizamos esse descritor para analisar seus resultados na recuperação de formas. Os experimentos mostram o potencial da técnica.

## 9. Referências

- [1] R. da S. Torres, E. M. Picado, A. X. Falcão, and L. da F. Costa, "Effective Image Retrieval by Shape Saliences". SIBGRAPI, 2003, 166-174, 12-15.
- [2] Falcão, A. X., Costa, L. F., and Cunha, B. S. "Multiescala Skeletons by Image Foresting Transform and its applications to neuromorphometry". *Pattern Recognition*, 35. pp. 1571-1582, 2002.
- [3] R. da S. Torres, A. X. Falcão, and L. da F. Costa. "A graph-based approach for multiscale shape analysis". *Pattern Recognition*, 37(6), pp. 1163-1174, 2004.
- [4] Torres, R. S., and Falcão, A. X. "Contour Saliency Descriptors for Effective Image Retrieval and Analysis". *Image and Vision Computing journal*, pp. 1-11, 2006.
- [5] Hoff III, K. E., Culver, T., Keyser, J., Lin, M. e Manosha, D.. "Fast computation of generalized Voronoi diagrams using graphics hardware". *Proceedings of ACM SIGGRAPH using graphics hardware*, 2000.
- [6] Costa, L. da F., Campos, A. G., Manoel, E. T. M.: "An Integrated Approach to Shape Analysis: Results e Perspectives". *International Conference on Quality Control by Artificial Vision*, Le Cresot, France, (2001) pp. 23-24.
- [7] Costa, L. da F., Estrozi, L. F.: "Multiresolution Shape Representation without Border Shifting". *Electronic Letters*, Vol.35 (1999) pp. 1829-1830.
- [8] Arica, N., Vural, F. T. Y.: "A Perceptual Shape Descriptor". In *International Conference on Pattern Recognition*, (2002) pp. 375-378.