

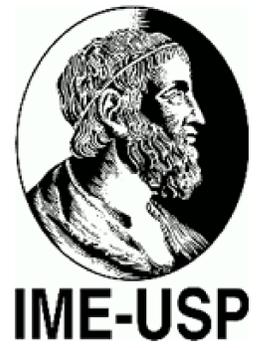
Complex Network Spectrum Signature - Um novo descritor de formas

Rodrigo Zerbini

Supervisor: Prof. Dr. Paulo A. V. de Miranda

Universidade de São Paulo - Instituto de Matemática

rodrigo.zerbini@usp.br — (11) 94294 1261



Resumo

Neste trabalho, o foco será em um sistema de visão computacional na área de recuperação de imagens baseada em conteúdo, conhecida como CBIR (*Content-Based Image Retrieval*). Ela pode ser entendida como qualquer tecnologia que ajude a buscar e organizar uma coleção de imagens digitais através de seus conteúdos visuais. Neste contexto, o desafio está em desenvolver descritores de formas e medidas de similaridades com alta acurácia. Algumas aplicações destes descritores são, por exemplo, correspondência de imagens, classificação de forma e reconhecimento de caracteres.

Introdução

Em 2015, três pesquisadores do Rio Grande do Sul propuseram um novo método descritor de formas: o *Complex Network Spectrum Signature* (CNSS) [1], que é baseado em redes complexas e teoria espectral de grafos. O CNSS cria, para cada imagem, uma rede de pixels interconectados para o qual um vetor de características é calculado. A comparação entre os vetores de características, os quais resumem as principais características da forma do objeto, é feita por meio de uma função de distância. Quanto menor a distância entre dois vetores de características, mais as imagens são similares.

O CNSS, além de apresentar um alto grau de precisão nas bases testadas, é bastante rápido quando comparado com outros métodos: sua função de distância tem complexidade linear.

Objetivos

1. Fazer uma revisão bibliográfica sobre alguns conceitos teóricos importantes para este trabalho tais como descritores de forma, redes complexas e teoria espectral de grafos.
2. Implementar o *Complex Network Spectrum Signature*.
3. Repetir os testes para algumas bases presentes no artigo e comparar os resultados obtidos com os originais.
4. Melhorar a performance do método e aplicá-lo a outros tipos de bases de imagens.

Modelagem da rede

Neste trabalho, os objetos são representados por imagens binárias, isto é, a região da forma do objeto é dada por um conjunto de pixels brancos sobre um fundo preto. Primeiramente, deve-se extrair os pixels que formam o contorno da forma do objeto.



Figura 1: Região da forma

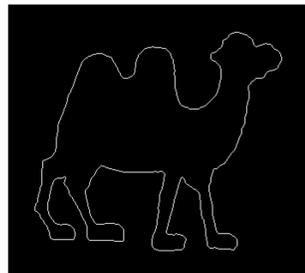


Figura 2: Contorno da forma

Um grafo então é formado através da associação de cada pixel do contorno a um vértice do grafo e a criação de uma aresta entre todos os pares de vértices. Esse tipo especial de grafo, onde todo vértice é adjacente a todos os outros vértices, é chamado de grafo completo.

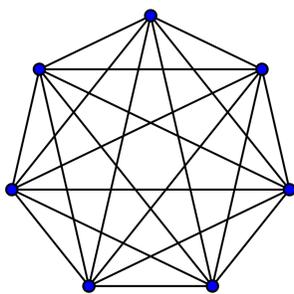


Figura 3: Exemplo de grafo completo

O próximo passo é associar um peso a cada uma das arestas do grafo. O valor do peso da aresta e_{ij} será dado pela distância euclidiana $d(p_i, p_j)$ entre os vértices p_i e p_j . Sabe-se que a distância euclidiana é invariante a transformações ortogonais, ou seja, seu valor não é afetado se a forma sofre translação ou rotação. Dado dois vértices $p_i = (x_i, y_i)$ e $p_j = (x_j, y_j)$ a distância euclidiana entre eles pode ser calculada usando a seguinte fórmula:

$$d(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Essas distâncias podem ser organizadas em uma matriz de pesos W , onde cada entrada w_{ij} é a distância entre os vértices p_i e p_j . A fim de evitar efeitos de escala, serão utilizadas as distâncias relativas entre os vértices. Para isso, os pesos da matriz W devem ser normalizados em relação à maior distância entre dois vértices:

$$w_{ij} = \frac{d(i, j)}{\max(W)}$$

A partir da nova matriz de pesos normalizada dez matrizes de adjacência podem ser construídas. Uma vez que todas as distâncias normalizadas encontram-se no intervalo $[0, 1]$, cada uma das dez matrizes de adjacência será associada a um décimo desse intervalo. Matematicamente, as entradas $m(i, j)_t$ da matriz de adjacência M_t podem ser obtidas com a equação abaixo:

$$m(i, j)_t = \begin{cases} 1, & \text{se } \frac{t-1}{10} \leq w(i, j) \leq \frac{t}{10} \\ 0, & \text{caso contrário} \end{cases} \quad (1 \leq t \leq 10)$$

Os autovalores das matrizes de adjacência podem ser encontrados através da fórmula do polinômio característico:

$$p(\lambda_t) = \det(\lambda_t I - m_t) \quad (1 \leq t \leq 10)$$

Finalmente, a assinatura da forma pode ser encontrada. Ela é representada por um vetor de 60 números reais formado da seguinte maneira: tomam-se os 3 maiores e os 3 menores autovalores associados a cada matriz de adjacência, aplica-se o módulo ao vetor e, por último, o mesmo é normalizado em relação à soma de todos seus elementos.

Similaridade entre formas

Agora que as formas dos objetos estão caracterizadas por suas assinaturas, é necessário um método para realizar a comparação entre duas assinaturas. Isso será feito através da distância de Hellinger, que é função do coeficiente de Bhattacharyya. O coeficiente de Bhattacharyya $bc(sn_1, sn_2)$ para um par de assinaturas sn_1 e sn_2 é calculado pela equação abaixo:

$$bc(sn_1, sn_2) = \sum_k \sqrt{sn_1(k)sn_2(k)}$$

onde os termos $sn_1(k)$ e $sn_2(k)$ são o k -ésimo valor das assinaturas das formas.

Com o coeficiente de Bhattacharyya determinado é possível calcular a distância de Hellinger $dh(sn_1, sn_2)$ entre as assinaturas das formas:

$$dh(sn_1, sn_2) = \sqrt{1 - bc(sn_1, sn_2)}$$

A distância de Hellinger é um valor entre zero e um e sua função é medir a similaridade entre duas assinaturas de formas. Assim, quanto mais próxima de zero for a distância, mais similares as formas dos objetos devem ser. Por outro lado, quanto maior for a distância, mais distintas devem ser as formas.

Resultados

- Os índices de acertos *bull's eye* para o MPEG-7 e a base de folhas foram respectivamente 61.15% e 62.63%.
- Nas figuras abaixo é possível ver as distâncias entre as imagens das bases MPEG-7 e da base das folhas. Nota-se que a variância das distâncias entre as folhas foi maior do que a observada na MPEG-7.

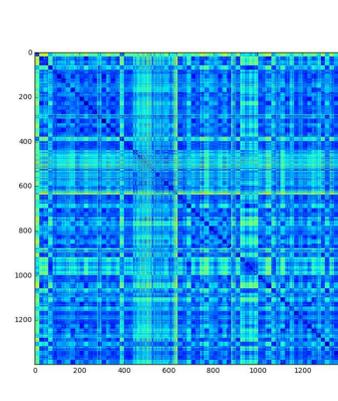


Figura 4: Distâncias entre imagens da MPEG-7

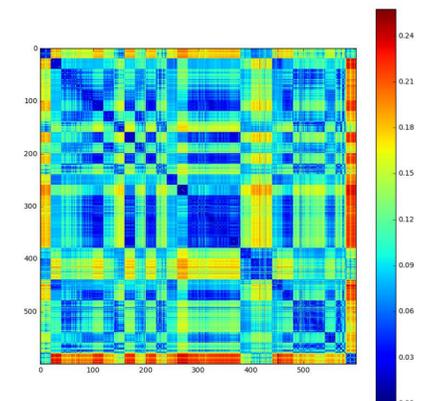


Figura 5: Distâncias entre imagens do banco de folhas

Conclusões

- Os resultados mostram que o algoritmo implementado não teve a mesma acurácia que aquela proposta no artigo (89,47% para MPEG-7 e 82,15% para folhas).
- A razão da diferença entre os valores obtidos e os da literatura é desconhecida. Uma outra implementação do mesmo método ainda foi feita em outra linguagem de programação, mas os resultados obtidos foram muito próximos aos da primeira implementação.
- Entrou-se em contato com os autores do artigo sobre o CNSS, mas não foi possível ter acesso ao código fonte no momento.

Referências

- [1] de Oliveira A. B., da Silva P. R., and Barone D. A. C. A novel 2d shape signature method based on complex network spectrum. *Pattern Recognition Letters*, 63:43–49, 2015.