
MAPA DE CARACTERÍSTICAS AUTO-ORGANIZÁVEL PARA SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

L. A. Silva*, R.A.Moreno**, E.Del-Moral-Hernandez*, S. S. Furuie**

*Departamento de Engenharia de Sistemas Eletrônicos/Escola Politécnica da
Universidade de São Paulo, São Paulo, Brasil

**Instituto do Coração (InCor)/Faculdade de Medicina da Universidade de São Paulo,
São Paulo, Brasil

e-mail: leandro.augusto@ieee.org

Abstract: This work addresses a method of employing the Self-Organizing Map (SOM) neural network in the ranking of feature vector, extracted from medical image using the multiresolution wavelet, for a content-based image retrieval (CBIR) system. CBIR systems usually use a query image for retrieval of similar images. The feature vectors are ranked in SOM map using the mean quantization error. The experiments were carried out to evaluate the wavelet coefficient with minor final mean quantization error and minor topographic error and to evaluate retrieval performance using the proposed method with the recall-precision curve.

Palavras-chave: Artificial Neural Network, Self-Organizing Map, Content-based Image Retrieval, Wavelets.

Introdução

A prática moderna em saúde mostra que, o uso de imagens médicas se tornou instrumento essencial para o diagnóstico de doenças. O uso da modalidade adequada da imagem, como raio-X, ressonância magnética (RM), ultra-som, entre outras, pode fornecer evidências físicas sobre a doença e até mesmo sua origem. Por esta razão, o volume de imagens produzidas diariamente é grande e o seu armazenamento é fundamental [1].

Estas imagens armazenadas podem ser usadas como referências em diagnósticos ou até mesmo no auxílio a tomada de decisão. Para isso, o uso de sistemas de recuperação de imagens por conteúdo, ou CBIR (*Content-Based Image Retrieval*) como são comumente conhecidos, é necessário.

Em CBIR, as imagens são representadas por características (conteúdo) como cor, forma e textura [1],[2]. As características formam um vetor (vetor de características) que podem ser usados individualmente, uma única característica, ou combinadas com outras características. A consulta por imagens no CBIR é feita em relação aos vetores de características. Para que isto seja possível, em termos computacionais, os vetores de características devem ser organizados e estruturados. Em geral, esta tarefa é feita usando-se técnicas de indexação multidimensional como *quadrees*, *k-dimensional trees* (k-d-

trees), *R-trees*, etc [3]-[6] ou técnicas de inteligência computacional como redes neurais artificiais [7]-[9]. A rede neural de mapa auto-organizável (SOM, de Self-Organizing Map) ou rede de Kohonen é a mais utilizada, por suas funcionalidades como agrupamento e preservação topológica dos vetores de características, além de trabalhar bem com vetores em alta dimensão [10].

Neste trabalho, as imagens utilizadas são de ressonância magnética. Como extrator de características será utilizada a técnica de *wavelets*, pois caracteriza as propriedades locais da imagem e comprime a imagem original [11]. A função de base *wavelets* empregada é a Haar, pois exige menor esforço computacional em relação às outras funções de base e mantém a representação da imagem original. Os vetores de características serão ordenados nos neurônios artificiais do mapa auto-organizado, a partir da medida do erro médio de quantização. Este método é proposto para potenciais aplicações em sistemas CBIR, sobretudo na busca pelas imagens mais semelhantes.

O artigo é organizado da seguinte forma: é apresentada uma discussão sobre recuperação de imagens por conteúdo, extração de características por *wavelets*, construção do mapa auto-organizável, bem como o método proposto para ordenação dos vetores de características e conseqüentemente, os resultados experimentais são apresentados. Por fim, a última seção apresenta as principais conclusões do trabalho.

Recuperação de Imagens por Conteúdo

Recuperação de imagens por conteúdo é campo de pesquisa desde 1970 [4]. Nas primeiras tentativas, as imagens eram ordenadas por anotações que descreviam o conteúdo da imagem e o usuário formulava a recuperação com base nas anotações. No início dos anos 90, a tecnologia de armazenamento era bastante evoluída e o uso de imagens digitais se tornava comum, o que resultou em um aumento significativo no volume de imagens armazenadas diariamente. Com isso, escrever anotações a partir do conteúdo da imagem se tornava impossível, não só pelo grande volume de imagens, mas pela subjetividade humana nas anotações. Atualmente, o conteúdo das imagens é extraído usando técnicas automáticas ou

semi-automáticas de processamento de imagens. Para facilitar a ordenação e a recuperação destas imagens, técnicas usadas em recuperação de informação são necessárias. Neste trabalho não serão feitas discussões sobre tais técnicas, para um maior aprofundamento no assunto veja [3]-[9]. Na seção seguinte será discutida a técnica aplicada neste trabalho para extração de características.

Extração de Características

A primeira fase, e a mais importante, tanto para um sistema CBIR como para um sistema de classificação de imagens é a extração de características. Em geral, as características comuns de serem extraídas são: cor, textura e forma [1]. Na extração de características, as imagens da base de dados (referências) são transformadas em vetores de características. Uma maneira alternativa de extração de características se faz através do uso de técnicas de processamento de sinais. As duas principais vantagens no uso destas técnicas são a efetiva caracterização de propriedades locais da imagem e a capacidade de compressão de imagens. A transformada de *wavelets* é um exemplo de técnicas de processamento de sinais. *Wavelets* são usadas para representar diferentes níveis de detalhes em uma sub-imagem. Cada sub-imagem contém informações como escala e orientação, que são convenientemente separadas. A implementação da análise de multiresolução, da qual *wavelet* é um tipo, foi derivada por Mallat [11]. A *wavelet* Haar é a mais simples das transformadas *wavelets*. Existe uma variedade de bases populares de *wavelets*, incluindo *Daubechies wavelets*, *Mexican Hat wavelets* e *Morlet wavelets* [11]. Porém estas últimas têm a desvantagem de necessitar de um maior esforço computacional em relação a *wavelet* Haar, que traz também como vantagem, a melhor representação da imagem original.

A transformada *wavelet* decompõe a imagem em coeficientes. O coeficiente de aproximação (cA_n) representa as baixas frequências da imagem e os coeficientes de decomposição cD_n^h (h-horizantal), cD_n^v (v-vertical) e cD_n^d (d-diagonal) representam as componentes de alta frequência. A decomposição pode ser feita em vários níveis (n), sendo que o coeficiente cA é o remanescente para as demais decomposições. Assim, para uma imagem de tamanho 256×256 , em um primeiro nível de decomposição, $n=1$, ela é decomposta em quatro coeficientes, com tamanho reduzido para 128×128 . Detalhes da escolha do número de decomposição e do coeficiente que representa melhor a imagem original serão discutidos na seção resultados experimentais. A seguir, uma breve descrição do mapa auto-organizável.

Mapa Auto-Organizável

O Mapa Auto-Organizável (*Self-Organizing Maps* ou SOM) é uma arquitetura de rede neural, com aprendizado do tipo não-supervisionado [10]. O SOM é uma rede em duas camadas que transforma o espaço dos vetores de características de dimensão d , em um mapa discreto

de duas dimensões, com preservação das relações topológicas entre as diferentes características. Os neurônios da primeira camada recebem do extrator de características, o coeficiente da transformada *wavelets* (vetor de características - \mathbf{x}). A segunda camada de neurônios é organizada em um mapa 2-D, que preserva as relações topológicas de \mathbf{x} . Para isto, existe uma relação de vizinhança entre os neurônios do mapa (veja mais detalhes em [10]). Os neurônios do mapa ou unidades do mapa são caracterizados por sua posição u_x e u_y , ou seja, $\mathbf{u}=[u_x \ u_y]$. Cada \mathbf{u} tem associado um vetor protótipo (ou modelo) $\mathbf{m}_u = [m_{u1}, m_{u2}, \dots, m_{ud}] \in \mathfrak{R}^d$ sendo d a dimensão do protótipo, a mesma das características de entrada.

Existem duas boas razões para o uso do SOM na organização dos vetores de características. A primeira é devida sua capacidade de separar em grupos o espaço dos vetores de características, adaptativamente. Assim, cada \mathbf{m}_u representa o centro destes grupos. A segunda motivação é que os neurônios de saída são ordenados topologicamente e com isso, neurônios vizinhos correspondem a grupos similares do espaço de vetores de características original.

Os vetores protótipos da rede são adaptados pelo algoritmo de treinamento do SOM em um processo iterativo e também chamado de competitivo. Inicialmente, em $\mathbf{m}_u(0)$, o vetor protótipo deve ser iniciado com algum valor. Comumente define-se a inicialização como sendo randômica, mas existem outras formas de se fazer isso, como mostrado em [10]. Em cada passo do processo, \mathbf{x} é randomicamente escolhido do conjunto de vetores de características (ou conjunto de treinamento). A distância, geralmente euclidiana, entre o vetor de entrada e todos os vetores protótipos é calculada. O neurônio com menor distância é chamado de neurônio vencedor ou BMU (de *Best-Matching Unit*), ou seja, é o \mathbf{u} com protótipo \mathbf{m} mais próximo a \mathbf{x} . Em seguida, os vetores protótipos são atualizados, conforme a regra vista em [10]. O \mathbf{m} do BMU e da sua vizinhança são movidos para próximos a \mathbf{x} , como se fosse um "arraste", para que vetores semelhantes tenham como neurônios vencedores, unidades vizinhas. Desta forma, o mapa preserva a representação topológica dos vetores de características. Para análise de desempenho do SOM existem três medidas, as quais serão apresentadas a seguir.

Medidas de Desempenho do SOM

As medidas usadas para calcular desempenho do SOM são: erro médio de quantização (mqe), erro médio final de quantização (MQE) e erro topográfico (TE) [12].

O erro médio de quantização (mqe) mede a distância média entre os vetores de entrada com seu respectivo protótipo \mathbf{m}_u . O cálculo dessa medida é:

$$mqe_u = \frac{1}{n} \sum_{x \in N} \{d(x(n), m_u)\} \quad (1)$$

O erro médio final de quantização (MQE) do mapa é calculado a partir dos mqe de cada \mathbf{u} :

$$MQE = \frac{1}{u} \sum_{m=1}^u m q e_m \quad (2)$$

A segunda medida de qualidade é o erro topográfico (*TE*) que quantifica a capacidade do mapa em representar a topologia dos vetores de características de entrada. Para cada x calcula-se o primeiro BMU e o segundo BMU, e então é calculado o *TE* de acordo com [12]:

$$TE = \frac{1}{N} \sum_{n=1}^N u(x(n)) \quad (3)$$

Onde N é o número total de vetores de características e $u(x(n))$ é igual a 1, se o primeiro BMU não for adjacente ao segundo BMU e 0 do contrário.

Estas duas últimas medidas (Equações (2),(3)) serão usadas para definir o coeficiente *wavelets* usado como vetores de características das imagens. Por outro lado, a medida *mqe* (Equação (1)) será utilizada para ordenar os vetores de características das imagens junto ao seu protótipo, o que simplificará a revocação por similaridade, como será apresentado a seguir.

Ordenando os vetores de características

A contribuição deste trabalho está na tentativa de ordenar (ou organizar) os vetores de características em relação aos protótipos do SOM. Para isto, os vetores de características de cada protótipo são organizados em relação à medida *mqe*. Primeiramente, o SOM deve ser treinado com as imagens de referência (ou treinamento), ou seja, gera-se o mapa. Por conseguinte, todas as imagens, representadas e associadas aos respectivos vetores de características envolvidos no treinamento do SOM são projetados no mapa (veja Figura1).

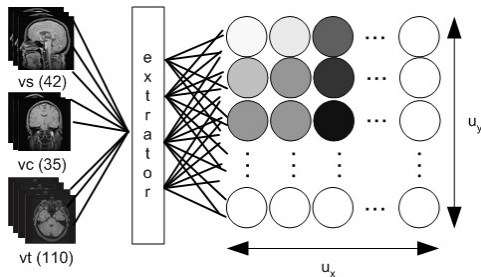


Figura1: As imagens de entrada de diferentes classes (vs (42) – 42 imagens de vista sagital, vc (35) – 35 imagens de vista coronal e vt (110) – 110 imagens de vista transversal). Cada imagem passa pelo extrator e os vetores de características são extraídos e usados para gerar o mapa (treinamento SOM).

As imagens são ordenadas a partir do *mqe* de cada protótipo. Deste modo, ao apresentar uma imagem nova, o seu vetor é extraído e comparado com os protótipos. A ordenação dos neurônios vencedores (BMU) é efetuada. Desta maneira, caso não tenha a quantidade de imagens a ser revocadas no neurônio vencedor (1ºBMU), consequentemente revoca-se as imagens do segundo neurônio vencedor e assim sucessivamente, até o último neurônio vencedor ou até que o número desejado de revocação seja atingido. O esquema desta metodologia para revocação é ilustrado na Figura 2.

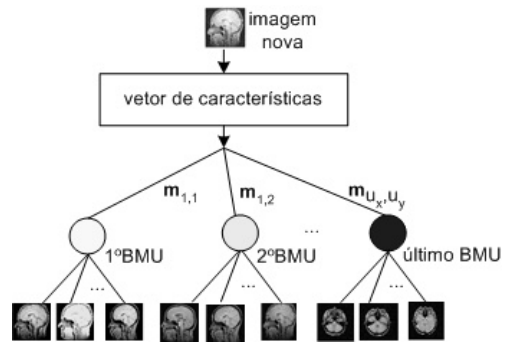


Figura 2: Ao apresentar uma imagem nova, o vetor de característica é extraído e então comparado com os vetores protótipos. De acordo com a similaridade, ordena-se os neurônios vencedores. Cada neurônio tem associado imagens de referência utilizadas no treinamento ordenadas em relação a *mqe*.

Resultados Experimentais

Foram selecionadas para os experimentos 187 imagens de ressonância magnética. Estas imagens foram separadas em três grandes grupos: vista sagital (42), vista coronal (35) e vista transversal (110) todas as vistas da cabeça. As imagens estão no formato DICOM e todas de tamanho 256x256 pixels. Os experimentos foram realizados em um PC Athlon XP 2200 MHz usando o software MATLAB.

Foi aplicada a transformada *wavelet* de base Haar, em cada uma das imagens. O nível de decomposição aplicado foi de 3 ($n=3$), que resulta em imagens de tamanho 32x32 em cada um dos quatro coeficiente. Isto implica em dimensão de 4096 elementos para cada vetor de característica. Para a escolha do coeficiente *wavelet* a ser usado como vetor de característica, realizou-se experimentos com SOM de 100 neurônios ($u_x=10$ e $u_y=10$). A partir destes experimentos, cujos resultados estão na Tabela 1 foi definido o coeficiente cA_3 para representar o vetor de característica, escolhido por apresentar menor valor médio entre *MQE* e *TE*.

Tabela 1: Estudo comparativo entre os coeficientes *wavelets* para escolha do vetor de característica.

	cA_3	cD_3^h	cD_3^v	cD_3^d
<i>MQE</i>	10,60	20,71	20,68	22,74
<i>TE</i>	0,02	0,02	0,00	0,01

Com o uso do cA_3 como vetor de característica, gerou-se o mapa SOM. Para análise de desempenho deste método de ordenação usou-se a curva revocação-precisão [1]. A revocação mede a proporção de imagens relevante presentes nas imagens de referência que foram recuperadas. Imagens relevantes são aquelas semelhantes à imagem nova consultada. A precisão mede a proporção das imagens revocadas em uma consulta, que são relevantes. O melhor resultado é precisão elevada para todos os níveis de revocação, como nas Figura 3b2 e Figura 3b3. Em geral, quando a revocação aumenta, a precisão diminui, como Figura 3b1.

Usando imagens novas de cada grupo de imagem, a revocação foi conduzida para verificar a eficiência de ordenação com o método proposto (veja Figura 2). O resultado de 20 imagens de referência revocadas para cada imagem nova, com a respectiva curva de revocação-precisão é ilustrado por cada Figura 3b).

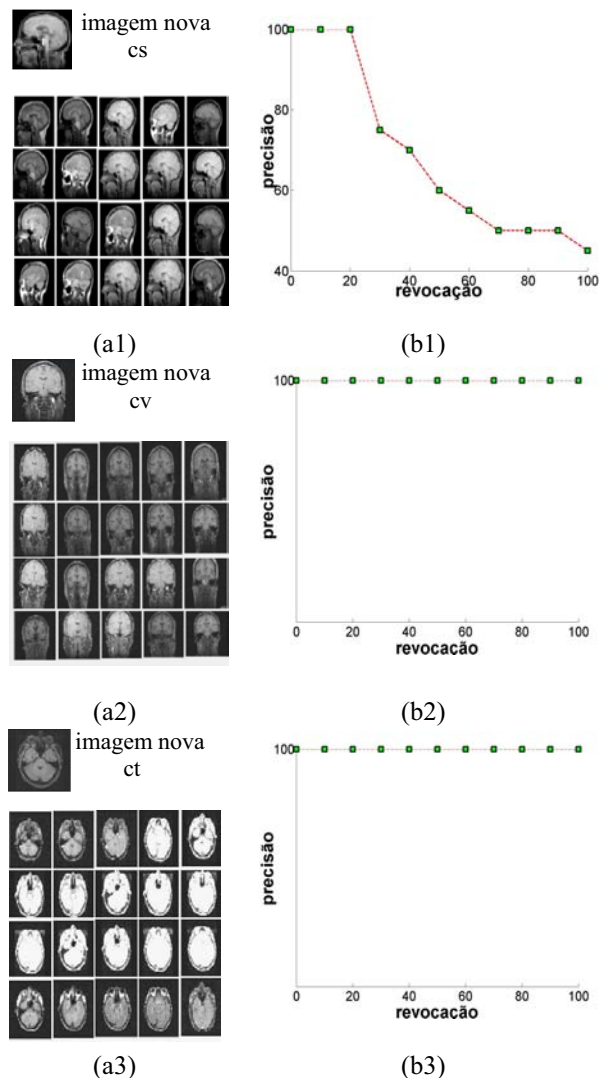


Figura 3: Cada imagem (a) ilustra a revocação de vinte imagens de referência, a partir de uma imagem nova. A ordem de semelhança segue o sentido esquerdo para direito. Os gráficos (b) são as curvas de recuperação-precisão para cada imagem nova.

Conclusões

Este trabalho apresentou um estudo exploratório do uso da rede neural artificial SOM com o uso do erro médio de quantização (Equação (1)) para a ordenação de imagens em relação ao seu conteúdo. Apresentou-se também o uso de coeficientes de wavelets com função de base Haar como método alternativo para a extração de características de imagem, bem como a seleção dos coeficientes (aproximação e detalhamento), escolhidos a partir do menor valor médio entre o erro final de quantização (Equação (2)) e o erro topográfico (Equação (3)).

Para ilustrar a eficiência da ordenação proposta neste trabalho, os experimentos de revocação de imagens semelhantes a uma imagem nova foram conduzidos e os resultados indicam sua potencialidade em sistemas CBIR.

Trabalhos futuros usando mais grupos e base de imagens mais extensa serão realizados. Estudos para verificar o uso desta técnica para encontrar grupos de imagens semelhantes de forma a organizar ou categorizar uma grande base de imagens, para a construção de um sistema CBIR deverão ser considerados.

Referências

- [1] Müller H., Michoux N., Bandon D. and Geissbuhler A., *A review of content-based image retrieval systems in medical applications-clinical benefits and future directions*, International Journal of Medical Informatic, vol. 73, pp. 1–23, 2004.
- [2] Deb S. and Zhang Y., *An overview of content-based image retrieval techniques*, Advanced Information Networking and Applications, vol.1, pp. 59–64, 2004.
- [3] White D. and Jain R., *Similarity indexing: Algorithms and performance*, San Diego, USA., 1996, vol. 2670.
- [4] Rui Y. and Huang T.S., *Image retrieval: Current techniques, promising directions, and open issues*, Journal of Visual Communication and Image Representation, vol. 10, pp. 39–62, 1999.
- [5] Böhm C., Berchtold S., and Keim D., *Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases*, 2001, vol. 3, p. 33.
- [6] Weber R., Schek H.-J., and Blott S., *A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces*, New York City, NY, USA., 1998, p. 194-205.
- [7] Wu S., Rahman M.K.M. and Chow T.W.S., *Content-based image retrieval using growing hierarchical self-organizing quadtree map*, The journal of Pattern Recognition Society, vol. 38, pp. 707–722, 2005.
- [8] Oh K.S., Kaneko K. and Makinouchi A., *Image classification and retrieval based on wavelet-som*, in DANTE '99: Proceedings of the 1999 International Symposium on Database Applications in Non-Traditional Environments, Washington, DC, USA, 1999, pp. 164–167, IEEE Computer Society.
- [9] Laaksonen J., Koskela M.; Laakso S., and Oja E., *Picsom – contentbased image retrieval with self-organizing maps*, Pattern Recognition Letters, vol. 21, no. 13, pp. 1199–1207, 2000.
- [10] Kohonen T. *Self-Organizing Maps*, Springer Series in Information Sciences, Vol. 30, third edition. Springer, Berlin, Heidelberg, New York, 2001.
- [11] Mallat S.G., *Multifrequency channel decompositions of images and wavelet models*, Transaction on Acoustics, Speech, and Signal Processing, vol. 37, no. 12, pp. 2091–2110, December 1989.
- [12] Kimmo Kiviluoto, *Topology preservation in self-organizing map*, Tech. Rep., Helsinki University of Technology, Laboratory of Computer and Information Science, Finland, 1995.