

Classificação de texturas usando redes neurais artificiais e o modelo de *Spot Noise*

Antonio Cesar Germano Martins, Alexandre da Silva Simões, Gustavo Imbrizi Prado

LAPI – Laboratório de Automação e Processamento de Imagens. Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus de Sorocaba. Avenida Três de Março, 511, Alto da Boa Vista – 18087-180 - Sorocaba – SP – Brasil.

amartins@sorocaba.unesp.br, assimoes@sorocaba.unesp.br,
gustavoprado@grad.sorocaba.unesp.br

Abstract. *This work presents a texture classification procedure based on the analysis of parameters from the power spectrum by an artificial neural network. The Spot Noise model is used as a guideline for: interpretation of random texture, parameter definition in the frequency domain and the analysis of the obtained results. Details about the neural network architecture are provided, as well as the tests carried out.*

Resumo. *Este trabalho apresenta um processo de classificação de texturas aleatórias utilizando uma rede neural artificial para a análise de um conjunto de parâmetros característicos do espectro de potência. O modelo de Spot Noise é utilizado como guia para a interpretação de texturas aleatórias, fundamentação para a definição dos parâmetros e análise dos resultados obtidos. São também discutidos detalhes sobre a definição da arquitetura da rede neural artificial empregada bem como os testes realizados.*

1. Introdução

Textura é uma importante característica que pode ser utilizada na classificação de imagens, com aplicações em diversos segmentos, tais como: análise e classificação de carne bovina [Basset 2000], análise de superfícies de alimentos [Quevedo 2002], determinação da composição de misturas de grãos [Ojala 1996], segmentação automática de terrenos através de fotos de satélites [Paget 1996] e diagnósticos médicos [Christodoulou 2003; Madablushi 2003].

Pode-se definir textura como sendo a qualidade visual de uma imagem relacionada à sua variação espacial de tons de cinza ou de cor [Rao 1990]. Os métodos de análise de texturas em imagens procuram avaliar a correlação entre estas variações e podem ser implementadas no domínio espacial ou de frequência, sendo que a principal motivação para o uso do segundo advém do fato de que cada ponto no domínio de frequência possui informação de todos os pixels da imagem.

Neste sentido, Weszka et al. (1976) apresentam um método para gerar parâmetros de uma textura a partir das somas das energias presentes em regiões do espectro de potência. Ao analisar estes parâmetros, os autores mostraram ser possível a discriminação de diferentes texturas.

Jernigan e D´Astous (1984) procuram tirar informações adicionais do espectro de potência de uma textura calculando a entropia dos valores de energia ao invés da soma.

Outras abordagens utilizam redes neurais artificiais associadas à aplicação de filtros de Gabor [Kachouie 2003] ou como forma de implementar um banco de filtros [Jain 1996a] para a classificação de texturas.

Existem ainda na literatura trabalhos que discutem o desempenho de diferentes métodos de caracterização de textura, tais como: a soma das energias, filtros espaciais, filtro de Gabor, estatísticas de tons de cinza e wavelet [Christodoulou, 2003]. Chen e Chen (1999), por exemplo, mostram que a análise de frequência consome menos tempo de processamento do que a baseada em filtros de Gabor.

Em visualização científica, o conceito de textura pode ser usado para mapear campos de velocidade e pressão sobre superfícies. Para esta aplicação, Wijk (1991) desenvolveu um método de síntese de textura chamado de *Spot Noise*. Martins et. al. (2001) usaram este modelo para estabelecer uma correlação entre uma textura em uma imagem e sons fricativos a partir da interpretação dos espectros de potência.

Este trabalho apresenta um método de classificação de texturas baseado no uso de uma rede neural artificial para a análise do espectro de potência, usando parâmetros característicos que são obtidos pela interpretação deste espectro a luz do modelo de *Spot Noise*. São apresentados detalhes da arquitetura da rede utilizada, resultados em imagens sintéticas e discutidas possíveis aplicações.

2. O modelo de *Spot Noise* como guia para a interpretação do espectro de potência

Dada uma imagem onde os valores dos pixels seguem uma distribuição randômica e um *spot*, que literalmente é uma mancha e neste contexto é uma figura que atua como um filtro, pode-se gerar uma textura aleatória a partir da operação de convolução [Wijk 1991]. A Figura 1 apresenta texturas geradas usando-se este modelo.

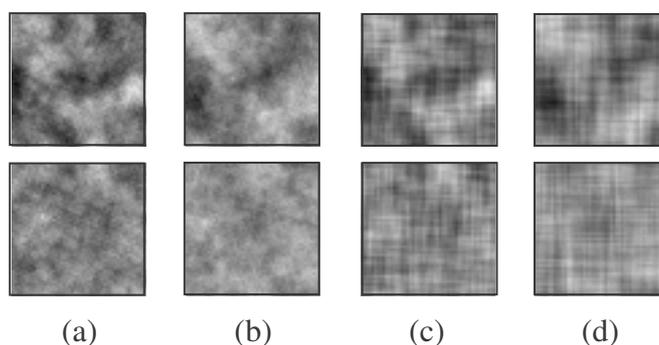


Figura 1. Imagens de 50 x 50 pixels como exemplos de texturas geradas com os seguintes spots: (a) círculo de diâmetro 12; (b) círculo de diâmetro 20; (c) quadrado de lado 12; (d) quadrado de lado 20.

Como é bem conhecido, dadas duas funções $f(x,y)$ e $g(x,y)$ a transformada de Fourier (TF) da convolução destas funções é igual a multiplicação das TFs de $f(x,y)$ e $g(x,y)$:

$$\mathfrak{F}(f(x, y) * g(x, y)) = \mathfrak{F}(f(x, y)) \cdot \mathfrak{F}(g(x, y)) \quad (1)$$

onde $\mathfrak{F}(\)$ denomina a operação da TF.

Se $f(x,y)$ representa um *spot* e $g(x,y)$ é um campo aleatório de impulsos de diversas intensidades, o espectro de potência da imagem obtida pela convolução de f e g é uma versão com ruídos do espectro de potência de $f(x,y)$, uma vez que o espectro de potência de um campo aleatório é essencialmente constante no caso de uma imagem digital. Deste modo, uma textura aleatória pode ser modelada como uma versão filtrada de um ruído, onde o filtro é representado por um *spot* de certa forma e tamanho, geralmente muito menor que o da imagem. Assim, assumindo que para cada textura exista um *spot* associado, o modelo de *Spot Noise* pode ser utilizado para se analisar texturas, conforme sugerido por Wijk (1991), ao se buscar características do *spot* no espectro de potência.

3. O método de classificação proposto

Para se obter informações sobre as características de simetria e tamanho do *spot* relacionado a uma textura, o espectro de potência é dividido em 8 setores circulares, com ângulo interno de 45° cada um, e calculado o primeiro momento destas sub-regiões, dado por:

$$m_i = \frac{1}{N} \sum_{(u,v) \in R_i} r_{u,v} \cdot P(u,v) \quad (2)$$

onde $P(u,v)$ é valor do espectro de potência na posição (u,v) , R_i é um setor circular com $i=1$ correspondendo a sub-região entre -22,5° e 22,5° (ver Figura 2) e N é um fator de normalização.

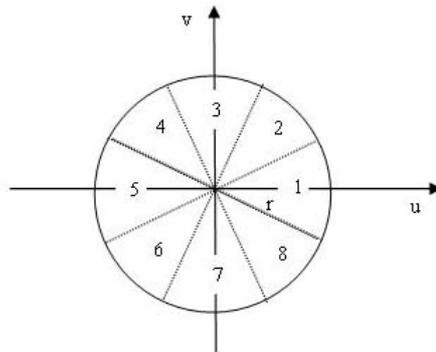


Figura 2. Setores circulares utilizados na análise do espectro de potência. Os números apresentados representam os índices para cada setor

Os valores destes momentos alimentam uma rede neural artificial com um neurônio na camada de saída, realizando, assim, um mapeamento $\mathbb{R}^8 \rightarrow \mathbb{R}$. Na fase de treinamento, são escolhidos valores apropriados para identificar cada uma das texturas apresentadas e na fase de classificação, as texturas são avaliadas de acordo com a proximidade do valor de saída e aqueles utilizados na fase de treinamento.

3.1. Determinação da arquitetura da rede neural utilizada

Como não se conhece a distribuição dos momentos no espaço R^8 decidiu-se por usar uma rede de quatro camadas, levando-se em consideração que as regiões a serem agrupadas possuem formatos genéricos [Jain 1996b].

A função de ativação utilizada foi:

$$G = a \cdot \tanh(bv) \quad (3)$$

onde v é a soma ponderada das entradas em um dado neurônio e os parâmetros a e b assumem, respectivamente, os valores 1,7159 e $2/3$, como sugerido por Haykin (1999).

Para o processo de treinamento desta rede foi utilizado o algoritmo de retropropagação, adotando-se o modo batch para a atualização dos pesos sinápticos, descartando-se as atualizações que levassem a um aumento do erro total. Desta forma, se um peso sináptico é atualizado e o erro total aumenta, mantém-se o valor anterior. Com isso, melhora-se o padrão de convergência da rede, não permitindo que os valores do erro total fiquem oscilando.

Cada linha da tabela de treinamento continha os 8 momentos associados a uma imagem e a saída desejada. Durante o processo de treinamento, o algoritmo acessava as linhas da tabela de forma seqüencial, de forma a atualizar os pesos sinápticos.

Para se definir o número de neurônios em cada uma das camadas da rede, realizou-se um conjunto de testes com diversas configurações. Procurou-se estudar a correlação do número de acessos à tabela de treinamento e o erro total obtido ao se atingir o critério de parada, sendo este o valor para que todos os erros das linhas da tabela de treinamento sejam menor que 0,0001. A taxa de aprendizagem foi definida como sendo 0,001 nos experimentos realizados.

A Figura 3 apresenta os resultados obtidos a partir da variação do erro total ao final do processo de treinamento em função do número de neurônios nas duas camadas intermediárias e a Figura 4, o número de vezes que a tabela de treinamento foi acessada. A configuração de 6 neurônios na primeira camada oculta e 14 na segunda não convergiu para o critério de parada e taxa de aprendizado utilizado, bem como configurações com menos de 6 neurônios nas camadas ocultas. A configuração ideal seria aquela em que o erro total e o número de acessos a tabela de treinamento fossem o menor possível. No entanto, como não se obteve tal configuração com os parâmetros utilizados, escolheu-se aquela em que o erro final era o menor.

Embora não tenham sido medidos os tempos de treinamento, notou-se que este aumenta consideravelmente com o aumento do número de neurônios, devido ao maior número de atualizações sinápticas.

Assim, definiu-se uma rede neural formada por uma arquitetura de perceptrons com quatro camadas, sendo duas ocultas, com 8 neurônios na primeira e 6 na segunda, uma camada de entrada com 8 neurônios e uma camada de saída com apenas um neurônio. As entradas do sistema são os oito momentos gerados a partir das imagens. A Figura 5 apresenta a arquitetura usada.

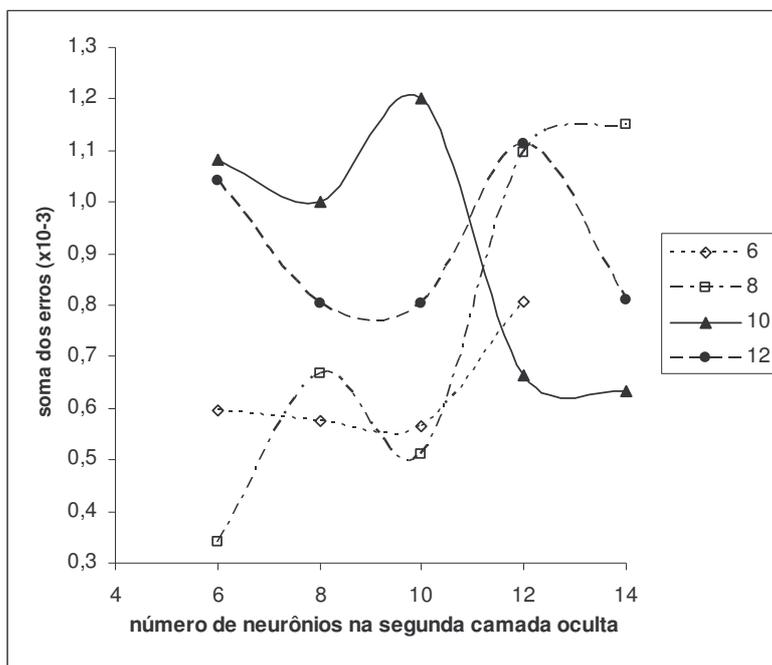


Figura 3. Estudo da correlação do número de neurônios e o erro total de saída. Os valores na legenda referem-se ao número de neurônios na primeira camada oculta. A configuração de 6 neurônios na primeira camada oculta e 14 na segunda não convergiu para o critério de parada e taxa de aprendizado utilizado.

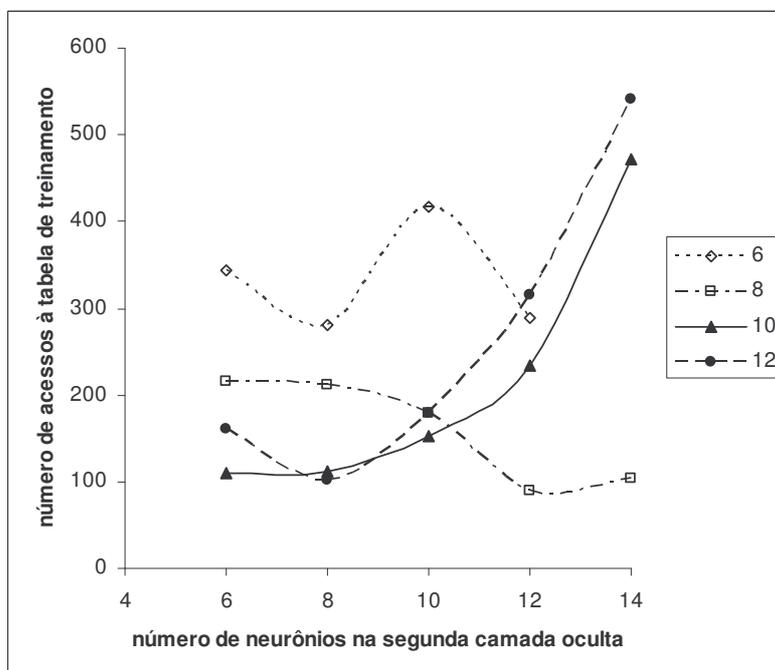


Figura 4. Estudo da correlação do número de neurônios e o número total de acessos à tabela de treinamento. Os valores na legenda referem-se ao número de neurônios na primeira camada oculta. A configuração de 6 neurônios na primeira camada oculta e 14 na segunda não convergiu para o critério de parada e taxa de aprendizado utilizado.

4. Resultados

4.1 Fase de Treinamento

Para se testar o método proposto utilizou-se o modelo de *Spot Noise* para sintetizar 4 imagens de 512 x 512 pixels com círculos de diâmetros 12 e 20 pixels e quadrados de lados 12 e 20 pixels como *spots*. Para cada uma destas imagens foram obtidas 16 sub-imagens de 256 x 256 pixels a partir de uma máscara móvel com deslocamento de 50 pixels na horizontal e na vertical. Assim, a tabela de treinamento foi composta por 64 linhas, resultado dos 4 tipos de *spot* e 16 sub-imagens cada, sendo que cada linha foi composta por um vetor com os 8 momentos.

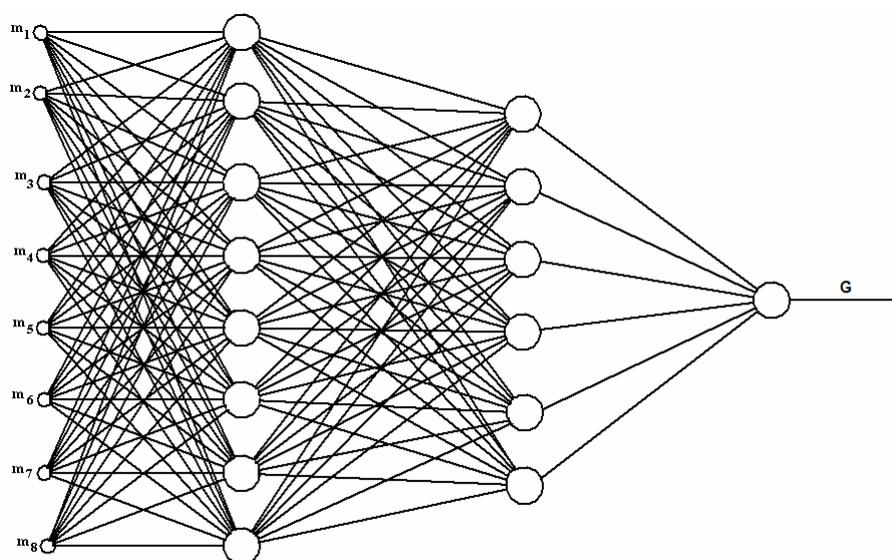


Figura 5. Arquitetura da rede utilizada.

O uso do modelo de *spot noise* permite a ordenação das imagens de acordo com o formato e tamanho dos *spots*. Desta forma, definiu-se como valores de treinamento aqueles apresentados na Tabela 1.

Tabela 1. Valores de treinamento.

Tipo de <i>spot</i>	Tamanho (pixels)	Valor de saída
Círculo	20	-1
	12	-0,5
Quadrado	12	0,5
	20	1

O treinamento convergiu para todas as imagens e a Figura 6 apresenta o comportamento da evolução do valor da soma do erro total no processo.

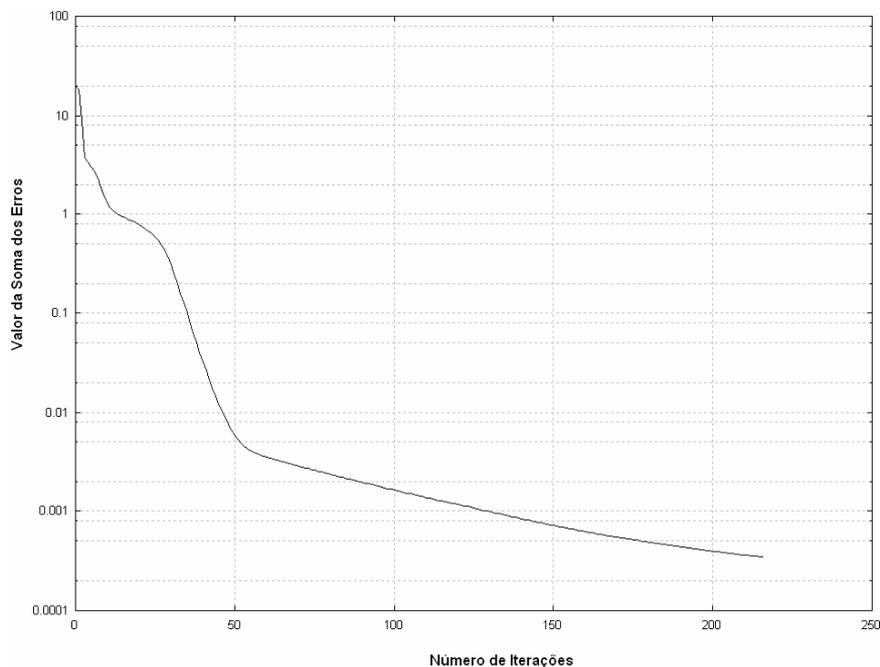


Figura 6. Gráfico da variação do valor da soma dos erros no processo de treinamento.

4.2. Fase de validação e verificação da generalização da rede

Para se verificar a eficácia da rede no processo de classificação utilizou-se 16 imagens distintas daquelas da fase de treinamento, porém mantendo-se os círculos de diâmetros 12 e 20 pixels e quadrados de lados 12 e 20 pixels como *spots*. Os resultados são apresentados na Figura 7 na forma de histogramas das saídas da rede neural, onde se pode notar que não ocorrem superposições dos valores de cada uma das classes de imagens definidas de acordo com o *spot* utilizado para a geração das texturas, levando a uma classificação correta de 100% dos padrões apresentados. Estes resultados mostram que a rede neural conseguiu identificar corretamente as características do *spot* presente no espectro de potência das imagens.

Para a verificação da generalização da rede utilizou-se um conjunto de imagens com *spots* de tamanhos distintos aos da fase de treinamento, mas com valores intermediários. Assim, foram utilizados círculos de diâmetros 8 e 16 pixels e quadrados de lados 8 e 16 pixels como *spots*. Os resultados obtidos encontram-se na Tabela 2 e mostram que, como desejado, as saídas da rede neural foram valores intermediários aos da fase de treinamento, levando a generalização da análise das características do *spot*, respeitando o formato e a ordem de tamanho.

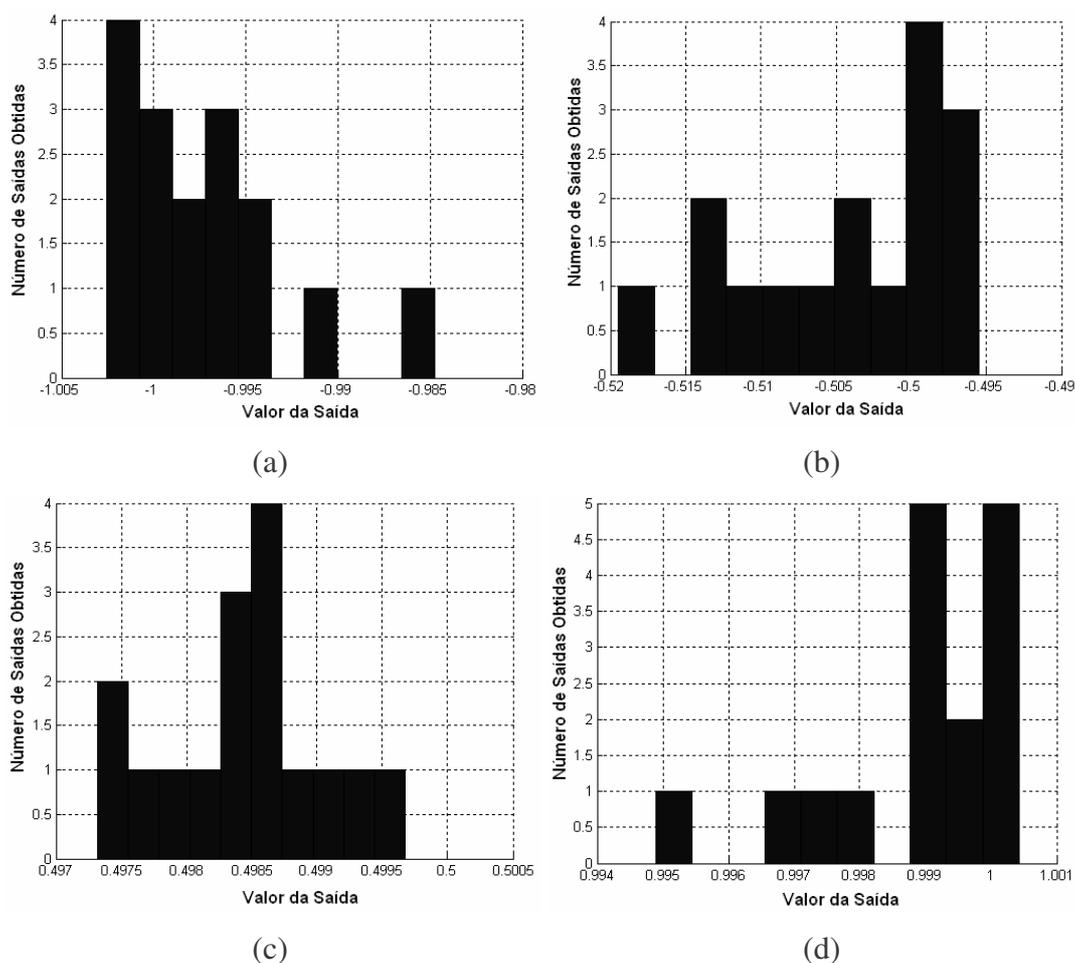


Figura 7. Resultados obtidos com a classificação de imagens diferentes das utilizadas na fase de treinamento com: (a) círculo de diâmetro 12; (b) círculo de diâmetro 20; (c) quadrado de lado 12; (d) quadrado de lado 20. Deve-se notar que os histogramas possuem intervalos diferentes no eixo “Valor de Saída”.

Tabela 2. Resultados da aplicação da rede neural treinada com um conjunto de imagens diferentes, porém com os mesmos *spots* utilizados para a síntese.

Tipo	Tamanho do <i>spot</i> (pixels)	Maior valor de saída	Menor valor de saída	Média dos valores de saída
Círculo	20	-1,005510	-0,985925	-0,995718
	16	-0,891800	-0,794650	-0,843225
	12	-0,511169	-0,492994	-0,502082
	8	-0,482046	-0,481796	-0,481921
Quadrado	8	0,203718	0,026777	0,115247
	12	0,501358	0,497762	0,499560
	16	0,931479	0,753768	0,842624
	20	1,001970	0,998231	1,000101

5. Comentários finais

Este trabalho apresentou um processo de classificação de texturas a partir da análise do espectro de potência por uma rede neural artificial.

O Modelo de *Spot Noise* permitiu a interpretação do espectro de potência e a definição dos momentos como forma de caracterizar a textura sendo analisada.

Os resultados mostram que não só a rede foi capaz de corretamente classificar texturas semelhantes às utilizadas na fase de treinamento, mas também conseguiu generalizar o aprendizado e classificar texturas com *spots* intermediários. Novamente aqui, o modelo de *Spot Noise* foi importante no sentido de permitir a definição, de forma quantitativa e objetiva, de uma ordem nas texturas sendo estudadas.

O processo de classificação apresentado pode ser utilizado em sistemas de análise por visão computacional que envolvem imagens com texturas aleatórias, tais: como grãos e outros alimentos em esteiras de carregamento, tecidos em fábricas e imagens de satélites.

Referências

- Basset, O., Buquet, B., Abouelkaram, S., Delachartre, P. e Culioli, J. (2000) "Application of texture image analysis for the classification of bovine meat", *Food Chemistry*, 69, p. 437-445.
- Chen, C. C. e Chen, C. C. (1999) "Filtering methods for Texture Discrimination", *Pattern Recognition Letters*", v. 20, p. 783-790.
- Christodoulou, C. I., Pattichis, C. S., Pantziaris, M. e Nicolaidis, A. (2003) "Texture-based classification of atherosclerotic carotid plaques", *IEEE Transactions on Medical Imaging*, v. 22, no. 7, p. 902-912.
- Haykin, S. (1999). "Neural Networks: a comprehensive foundation" 2 ed., Upper Saddle River: Prentice-Hall.
- Jain, A. K. e Karu, K. (1996a) "Learning Texture Discrimination Masks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 18, no. 2.
- Jain, A. K., Mao, J. e Mohiuddin, K.M. (1996b) "Artificial Neural Networks: A Tutorial", *Computer*, v. 9, no. 3, p. 31-44.
- Jernigan, M. E. e D'Astous, F. (1984) "Entropy-Based Texture Analysis in the Spatial Frequency Domain", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, v. PAMI-6, no. 2.
- Kachouie, N. N. e Alirezaie, J. (2003) "Texture Segmentation using Gabor Filter and Multi-Layer Perceptron", *Proceedings of the IEE International Conference on Systems, Man and Cybernetics*, v.3, p. 2897-2902.
- Madablushi, A. E. e Metaxas, D. N. (2003) "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions", *IEEE Transactions on Medical Imaging*, v. 22, n. 2, p.155-169.
- Martins, A .C .G, Rangayyan, R.M e Ruschioni, R.A. (2001) "Audification and sonification of texture images", *Journal of Electronic Imaging*, 10(3), p. 690-705.

- Ojala, T., Pietikäinen, M. e Nisula, J. (1996) "Determining composition of grain mixtures by texture classification based on feature distribution", *International Journal of Pattern Recognition and Artificial Intelligence*, 10 (1), p. 73-81.
- Paget, R. D. e Longstaff, D. (1996) "Terrain mapping of radar satellite images", *Electronic Imaging*, 6(2), p. 6-7.
- Quevedo, R., Carlos, L, Aguilera, J. M. e Cadoche, L. (2002) "Description of food surfaces and microstructural changes using fractal image texture analysis", *Journal of Food Engineering*, 53, p. 361-371.
- Rao, A. R. (1990) "A Taxonomy for texture description and identification", Nova York: Springer-Verlag.
- Weszka, J. S., Dyer, C. R. e Rosenfeld, A. (1976) "A Comparative Study of Texture Measures for Terrain Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-6, no. 4, p. 269-285.
- Wijk, J. J. (1991) "Spot Noise", *Computer Graphics*, 25(4), p. 309-318.