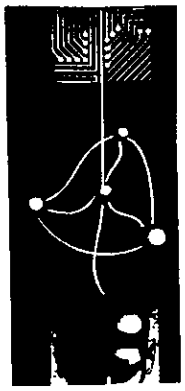

1° Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajubá
Itajubá, 24 a 27 de outubro de 1994

Uma Solução Backpropagation Invariante a Escala, Rotação e Translação para o Reconhecimento de Caracteres

Luiz Eduardo Seabra Varella [1,2]
Emmanuel Piseces Lopes Passos [2]
Márcio Azevedo Santos [1]
Ricardo Lomba de Araujo [1]

[1] PETROBRÁS-Petróleo Brasileiro S.A.
Depex-Departamento de Exploração
Av. Chile 65, sala 1402
20035 Rio de Janeiro, RJ, Brasil
g073@c53000.petrobras.anrj.br

[2] IME-Instituto Militar de Engenharia
Seção de Sistemas e Computação/Cartografia
Praça General Tibúrcio, 80
22290 Rio de Janeiro, RJ, Brasil
pass@lncc.bitnet

Abstract: Character recognition represents one of the steps of major complexity in systems designed for the recognition of graphic documents (maps). The method suggested in this paper uses neural networks to recognize characters with variations of scale, rotation and translation. The method primarily consists of the extraction of invariant moments of image. These moments are subsequently submitted to a system of Backpropagation (multi-layer Perceptron and Backpropagation learning) networks. Such system consists of 466 networks, which classify the image characters. Results thus obtained reveal a high degree of accuracy to the extent that 97% of trained patterns were duly recognized.

1.0-Introdução

A disponibilidade de equipamentos matriciais de alta resolução (*Scanners*) e o crescente aumento da velocidade de processamento dos computadores, vêm estimulando nos últimos anos a investigação de procedimentos destinados ao reconhecimento de informações pictóricas. Tais procedimentos permitem que gráficos impressos em papel sejam convertidos em arquivos digitais codificados em formato vetorial.

A transferência de gráficos impressos para arquivos digitais, com objetivo de se alimentar Sistemas Gráficos, permite não só uma maior segurança dos dados, mas também facilita a reprodução, manipulação e análise das informações. Neste sentido, o presente trabalho propõe um procedimento automatizado que visa agilizar a aquisição de grandes volumes de dados espaciais.

O texto do trabalho foi organizado em três partes. A primeira, introduz o problema do reconhecimento de caracteres nos sistemas de digitalização de documentos gráficos. A segunda, apresenta uma solução, fundamentada em redes neurais, para o reconhecimento de caracteres com variações de escala, rotação e translação. Finalizando, a terceira parte descreve os resultados alcançados com a implementação.

2.0-Sistemas de Reconhecimento de Imagens

De maneira geral, os sistemas que realizam reconhecimento em imagens definem três etapas bem distintas: Pré-Processamento da imagem; Extração de *features* e Classificação da imagem. O Pré-Processamento se constitui da eliminação dos ruídos da imagem digitalizada, do afinamento (geração do esqueleto), da vetorização e da classificação dos elementos gráficos em duas categorias (linhas e caracteres ou símbolos). A Extração de *features* tem como objetivo mapear um espaço de medidas (imagem) em um outro, denominado espaço de *features*, de dimensões reduzidas. O objetivo principal da transformação não é apenas a redução da dimensionalidade do espaço de medidas, mas também, a descoberta de características que permitam a classificação da imagem, em particular os textos.

O sistema de reconhecimento de caracteres, aqui proposto, inicia-se com o cálculo dos seis primeiros momentos invariantes, mapeando-se os caracteres da imagem em um espaço de seis variavéis. Em seguida, estes momentos são aplicados em um sistema de redes neurais, fundamentado no paradigma *Backpropagation*, que responde a este estímulo de entrada com a classificação dos caracteres. As sub-seções seguintes descrevem, respectivamente, o extrator de característica utilizado e a arquitetura de redes neurais adotada para o reconhecimento de caracteres.

3.0-Momentos Invariantes

O uso dos momentos invariantes para extração de características fundamentais é uma técnica que remonta a alguns anos, sendo introduzida inicialmente por [Hu (1962)]. Sua utilização neste trabalho foi motivada pelas publicações de [Dudani (1977)] e [Khotanzad (1988)].

Conceitua-se por momentos invariantes o conjunto de funções não lineares invariantes à escala, translação e rotação, obtidas a partir dos momentos geométricos da imagem. Dada uma imagem digital g(x,y) de dimensões M x M, $\{g(x,y), x,y = 0,\ldots M\text{-}1\}$, o $(p+q)$-ésimo momento geométrico é dado por:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} x^p y^q g(x,y) \quad \text{para } p,q = 0,1,2,\ldots .$$

Como são tratados caracteres de diferentes tamanhos, faz-se necessário a normalização do domínio de m_{pq}, que é feito mapeando-se a imagem plana M x M em um quadrado definido por $x \in [-1,+1]$ e $y \in [-1,+1]$. Conseqüentemente, a definição de m_{pq} fica:

$$m_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} x^p y^q g(x,y).$$

Fazendo os momentos invariantes à translação, têm-se a seguinte formulação para os momentos centrais da imagem:

$$\mu_{pq} = \sum_{x=-1}^{+1} \sum_{y=-1}^{+1} (x-\bar{x})^p (y-\bar{y})^q g(x,y), \text{ onde}$$

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \text{e} \quad \bar{y} = \frac{m_{01}}{m_{00}}.$$

Normalizando os momentos centrais a fim de torná-los invariantes à escala chega-se a:

$$\eta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^{\gamma}}, \text{ onde } \gamma = \frac{p+q}{2} + 1.$$

O conjunto de seis funções não lineares, invariantes à escala, rotação e translação, é definido sobre η_{pq} da seguinte maneira:

$$\phi_1 = \eta_{20} + \eta_{02} ;$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4(\eta_{11})^2 ;$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 ;$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 ;$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] ;$$

$$\phi_6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}) .$$

Como os valores de ϕ_1 a ϕ_6 são extremamente pequenos, é comum, ao final do cálculo dos momentos, aplicar-se a função logarítmica a fim de se evitar problemas de precisão numérica. Logo, as características extraídas dos caracteres são definidas por: $\log |\phi_i|$, $i = 1,...,6$.

4.0-Descrição do Método de Reconhecimento de Caracteres

Antes de se descrever a solução, é de grande importância que se analise alguns aspectos. O primeiro deles refere-se à geometria e disposição dos padrões. Os textos de um documento gráfico apresentam-se de diversas formas, tamanhos e inclinações, portanto, deve-se chegar a uma solução que contemple tais características. O segundo diz respeito às espessuras. Vale lembrar que a imagem encontra-se representada por seu esqueleto, com todos os elementos apresentando espessuras unitárias. Esta particularidade deve ser levada em conta, uma vez que, a esqueletização da imagem produz sensíveis alterações na forma dos caracteres com inclinações não múltiplas de 45 graus. O terceiro, e último, refere-se à similaridade dos padrões. É importante que se desenvolva uma técnica que discrimine adequadamente caracteres semelhantes, por exemplo, I do 1, O do 0, B do 8.

Como em quase todos os sistemas neurais, baseados no modelo *Backpropagation*, faz-se necessário a realização de um conjunto de testes para se chegar a uma arquitetura definitiva. A título de simplificação para a implementação dos testes, foi digitalizado apenas um fonte de tamanho 0.4 cm, constituído de 36 caracteres. Para cada caracter foram geradas 13 imagens com inclinações múltiplas de 15 graus, variando de -90 à 90 graus, vide Figura 1. A seguir são descritos os testes realizados em um computador IBM/3090.

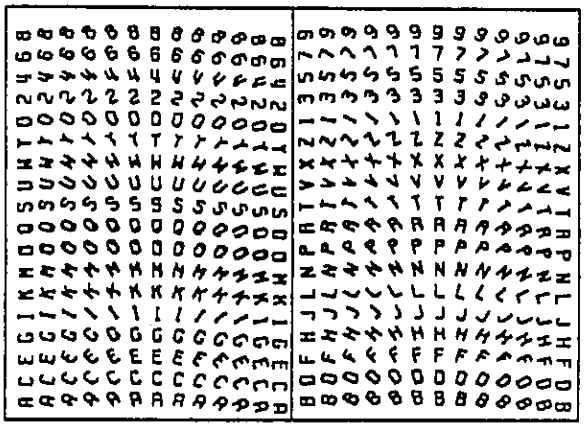

Figura 1 : Fonte Digitalizado

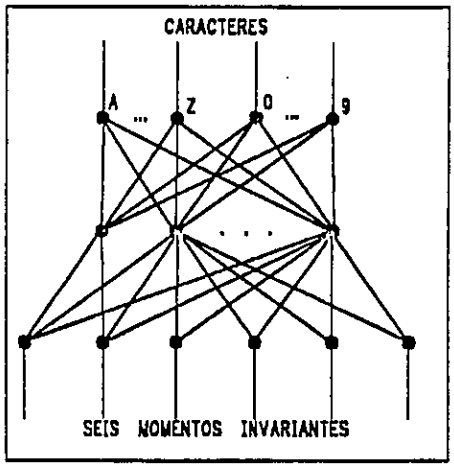

Figura 2: Teste 1 - Rede Única

4.1-Teste 1

O primeiro sistema que se pensou foi de apenas uma rede *Backpropagation* com seis elementos de processamento na camada de entrada, um elemento para cada momento invariante, e trinta e seis elementos na camada de saída, cada elemento respondendo por um caracter, vide Figura 2. Com intuito de se estimar um número inicial para as unidades de processamento da camada intermediária, foi treinado um conjunto de redes para o reconhecimento de caracteres com inclinação igual a 0 graus. Para este teste foram experimentadas 22 arquiteturas de redes, treinadas 20000 vezes com os padrões a 0 graus.

A tabela 1 mostra para cada número de elementos da camada intermediária o respectivo percentual de reconhecimento dos padrões

treinados. Foi adotada para a taxa de aprendizado, η, o valor 0.2 e para a constante *momentum*, α, o valor de 0.9. A realização deste teste consumiu aproximadamente 2h 10min de CPU.

De posse desse resultado, foram treinadas três redes com 250, 270 e 300 unidades de processamento na camada intermediária, visando o reconhecimento dos caracteres com todas as inclinações. Do conjunto de treinamento, 468 amostras, Figura 1, foram eliminados os dois W da extremidade (-90 e 90 graus) por serem idênticos aos padrões da letra M (90 e -90 graus). O treinamento destas redes consumiu mais de 32h de CPU, sem contudo apresentar resultados que justificassem o aprimoramento da solução. Este longo tempo de treinamento, sem dúvida, inviabiliza qualquer solução neste sentido.

Tabela 1 : Resultados do Teste 1.

Un. Inter.	% Rec.	Un. Inter.	% Rec.	Un. Inter.	% Rec.
10	2.5	108	61.5	216	71.7
16	2.5	120	66.6	230	82.0
32	10.2	130	79.4	250*	97.4*
40	7.6	140	71.7	260	87.1
56	25.6	150	74.3	280	84.6
68	30.7	170	79.4	300	79.4
80	51.28	186	71.79	-	-
92	48.71	200	82.05	-	-

4.2-Teste 2

Uma segunda alternativa foi a implementação de 36 redes, uma para cada caracter, com seis elementos de processamento na camada de entrada (seis momentos invariantes) e dois elementos na camada de saída, indicando a aceitação ou rejeição do padrão de entrada, vide Figura 3. Segundo esta arquitetura, o reconhecimento é feito aplicando-se os seis momentos invariantes do padrão em todas as redes. Conseqüentemente, a classificação é definida pela rede que apresentar o maior valor de ativação do elemento de saída responsável pela aceitação. Para o teste, as 36 redes foram treinadas com 466 amostras, sendo apresentadas 5000 vezes. A taxa de aprendizado (0.2) e o *momentum* (0.9) foram os mesmos do teste anterior. Para cada uma das redes foi pesquisado o melhor número de elementos da camada intermediária.

A tabela 2 mostra o resumo do teste, apresentando, para cada rede, o número de unidades escondidas e os percentuais *aproximados* de reconhecimento, indicados pelos escores de aceitação e rejeição dos padrões. Nesta tabela, os escores de aceitação e rejeição não expressam o resultado global do reconhecimento. Estes valores indicam que, por exemplo, para a rede que corresponde a letra A, 100% dos padrões A são reconhecidos e 99% dos demais padrões são ditos não A. Este índice de rejeição indica que existem alguns padrões não A que são classificados como A. Como apenas quatro redes apresentaram os percentuais ideais, ou seja, 100% para aceitação e 100% para rejeição, e algumas delas mostraram valores inferiores a 70%, é de se esperar que o resultado final aponte para a busca de uma outra solução. Apesar disso, 68% das amostras treinadas foram reconhecidas e o tempo de treinamento das 36 redes foi bem inferior (1h 25min) ao do teste de uma única rede, mostrando que é mais rápido treinar um conjunto de pequenas redes do que apenas uma com elevado número de elementos de processamento.

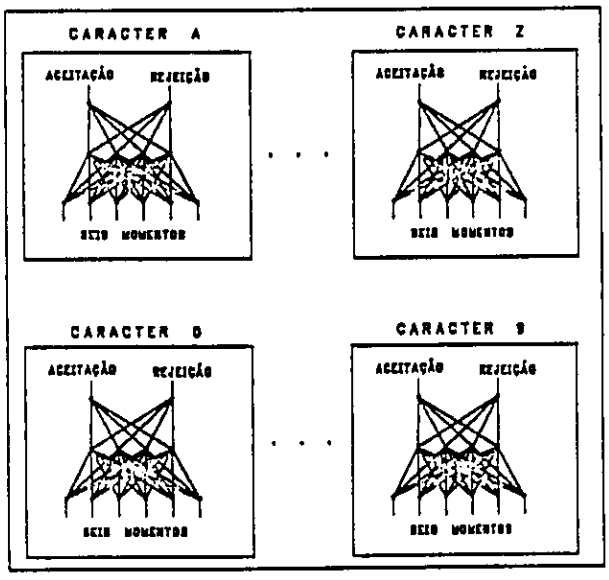

Figura 3: Teste 2 - 36 Redes.

Tabela 2 : Resultados do Teste 2.

Rede	% Ac.	% Rej.	Rede	% Ac.	% Rej
A-8	100	99	B-8	100	98
C-8	100	100	D-24	100	97
E-9	76	100	F-9	100	98
G-9	100	100	H-8	100	99
I-8	100	98	J-9	53	96
K-9	100	99	L-8	38	94
M-8	100	98	N-8	100	100
O-10	84	98	P-8	84	98
Q-8	84	95	R-8	100	100

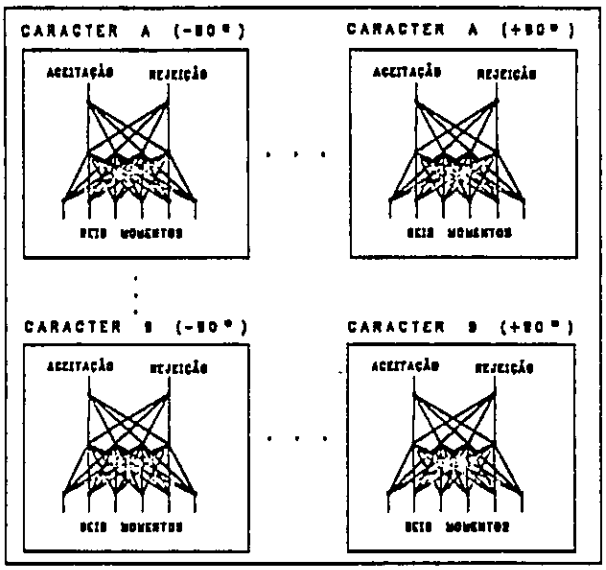

Figura 4: Teste 3 - 466 Redes.

4.3-Teste 3

Continuando nesta linha, com o propósito de se aumentar o percentual de reconhecimento, foi treinado um conjunto de 466 redes (36 caracteres * 13 inclinações de 15 em 15 graus - 2 W), com arquiteturas idênticas: seis elementos de processamento na camada de entrada, quatro elementos na camada intermediária e dois na camada de saída. Cada uma destas redes é responsável pelo reconhecimento de apenas uma inclinação de um dado caracter. A Figura 4 mostra o esquema de arquiteturas do sistema neural.

Quanto à estratégia de reconhecimento, o sistema segue o mesmo procedimento do teste anterior, ou seja, aplicam-se os momentos invariantes em todas as redes, e aquela que apresentar o maior valor de ativação no elemento de aceitação do padrão, classificará o caracter. O treinamento deste sistema consumiu 6h 25min de CPU. Para cada rede foram apresentadas 5000 vezes o conjunto de amostras, com a taxa de aprendizado igual a 0.2 e *momentum* igual a 0.9.

O resultado deste modelo, com 466 redes, mostrou avanços significativos, reconhecendo 97% dos padrões treinados. De maneira a se medir a capacidade de generalização do sistema, isto é, o reconhecimento de padrões não treinados, foram realizados outros testes descritos a seguir.

4.4-Teste de Generalização

Foi então gerada uma massa de padrões com caracteres rotacionados de 7.5 graus a partir dos padrões treinados. Para este conjunto de padrões, "não treinados", o sistema respondeu reconhecendo 41% dos caracteres. Este índice deveu-se ao elevado intervalo de rotação aplicado em cada caracter (15 graus) do conjunto de treinamento. De maneira análoga, foi realizado um segundo teste de generalização com intervalos de 10 graus para o treinamento e de 5 graus para o reconhecimento. A generalização deste modelo atingiu percentuais da ordem de 59%. Estima-se que com intervalos de 5 graus para o treinamento, a generalização atinja valores superiores a 80%. A tabela 3 sintetiza os resultados destes testes de reconhecimento dos padrões não treinados, mostrando para cada intervalo de rotação os respectivos percentuais de generalização.

Com relação ao reconhecimento de caracteres de diferentes tamanhos (escala), os percentuais se mantiveram no mesmo nível para variações até a metade do padrão treinado. Isto significa dizer que para o fonte utilizado nos testes, de tamanho igual a 0.4cm, pode-se reconhecer caracteres que variam de 0.2 a 0.6cm.

Tabela 3 : Teste de Generalização.

Int. Trein.	Int. Gener.	% de Rec.
15°	7.5°	41
10°	5°	59
5°.	2.5°.	> 80

5.0-Conclusões

Concluindo, o método aqui proposto para o reconhecimento de caracteres com variações de escala, rotação e translação se dá através de um conjunto de pequenas redes neurais, do tipo *Backpropagation*, onde cada uma delas é responsável pela classificação de uma única posição do caracter. Todas as redes são constituídas de seis elementos de processamento na camada de entrada, quatro elementos na camada intermediária e dois elementos na camada de saída.

É importante ressaltar que a acuracidade do reconhecimento é função da similaridade do conjunto de caracteres, da deformação produzida pela esqueletização, da resolução do dispositivo de aquisição da imagem e da capacidade de generalização introduzida ao sistema (quantidade de redes).

Referências

• L.E.S. Varella, Reconhecedor de Elementos Gráficos Digitalizados via Scanners, Dissertação de Mestrado, Instituto Militar de Engenharia, RJ, 1992.

• T. Pavlidis, Algorithms for Graphics and Image Processing, Computer Science Press, Rockville, 1982.

• K. Ramachandran, Coding Method for Vector Representation of Engeneering Drawings, Proc. of the IEEE, 1980, Vol.68, Num.7.

• D. Ting & B. Prasada, Digital Processing Techniques for Encoding of Graphics, Proc. of the IEEE, 1980, Vol.68, Num.7.

• L.A. Fletcher & R. Kasturi, A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images, IEEE Trans. on Patt. Anal. and Mach. Int., 1988, Vol.10, Num.6, PP.910-918.

• M. Hu, Visual Pattern Recognition by Moment Invariants, IRE Trans. Inform. Theory, 1962, Vol.IT-8, PP.179-187.

• S.A. Dudani "et alli", Aircraft Identification by Moment Invariants, IEEE Trans. on Computers, 1977, Vol.C-26, Num.1, PP.39-46.

• A. Khotanzad & J. Lu, Distortion Invariant Character Recognition by a Multi-Layer Perceptron and Backpropagation Learning, IEEE International Conference on Neural Networks, 1988, Vol.1, PP.625-632.

• P.K. Simpson, Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations, Pergamon Press, 1990.

• Y. Le Cun "et alli", Handwritten Digit Recognition Applications of Neural Chips and Automatic Learning, IEEE Commun. Mag., 1989, Vol. 27, Num. 11, PP. 41-46.

• D.E. Rumelhart, G.E. Hinton & R.J. (1986 b), Learning Internal Representations by Error Propagation, em D.E. Rumelhart & J.L. McClelland, Parallel Distributed Processing, 1988, Vol. 1, Cambridge, M.A., The MIT Press.