

Reconhecimento de Caracteres Através de Redes Neurais Artificiais com Aplicação a Placas de Automóveis

Danilo Rodrigues Fontana

Instituto Tecnológico de Aeronáutica – ITA
Rua dos Girassóis, nº 380, apto. 62, Jd. Industrias, São José dos Campos - SP
Bolsista PIBIC-CNPq
danilofontana@uol.com.br

Luiz Roberto Marim

Instituto Tecnológico de Aeronáutica – ITA
Rua Ernesto Soares Filho 130, Parque Espacial, São Bernardo do Campo - SP
marim@ita.br

Resumo. Os sistemas de reconhecimento de caracteres são desenvolvidos para reproduzir a capacidade humana de ler textos e identificar objetos ou padrões. Por sua vez, as Redes Neurais são elementos interconectados em forma paralela com as quais se procura imitar uma certa concepção de funcionamento do cérebro humano. Com essas redes voltadas ao reconhecimento de caracteres, busca-se o aprendizado para posterior reconhecimento de padrões que são provenientes da digitalização de imagens de textos. Tais sinais são armazenados na rede, como forma de aprendizado, para posterior consulta no reconhecimento. A pesquisa e coleta de informações sobre o processamento de imagens é o principal enfoque do trabalho, a extração de pontos ou caracteres de uma imagem, mostrou-se promissora com diversos métodos de análise sobre o problema. Dentre as atividades de pesquisa e de aprofundamento no ambiente MATLAB, foi desenvolvido o método de análise da imagem digitalizada da placa do carro e com ela reconhecer a região em que a placa está localizada.

Palavras chave: identificação de caracteres, redes neurais artificiais, placas de carro

1. Introdução

A placa do carro nasce e morre com ele, ela é a sua identidade. Quando o carro é vendido para alguém de outro estado, a combinação de letras continua com o veículo, o que muda é a tarja com o nome do estado/UF.

Cada estado tem uma série inicial e outra final. Sabendo de onde o carro vem, será mais fácil verificar sua procedência e descobrir possíveis impedimentos que o automóvel possa ter.

Um grande problema encontrado foi a diversidade de fontes e tamanhos dos caracteres das placas dos automóveis. Mesmo sendo algo regulamentado em lei, a variação do tamanho da placa, fonte e espaçamento dos caracteres é tal que dificultou a forma de seleção e preparação das entradas da rede neural [Justiça, Conselho Nacional de Trânsito-CONTRAN, 1967].

Cada imagem obtida com a câmera fotográfica apresentava a placa em uma região diferente devido a diversos fatores como modelo do automóvel, inclinação da placa, formato da placa, formato dos caracteres e o material como a placa é feita. Algumas das imagens mostram placas feitas de um material mais opaco em algumas regiões, fato que pode dificultar a visualização de radares sob alta velocidade, outro ponto a se ressaltar é que as imagens têm uma qualidade muito boa, tal fato dificulta o processamento feito pelos filtros e como foi citado em [Fontana, D., 2009], quando a qualidade da imagem obtida é muito boa, como no caso das câmeras digitais, é preciso usar filtros cada vez mais poderosos para eliminar impurezas [Gonzalez, 2000].

2. Aquisição da Imagem e Pré Processamento

Primeiramente, a imagem adquirida pela câmera digital apresentava 960 linhas dispostas por 1280 colunas. O fato de a câmera ser de boa qualidade gera alguns problemas como luminosidade, reflexos e a alta qualidade dos pixels associada à dimensão da imagem diminuem a velocidade de processamento.

Em uma segunda abordagem utilizamos uma câmera digital com menor resolução, que tirava fotos em tons de cinza, facilitando o tempo de processamento computacional e eliminando possíveis impurezas e diferenças de luminosidade. A matriz utilizada apresenta 768 linhas e 1024 colunas, dispostas em três bandas de 256 tons de cinza. Para reduzir o tempo computacional e facilitar o tratamento da imagem, foi feita a média ponderada dos valores das 3 bandas de cor, para obtermos apenas uma banda de tons de cinza.

Como a imagem em tons de cinza possui apenas uma banda de cor, temos uma matriz menor a ser processada e com isso temos um ganho computacional maior sem ter que prejudicar a qualidade da imagem.



Figura 1 – Foto do carro com a câmera digital

O passo seguinte é tornar essa imagem de tons de cinza em uma imagem binária, de zeros e uns, para que possamos então fazer a varredura na imagem para encontrar a região da placa e por fim segmentar os caracteres de forma que serão reescritos como uma matriz coluna para ser então devidamente apresentada a rede. Tais processos descritos acima agilizam o processamento da imagem fazendo com que tenhamos um ganho em tempo computacional e diminuam o espaço físico ocupado na máquina

3. MATLAB

Os trabalhos com o MATLAB R2007a vêm sendo desenvolvidos pelo grupo de pesquisa desde 2008 com o projeto do Sistema de Autenticação/Identificação Pessoal Biométrica da Palma da Mão com Auxílio de Redes Neurais Artificiais [4] que apresentou bons resultados. Devido à facilidade do domínio da linguagem e pela variedade de bibliotecas contidas no software podemos trabalhar em diversas áreas com ambientes fáceis de usar onde problemas e soluções podem ser expressos de forma matemática mais clara e concisa. Contudo, o domínio das técnicas do MATLAB, é de extrema importância na solução de uma enorme gama de problemas.

4. Filtros

A filtragem aplicada a uma imagem digital é uma operação local que modifica os valores dos níveis digitais de cada pixel da imagem considerando o contexto atual do pixel. Pela filtragem, o valor de cada pixel da imagem é modificado utilizando-se uma operação de vizinhança, ou seja, uma operação que leva em conta os níveis digitais dos pixels vizinhos e o próprio valor digital do pixel considerado [Fontana, D, 2009].

O processo de filtragem da imagem tem por objetivo limpar impurezas, realçar contrastes e eliminar o máximo possível de informações que para determinada utilização podem vir a não ter relevância, ou seja, para utilização deste projeto deseja-se eliminar o veículo e o ambiente a sua volta deixando apenas os detalhes da placa realçados.

4.1 Filtro Gaussiano

O filtro gaussiano tem a finalidade de “borrar” a imagem para descartar pequenos ruídos como falhas na figura, linhas pouco evidentes e algumas imperfeições formadas a partir da imagem em tons de cinza, o filtro gaussiano percorre toda a vizinhança da imagem em forma de quadrado, centralizando um pixel e “borrando” os demais, fazendo este percurso por toda a imagem.

Como dito anteriormente, este filtro faz uma varredura na imagem e cria certa dificuldade para que imperfeições e detalhes minúsculos sejam contabilizados como possíveis regiões que contenham a placa do automóvel.

4.2 Binarização

O método de binarização consiste em transformar os pixels da imagem em tons de cinza que passaram pelo filtro gaussiano em pixels com um valor de zero ou um. O processo seguinte consiste em tornar a imagem binária, ou seja, transformar todas as informações contidas na matriz em zeros e uns, onde **0** significa preto e **1** significa branco. Este é um passo importante para a implementação dos demais algoritmos de tratamento da imagem, pois muitos destes trabalham melhor com imagens na forma binária.

Após isso a imagem foi filtrada em sua forma binária por outro algoritmo do MATLAB, um filtro que buscava os limites da imagem a fim de recuperar a qualidade da imagem que perdemos ao binarizar a imagem e também para eliminar traços mais fracos que não se destacavam, para isso, também foi implementado uma técnica de afinamento das linhas que será vista logo abaixo.

4.3 Afinamento

O processo de afinamento da imagem reforça o algoritmo de binarização fazendo com que várias linhas que formavam uma linha grossa sejam substituídas por apenas uma linha, por isso esse processo é conhecido como afinamento. Este processo suaviza a imagem e retira boa parte das impurezas facilitando assim o processo de tratamento na imagem. A varredura da imagem é feita ponto a ponto, examinando a vizinhança e verificando quando o pixel pode ou não ser apagado.

Quando um pixel é apagado, seu valor na matriz muda de 1 para 0 e a imagem é dita afinada. Após ser afinada a imagem passa por um processo de “limpeza”, um algoritmo passa baixa busca todos os pixels isolados e os remove analisando sua vizinhança acima, abaixo e aos lados para saber se tem ou não um valor 1 correspondente nas casas, reforçando assim o trabalho de afinamento da imagem e melhorando a qualidade da mesma.

A figura abaixo mostra a imagem após todo o processamento realizado sobre a imagem da figura 1.

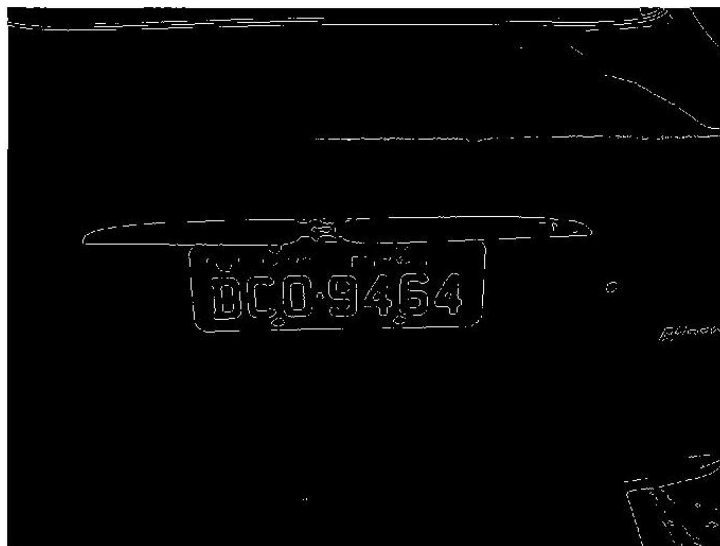


Figura 2 – Imagem do carro após a passagem dos filtros.

5. Redes Neurais Artificiais

A origem da teoria de Redes Neurais remonta aos modelos matemáticos e aos modelos de engenharia, de neurônios biológicos. Como qualquer célula biológica, o neurônio é delimitado por uma fina membrana celular que além da sua função biológica normal, possui determinadas propriedades que são essenciais para o funcionamento elétrico da célula nervosa. Basicamente o neurônio é composto de três partes fundamentais: o corpo celular, o axônio e os dendritos [Bishop, C.M., 1995].

O corpo celular, ou soma, é o centro dos processos metabólicos da célula nervosa, dele projetam-se extensões filamentosas, os dendritos, e o axônio.

Os dendritos são prolongamentos numerosos especializados na percepção dos estímulos do meio ambiente e na recepção de impulsos nervosos provenientes de outros neurônios. O papel do axônio é essencial uma vez que este é responsável pela condução de um impulso elétrico de uma célula a outra.

As manifestações elétricas de neurônios biológicos foram observadas pela primeira vez no século 19 por DuBois Reymond com auxílio de galvanômetros. No século 20, e mais especificamente nas duas décadas seguintes a 1944, passou-se a entender o neurônio biológico como sendo basicamente o dispositivo computacional elementar do sistema nervoso, que possuía entradas, muitas entradas e uma saída. As entradas ocorrem através das conexões sinápticas que conectam os dendritos aos axônios de outras células nervosas [Braga, 2007].

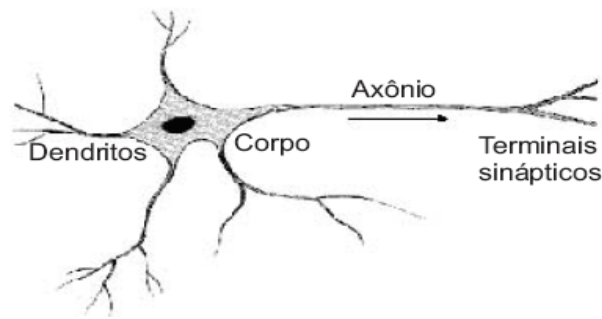


Figura 3 – Representação Simplificada de um Neurônio Biológico.

Os sinais que chegam por estes axônios são impulsos nervosos ou potenciais de ação, que constituem a informação que o neurônio processará, de alguma forma, para produzir como saída um impulso nervoso no seu axônio.

Dependendo da alteração no potencial elétrico da membrana, o resultado pode ser do tipo excitatório ou inibidor. Uma conexão excitatória contribui para a formação de um impulso nervoso do axônio de saída enquanto que uma conexão inibitória age no sentido oposto.

Em 1943, inspirado talvez pelos avanços na determinação experimental das propriedades eletrofisiológicas do neurônio, e pelos então recentes resultados de Alan Turing e John Von Neumann que indicavam ser booleana a natureza da inteligência, o que por sua vez resultava possivelmente da atividade neuronal, Warren McCulloch juntamente com o estatístico Walter Pitts publicaram no *Bulletin of Mathematical Biophysics* um artigo com o título: “*A Logical Calculus of the Ideas Immanent in Nervous Activity*” [Kovács, 2002] que é hoje a base para toda teoria das redes neurais artificiais.

O neurônio de McCulloch era um dispositivo binário: a sua saída poderia ser *pulso* ou *não pulso*, e as suas várias entradas tinham ganho arbitrário e poderiam ser *excitatórias* ou *inibitórias*. Para determinar a saída do neurônio, calculava-se a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação, positivos para aqueles que fossem excitatórios e negativos para aqueles que fossem inibitórios. Se o resultado fosse maior ou igual a um certo limiar então a saída do neurônio era *pulso*, e no caso o resultado fosse menor que o limiar era *não pulso*.

Assim, a essência da proposta de McCulloch e Pitts foi a seguinte: “*A inteligência é equivalente ao cálculo de predicados que por sua vez pode ser implementado por funções booleanas. Por outro lado, o sistema nervoso é composto de redes de neurônios, que com as devidas simplificações, tem a capacidade básica de implementar estas funções booleanas Conclusão: a ligação entre a inteligência e atividade nervosa fica estabelecida de forma científica.*” [Demuth, 1997].

Inicialmente, nesse trabalho foi implementada uma rede neural artificial do tipo *feedforward* devido ao algoritmo de *backpropagation*, ou seja, da retropropagação do erro para redes de neurônios de múltiplas camadas. Essa rede foi desenvolvida para reconhecer os números, tendo em sua arquitetura 3 camadas, dispostas de uma camada de entrada contendo 180 neurônios, uma camada intermediária com 30 neurônios e uma camada de saída com 10 neurônios na camada de saída. Tal arquitetura se mostrou eficiente pois eram atribuídos valores as respectivas posições da matriz de saída [Berg, 1993].

A cada imagem dos segmentos da placa, com 180 linhas, era atribuída uma saída com 10 linhas em que representava seu respectivo valor. As respostas então eram dispostas no formato da figura abaixo:

$$\begin{array}{c}
 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad 1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \dots \quad 9 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
 \end{array}$$

Figura 4 – Representação das saídas da Rede Neural Artificial

A rede neural foi treinada com vinte imagens de cada número, totalizando assim, duzentas imagens na matriz de entrada.

A fase de treino da Rede foi realizada com uma arquitetura do tipo *feedforward* com 180 neurônios na camada de entrada, 30 neurônios na camada oculta e 10 neurônios na camada de saída. Para o treinamento foi aplicada uma função com gradiente descendente. A função de ativação sigmóide tangencial (*tansig*) foi usada na rede na camada oculta e para a camada de saída uma função linear, também conhecida como *purelin*. A meta, ou o erro objetivo para o fim de treinamento escolhido foi de 8.10^{-3} e uma taxa de aprendizagem de 1.10^{-2} .

Para isso, foram realizados vários treinos e simulações que para os casos estudados, obtivemos os seguintes resultados:

Tabela 1 – Resultados obtidos pelas simulações da RNA

Dígitos	Porcentagem (%)	Dígitos	Porcentagem (%)
0	68	5	70
1	72	6	75
2	58	7	82
3	69	8	67
4	46	9	79

Tais resultados se mostram muito além daqueles esperados pelo grupo, uma vez que com maior capacidade computacional e com um aumento significativo das imagens no banco de dados, o número de dados a ser processado pela rede neural artificial traria uma melhora na qualidade dos resultados, favorecendo assim o desempenho de um sistema de reconhecimento de caracteres baseado em técnicas de rede neural artificial.

6. Algoritmo

O algoritmo desenvolvido baseia-se em todos os tratamentos descritos acima e tem o intuito de preparar a imagem para o input da rede neural. A dimensão inicial da imagem adquirida pela câmera é de 768 linhas por 1024 colunas, fato que consome muito tempo computacional e espaço físico na máquina. O processo mais trabalhoso é o de encontrar a região onde a placa está localizada, por isso, ao varrer a imagem temos que considerar diversos fatores como a inclinação, forma dos caracteres, tonalidade e forma como estes estão dispersos.

Embora exista uma lei que normaliza os padrões de letras e números presentes nas placas, algumas apresentam caracteres com fonte e tamanho diferentes do que manda a legislação, tal fato gera dificuldades pois padrões de letras distintos e com espaçamentos variados vem a dificultar mais e mais os efeitos da luminosidade e as falhas de tonalidades causadas pela deterioração do tempo e dos demais efeitos do meio ambiente.

As imagens feitas com a câmera de melhor resolução, que gerava uma matriz maior e com muitos detalhes devido à qualidade e a resolução adotada, apresentaram os seguintes resultados:

Tabela 2 – Quantidade de Imagens onde as placas são reconhecidas.

Banco de Dados de Automóveis até 22 de Janeiro de 2010	
Placas Reconhecidas	Placas Não Reconhecidas
55	16

Já com a nova câmera, ao varrer a imagem em busca de possíveis placas, temos uma região com 70 linhas e 360 colunas que com o presente banco de dados apresentou os seguintes dados.

Tabela 3 – Quantidade de Imagens onde as placas são reconhecidas.

Banco de Dados de Automóveis até 11 de Junho de 2010	
Placas Reconhecidas	Placas Não Reconhecidas
220	2

Com as imagens das placas isoladas é preciso agora isolar os caracteres, as 3 letras e os 4 números presentes na placa. Ao fazer o isolamento de cada caracter varremos a imagem de modo a demarcar 72 posições onde os caracteres podem estar presentes.

Logo, de cada placa extraímos 72 matrizes de 70 linhas por 40 colunas que apresentam as possíveis localizações da placa do automóvel.

Pelo fato de serem 7 termos que adquirimos da placa, para cada caracter, faz a varredura da imagem com um passo que determina a possível variação da posição do mesmo em relação à placa, isso se mostrou como uma solução para imperfeições na imagem, ângulo de inclinação da placa e também os efeitos da luminosidade, assim, imagens em que a foto não conseguiu focar a placa do carro ou apresentou um tamanho diferenciado devido ao ângulo de inclinação, podem ser reconhecidas e ter os caracteres extraídos.



Figura 4 – Foto da Placa do carro extraída após a passagem dos Filtros e varredura em busca da placa.

As imagens abaixo representam a extração dos caracteres da placa mostrada na Figura 3 e o passo do corte que é feito nas imagens de modo a obtermos 72 novas matrizes que condizem aos caracteres segmentados.



Figura 5 – Exemplos da placa segmentada.

No exemplo acima temos apenas 7 representações dos 72 cortes feitos na placa que nos ajudam a ajustar os caracteres para as placas que não ficam completamente enquadradas, auxiliando assim no processo de reconhecimento de caracteres e no melhoramento das chances de encontrar a placa. Para acelerar o processo de treinamento e simulação da rede neural, a imagem foi reduzida, ficando assim com 18 linhas e 10 colunas. Tal fato acelerou o processamento dos dados e ajudou a refinar os dados que seriam apresentados a RNA, fazendo assim, com que os dados de entrada para a rede neural fossem dispostos em matrizes de 180 linhas, em que cada placa gerava 32 matrizes, ao invés das 72 anteriores.

A placa dos automóveis pode ser identificada em 99 % dos casos estudados, sendo que o algoritmo foi capaz de identificar a placa em diversas condições. Dentre estas condições, o algoritmo realizou a varredura e detectou a placa em diferentes posições e distâncias.

7. Conclusões

Após desenvolver o algoritmo capaz de identificar as placas de automóveis nas imagens obtidas através da câmera digital, a taxa de reconhecimento das placas foi de 99%, sendo que dentre as placas não reconhecidas havia imagens de automóveis cujas placas estavam deterioradas, a ponto de não serem reconhecidas nem a olho nu. Um grande problema encontrado no projeto foi a imensa diversidade de fontes, tamanhos e formas dos caracteres encontrados nas placas. Em relação a posição da placa no automóvel, o algoritmo é capaz de encontrar a placa e selecioná-la para o estudo dos caracteres pela rede neural.

O trabalho de aquisição de dados para a implementação da rede neural consistiu em tirar fotografias de automóveis, em parte as imagens foram obtidas em estacionamentos de lugares públicos, porém, dada a dificuldade em adquirir imagens, associada a disparidade de letras devido a legislação nacional afirmar que no estado de São Paulo, as placas de automóveis tem série inicial definida por "BFA" e série final "GKI", o número de letras obtidas no banco de dados não se mostrou suficiente para aprimorar o treinamento da rede neural, fazendo assim com que os resultados da RNA não atingissem o erro desejado e o treinamento não convergisse.

Optamos por trabalhar com diferentes arquiteturas de redes para o tratamento das letras e dos números, uma vez que a rede do tipo *feedforward* não se mostrou eficiente para o reconhecimento das letras. Para tal, propomos a implementação de um mapa auto-organizável de Kohonen, em que dado o aumento do banco de dados, poderemos

montar uma rede neural em que teremos regiões de aglomerados de letras, obtendo assim, um centro de massa em que podemos definir a região de interesse para a identificação das letras.

8. Agradecimentos

Os autores agradecem ao CNPQ pelo auxílio financeiro e o incentivo a projetos científicos.

9. Referências

- Berg, A.; Muller, D.; Engel, P. Reconhecimento de Caracteres usando Redes Neurais, Universidade Federal do Rio Grande do Sul, curso de pós-graduação em ciências da computação. Processamento de Imagens por Redes Neurais. Porto Alegre, 1993.
- Bishop, C. M.; Neural Networks for Pattern Recognition, Oxford University Press, 1995.
- Braga, A. de Padua., Redes Neurais Artificiais, Teorias e Aplicações. Editora LTC, 2ª edição, Rio de Janeiro, 2007.
- Demuth, H.; Beale, M.; Neural Network Toolbox User's Guide Version 3.0, MathWorks, Inc, 1997.
- Fontana, D.; Marim, L.; Sistema de Autenticação/Identificação Pessoal Biométrica da Palma da Mão com Auxílio de Redes Neurais Artificiais, 2009.
- Gonzalez, R., Woods, Richard., Processamento de imagens digitais, Editora Edgard Blücher LTDA, 2000.
- Justiça, Conselho Nacional de Trânsito-CONTRAN, decreto de lei de 28 de fevereiro de 1967 <http://www.mj.gov.br/contran/>, 2009.
- Kovács, Z. L., REDES NEURAIAS ARTIFICIAIS, Fundamentos e Aplicações. Editora Livraria da Física, 3ª edição, São Paulo, 2002.
- MatLab, informações e suporte. Site acessado em 25 de Janeiro de 2010, <http://www.mathworks.com/technical-computing/>.
- WALTER, Marcelo. A Obtenção de Texturas na Síntese de Imagens Realísticas Num Ambiente Limitado. Porto Alegre, CPGCC-UFRGS, 1991.