Detecção Facial com Redes Neurais LVQ

Mário Barbosa e Tsang Ing Ren

[mbaj@cin.ufpe.br](mailto:mbaj@cin.ufpe.br), [tir@cin.ufpe.br](mailto:tir@cin.ufpe.br)

Universidade Federal de Pernambuco (UFPE)

A detecção facial é uma área importante no campo de processamento digital. Possuindo diversas aplicações práticas o que torna uma área de bastante interesse comercial. O objetivo deste trabalho é apresentar uma nova técnica de detecção facial a partir de imagens estáticas utilizando uma rede neural LVQ (Learning Vector Quantization). A implementação da técnica se inicia com o processamento da base de treinamento, extraindo suas características com um filtro Gabor, esses dados então são treinados em uma rede neural LVQ até atingir a convergência. Finalizado o treinamento da rede, o classificador é utilizado através de uma varredura piramidal para detectar as faces.

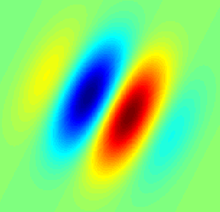
# Justificativa

Sistemas de detecção facial atualmente são bem difundidos, e podem ser encontrado facilmente em sistemas embarcados como câmeras digitais e *smartphones*. Um sistema de detecção facial para obter taxas relevantes de acerto precisa lidar com diversos problemas como múltiplas poses, iluminação não controlada e oclusão parcial por exemplo.

Para o uso em sistemas embarcados e de tempo real, é necessário uma técnica de classificação rápida, que é uma das principais características de redes neurais artificiais, apesar de possuir um processo custoso para o treinamento, porem com uma rápida classificação.

# Descrição da Técnica

As imagens são inicialmente processadas para extrair suas características. A técnica se baseia no filtro de Gabor, que consiste em um filtro linear bastante utilizado para a detecção de bordas. A frequência e a representação da orientação dos filtros de Gabor são similares ao sistema visual humano, e tem sido uma excelente ferramenta para representação de textura e discriminação [5].



Filtro de Gabor de duas dimensões

Em seguida, essas características são passadas como entrada de uma rede neural LVQ (*Learning Vector Quantization*).

A LVQ é um método de classificação de padrões que utiliza uma rede de camada única de neurônios artificiais posicionadas em duas ou mais dimensões para classificar as informações apresentadas em padrões ou classes distintas [6]. A LVQ utiliza aprendizagem do tipo supervisionado utilizando um conjunto de dados de entrada conhecidos [3].

Em uma rede LVQ cada vetor de entrada é apresentado a todos os neurônios. Um processamento de comparação é realizado em cada neurônio a fim de encontrar o neurônio que melhor responde o padrão, então o neurônio vencedor tem seus pesos atualizados seguindo as seguintes regras, caso o padrão que o neurônio representa seja da mesma classe do padrão de entrada, os pesos são atualizados para melhorar a resposta ao padrão, caso contrário são atualizados pra aumentar a diferença entre eles. Esse processo é descrito como uma aprendizagem por processo de competição, onde se sobressai o neurônio vencedor. Esse processo é repetido por diversos ciclos, e em cada ciclo a taxa de aprendizagem é reduzida para controlar a taxa de esquecimento da rede [3].

Após realizado o treinamento da rede neural, o sistema encontra-se pronto para ser utilizado, recebendo como entrada uma imagem que antes de tudo é processada para intensidades em escala de cinza. Então a imagem e varrida por uma algoritmo de multiscale piramidal que processa cada sub-imagem, extraindo as suas características com o filtro de Gabor descrito anteriormente e inserindo as características na rede neural para obter a classificação da sub-imagem. A rede então apenas responde se a sub-imagem dada na entrada é ou não uma face.

# Preparação da base de treinamento

Como já foi descrito, as imagens de treinamento não podem ser diretamente colocada para treinamento, o processo requer um pré-processamento antes.

Para deixar o sistema invariante a iluminação, as imagens de treinamento são processados por um método de filtragem baseado nos filtros de Gabor [4]. Inicialmente é gerado um conjunto de filtros de Gabor adequados para as frequências espaciais capaz de extrair características a partir de imagens afetada por diversas formas de iluminação. Os parâmetros utilizados por cada filtro é estimado a partir de *templates* de faces, empregando-se a transformada de Fourier. Cada filtro gera uma imagem filtrada que quantifica a frequência espacial definida no filtro.

A base de treinamento é processada sendo convertida para tons de cinza e reduzidas para o tamanho de 27 por 18 pixels. Em seguida é aplicado o filtro de Gabor a fim de extrair características invariante a iluminação. Com essas características, cada imagem é representada por um vetor de tamanho 2160.

Agora sim após processadas e convertidas para vetores de características estão prontas para serem inseridas na rede para o treinamento.

A base de dados de treinamento e testes foram oriundas da união entre três fonte de base de dados, duas com faces para amostras positivas e uma sem faces para amostra negativa [7, 8, 9]. Todas as faces positivas eram frontais, com diversas amostras de diferentes iluminações.

Para a conversão das imagens de treinamento e teste para níveis de cinza, e posteriormente a redução para os tamanhos especificados, foram utilizadas duas ferramentas, o Automator [11], que se trata de uma ferramenta de automação nativa do sistema operacional Mac que foi utilizada para selecionar as imagens e reduzi-las para o tamanho especificado e renomea-las de forma sequencial para a leitura do Matlab, depois foi utilizada a ferramenta ImageMagick [10] para converte-las para intensidades em tons de cinza.

# Implementação da Técnica

Para a implementação foi utilizado o software matemático Matlab em conjunto com o framework fornecido pela própria MathWorks, o Neural Networls Toolbox [1], e um algoritmo de filtro de Gabor [4].

Foram testados vários algoritmos de redes auto-organizáveis provenientes da Toolbox [2], mas foi preferível trabalhar com a Neural Networks Toolbox nativa do Matlab pela estabilidade, e confiabilidade.

Para a configuração da Rede Neural foram utilizados vários parâmetros até conseguir os que maximizassem a taxa de acerto da base de testes. Os melhores resultados foram obtidos com uma rede supervisionada LVQ com 40 neurônios na camada escondida, e uma distribuição de neurônios para representar os elementos positivos (face) e negativos (não face) no mapa auto-organizado de 60% e 40% respectivamente. A taxa de aprendizagem empregada foi de 0,01.

# Experimentos de Avaliação Técnica

O Neural Networks Toolbox possui várias ferramentas de avaliação. O gráfico de performance de erro por época de treinamento foi utilizando para acompanhar a evolução da convergência da Rede Neural.

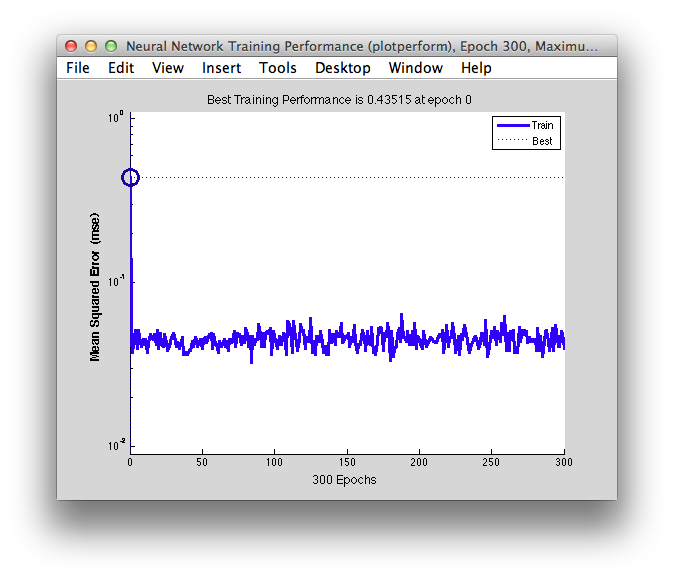
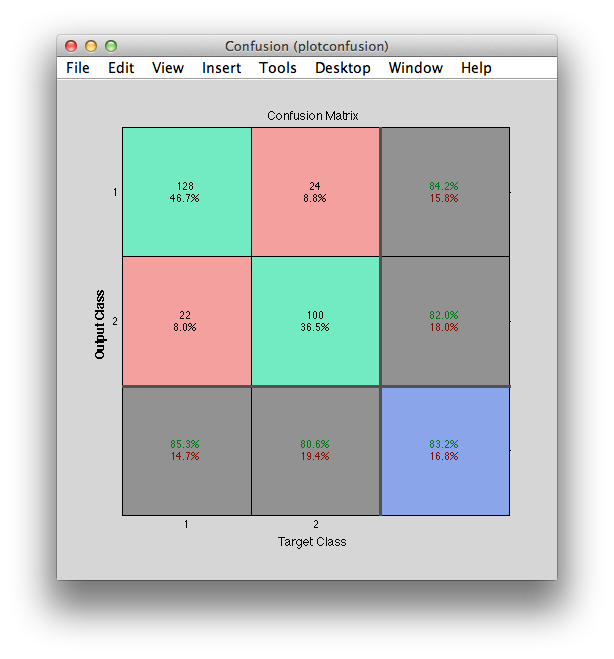


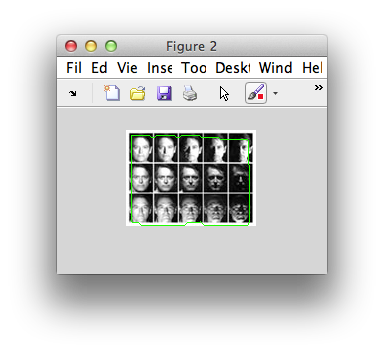
Gráfico de Performance de Treinamento

A matriz de confusão foi utilizada para avaliar as taxas de acertos por classe e global da base de treinamento e de testes. Os parâmetros das Redes Neurais eram modificados a fim de melhorar os resultados da matriz de confusão da base de testes.



Matriz de Confusão do Conjunto de Testes

Utilizando o sistema para detectar faces com imagens reais, ele apresentou um comportamento esperado, apresentando uma excelente invariância a iluminação porem com uma performance bastante afetada dependendo do angulo da face, só conseguindo detectar faces frontais.



Detecção Invariante a Iluinação

# Conclusão

O projeto ainda não se encontra finalizado, diversos problemas ainda precisam ser corrigidos. Apesar do sistema ser invariante a iluminação, resolvendo um dos principais problemas na área de detecção facial, ainda apresenta uma taxa de erro alta para encontrar faces que não estejam na posição frontal. Possuindo um excelente resultado na etapas de treinamento e testes, mas com taxas relevantes de erro com imagens reais, pois as imagens de treinamento e teste possuem apenas faces frontais, falhando em imagens com outras poses, não resolvendo completamente o problema de detecção.

Posteriormente esse trabalho será estendido para resolver um problema ainda maior, o de detecção de cabeças, adicionando ainda mais complexidade para ser resolvida, pois diferente de faces, que possui um padrão uniforme entre todos os padrões como o contorno dos olhos, uma cabeça possui muito mais possibilidades.

Referências

1. Matlab Trial Software (http://www.mathworks.com/products/matlab/).
2. SOM Toolbox for Matlab 5 (http://www.cis.hut.fi/somtoolbox/)
3. Learning Vector Quantization (http://en.wikipedia.org/wiki/Learning\_Vector\_Quantization)
4. Gabor Filter Algorithm (http://en.wikipedia.org/wiki/Gabor\_filter)
5. Neide Pizzolato Angelo, *Aplicação de Filtros de Gabor no Processo de Classificação de Imagens Digitais com Base em Atributos de Textura*, Porto Alegre, Brasil
6. Yasuo Kono, José Demiso Simões, Rafael Duarte. *Utilização de Rede Neual Para Previsão do Nível do Rio Paraguai*. São José dos Campos, Brasil
7. Department of Informatics and Mathematical Modeling. The IMM Face Database (<http://www2.imm.dtu.dk/~aam/>)
8. Psychological Image Collection at Stiling (http://pics.psych.stir.ac.uk/2D\_face\_sets.htm)
9. Negatives Database. Asian Institute of Technology. URL: <http://www2.cs.ait.ac.th/vgl/irshad/newsite/data>
10. Image Magick (<http://www.imagemagick.org/script/index.php>)
11. Software Automator (http://support.apple.com/kb/HT2488?viewlocale=pt\_PT)