

Projeção poligonal no auxílio à detecção de faces.

Adriano Martins Moutinho

Núcleo de Computação Eletrônica – Universidade Federal do Rio de Janeiro
adrianomm@posgrad.nce.ufrj.br

Antonio Carlos Gay Thomé

Núcleo de Computação Eletrônica – Universidade Federal do Rio de Janeiro
thome@ufrj.br

Resumo

Neste trabalho é sugerida uma modificação no processo de detecção neural de faces, proposto originalmente em [1], aplicando-se um método diferente para extração do vetor de características a ser apresentado ao modelo neural. O método de extração, denominado projeção poligonal, foi inicialmente sugerido em [7]. Este método se revelou proporcionar uma melhor discriminação entre padrões de face e de não-face.

Abstract

This paper suggests a modification in the neural face detection process as originally described in [1] by the introduction of a different method to extract the feature vectors to be presented to the neural model. This approach is based on an algorithm, called polygonal projection, initially introduced in [7]. The strategy provided a better discrimination among face and non-face patterns.

1. Introdução

Sistemas de segurança que exploram a possibilidade do reconhecimento de faces muitas vezes precisam lidar com a dificuldade inicial de localizar a posição da face em um contexto maior, distinguindo-a de outros objetos e do fundo.

Localizar faces é uma tarefa difícil, normalmente executada por um sistema à parte [1], cujo trabalho é separar as sub-partes de uma imagem que são candidatas à face, analisar a possibilidade de ser ou não uma face e a seguir descobrir a posição da mesma na figura original.

O sistema localizador de faces – *FaceLoc*, proposto neste trabalho e em [2] utiliza um sistema neural de detecção como base, conforme será descrito nas seções 2 e 3.

2. Visão geral do sistema localizador de faces

O processo de localização de faces implementado no *FaceLoc* é dividido em quatro etapas principais, como mostra a figura 1.

Figura original

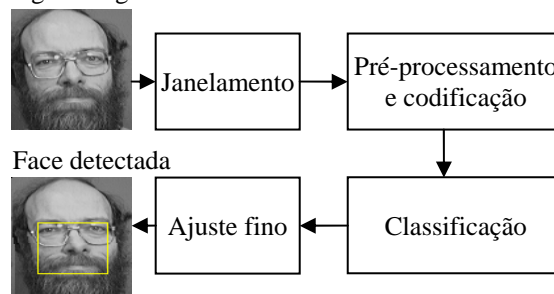


Figura 1 – Sistema Localizador de faces

O janelamento, descrito na seção 2.1, é o processo pelo qual a figura original é recortada em várias sub-imagens a serem avaliadas como candidatas à face.

O pré-processamento e a codificação, descritos na seção 2.2, consistem de um conjunto de algoritmos usados para a preparar e extrair o sinal a ser apresentado à entrada da rede neural. Nesta fase são aplicadas as novas técnicas de projeção poligonal descritas por este trabalho.

A etapa de classificação faz uso de uma rede neural do tipo Multi Layer Perceptron (MLP) [3], previamente treinada para distinguir o padrão facial [1]. Este bloco está detalhado na seção 2.3.

A quarta e última etapa do processo consiste de um ajuste fino onde se busca encontrar a dimensão e o posicionamento do retângulo que melhor circunscreva a face. Os procedimentos de refinamento são descritos na seção 2.4.

2.1 Janelamento

A partir da suposição de que qualquer parte da figura original, independentemente do seu tamanho e posição, pode conter uma face, o módulo de janelamento deverá extrair da imagem original todos os retângulos possíveis e distintos, enviando-os ao módulo seguinte. Em geral, para entrada em um modelo neural, estas janelas devem ser redimensionadas para um tamanho pré-determinado, que para este trabalho adotou-se a dimensão de 19x19 pixels.

Embora esta abordagem seja implementável, exigiria muito tempo de processamento, pois mesmo de uma imagem pequena pode-se extrair muitas sub-imagens retangulares distintas com sobreposição.

A abordagem finalmente adotada neste e em outros trabalhos prévios [2 e 8], extrai apenas janelas que contenham pelo menos um segmento de borda.

Mesmo segmentando a imagem apenas nos pontos contendo bordas, o número de sub-imagens ainda pode ser bastante elevado. Uma solução para minimizar esse problema é considerar inicialmente apenas janelas quadradas, e posteriormente fazer um ajuste fino, descrito na seção 2.4, nas imagens que forem reconhecidas como faces pela rede neural.

Após a segmentação todas as sub-imagens passam por uma seqüência de algoritmos de pré-processamentos, como será mostrado na seção 2.2.

2.2 Pré-processamento da imagem e codificação.

Para se obter um melhor desempenho na taxa de reconhecimentos corretos é normalmente necessário tratar a imagem antes de extrair os descritores que serão utilizados pela RNA [3].

No caso específico do reconhecimento de padrões faciais é necessário um processo que acentue as diferenças entre as imagens que contêm faces e as que não contêm [1].

Assim aplica-se ao banco de dados de imagens disponíveis uma seqüência de transformações incluindo a equalização de histograma [4] [5], máscaras ovais [1] [2], correção de iluminação [2] e a técnica de componentes principais (PCA) [3] [8].

Inicialmente é aplicada uma correção de iluminação. Este método, melhor descrito em [2], detecta a partir do gradiente imagem a parcela de iluminação que se supõe estar incidindo sobre a imagem.

A seguir o gradiente obtido é subtraído da figura original obtendo-se uma imagem com brilho ajustado em toda a sua extensão [2].

A motivação do uso deste tipo de correção é que embora diferentes tipos de iluminação possam modificar substancialmente uma figura, estas mudanças não são relevantes à classificação. Sendo assim, a aplicação da correção de iluminação tende a aumentar a generalização [1] [2].

A seguir é aplicada uma equalização de histograma, feita seguindo os métodos clássicos de equalização [4] [5], visando aumentar o contraste da imagem.

Finalmente, é aplicada a técnica de mascaramento oval [2], utilizada para retirar os pontos das bordas da imagem, pois os mesmos não são relevantes à classificação [1]. A figura 2 ilustra o processo de mascaramento oval.



Figura 2: Processo de mascaramento oval

A figura 3 mostra a seqüência completa do tratamento aplicado [2].

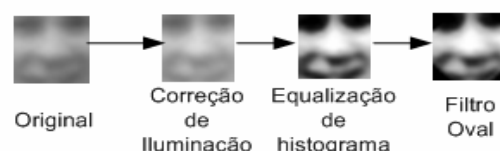


Figura 3 – Seqüência do Tratamento aplicado com vistas a obter maior generalização.

Antes da aplicação da estratégia dos Componentes Principais (PCA), usada para redução do espaço de dados [3] [8], é necessário realizar a extração de características.

Originalmente [1] [2], a extração de características era feita simplesmente colocando os valores de nível de cinza das linhas da figura lado a lado, de maneira a formar um vetor, levado então à técnica de componentes principais e finalmente à rede neural. Dessa forma, cada ponto da imagem era uma característica a ser levada em consideração pela rede neural.

Neste trabalho se propõe uma forma alternativa de extração de características com o intuito de tornar o sinal a ser classificado mais imune a variações de tamanho, posição e rotação. Esta nova técnica é conhecida como projeção poligonal de faces, sendo mais bem descrita na seção 3.

2.3 Rede Neural.

O uso de redes neurais para detectar faces foi originalmente proposto em [1]. Inicialmente, define-se face como o menor retângulo que corta da figura original os olhos, o nariz e a boca de um indivíduo, como mostra a figura 4.



Figura 4 – Exemplo de face.

O conceito básico da detecção neural de faces é utilizar os métodos de pré-processamento descritos na seção 2.2 com o intuito de fazer com que o padrão facial fique bem diferente do padrão não-facial. Para possibilitar a detecção, utiliza-se uma rede neural MLP [3] treinada previamente com os padrões mostrados na figura 5 (faces e não-faces).



Figura 5 – Padrões de face e não face [1].

2.4 Ajustes finos

Depois que a rede neural da seção 2.3 arbitra quais as figuras segmentadas no janelamento descrito na seção 2.1 correspondem a uma face, pode-se tentar fazer um ajuste mais fino para encontrar o retângulo melhor ajustado à face [1] [2].

Como o janelamento segmenta a imagem original com sobreposição, é bem provável que a face de um indivíduo seja detectada várias vezes de forma subsequente, em quadrados próximos.

Um possível ajuste fino que leva em consideração este fato é a redução por centróide, previamente descrita em [2] e [8].

Este método inicialmente descobre quais quadrados, detectados como faces pela RNA, têm o seu centróide dentro da área limitada por outros quadrados também detectados como faces, o que se pode chamar de “vizinhos de centróide” [2]. Todos os quadrados “vizinhos de centróide” podem ser reduzidos a apenas um quadrado a partir da média aritmética dos seus vértices.

Outro método de ajuste fino, também descrito em [2] e [8], é a recomparação sucessiva, onde se faz uma busca local com a rede neural para encontrar o máximo de saída da rede, o que se supõe obviamente ser a posição mais bem adaptada à face.

3. Extração de características

Consiste em extrair e selecionar descritores que melhor possibilitem a distinção de sub-imagens com padrões faciais daquelas sem padrões faciais.

3.1 Simples vetorização da imagem

Em primeira análise, uma possível extração de características, no que concerne à classificação de imagem, é colocar todas as linhas da figura lado a lado, de maneira a formar um vetor de treinamento.

Esta vetorização se faz necessária pois a rede neural MLP aceita como entrada vetores de valores considerados como características dos exemplos presentes no espaço de dados [3].

Embora a vetorização seja um método bastante utilizado não é a forma mais adequada, pois não permite nenhuma normalização quanto a qualquer variação de posição, rotação ou escalonamento da imagem.

Um simples deslocamento de um ponto para a esquerda em toda a figura, por exemplo, o que muito possivelmente não mudaria a classe de face para não-face, poderá fazer com que o vetor de dados mude completamente, complicando em muito o trabalho de classificação da rede neural.

3.2 Projeção poligonal de faces

Com o intuito de melhorar a generalização da rede neural, propõe-se um método de projeção poligonal adaptado de [7] para possibilitar o seu uso em imagens em níveis de cinza, ou seja, não binarizadas.

Em [7], um polígono circunscreve um padrão de caractere manuscrito e são extraídas diferentes distâncias do perímetro do polígono até o primeiro ponto preto do caractere. Este conjunto de distâncias é então normalizado e levado à rede neural para ser treinado ou classificado.

No entanto, este método não poderia ser diretamente aplicado no caso de segmentação de faces pela dificuldade em se arbitrar qual o primeiro ponto preto da imagem em uma determinada direção de projeção.

Para possibilitar o uso deste método, que permite um aumento na capacidade de reconhecimento e generalização [7], propõe-se um conceito de energia de projeção, a ser definido da seguinte forma:

A partir de uma borda lateral da imagem, escolhe-se uma direção de projeção e arbitra-se um valor qualquer para energia, conforme mostra a figura 6.

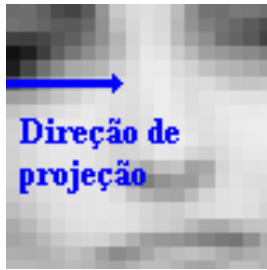


Figura 6 – Direção de projeção em uma face.

Arbitrando a energia inicial como sendo “1”, verifica-se o valor do nível de cinza dos pontos na direção de projeção, subtraindo um após o outro da energia inicial até que a mesma atinja zero ou menos. Neste ponto obtém-se a distância da projeção poligonal.

O processo de projeção poligonal se assemelha, com esta modificação, a forma com que são extraídas características de um objeto usando raios X [4]. Neste caso, um objeto é colocado à frente de um anteparo e uma emissão de raios X sensibiliza um filme. O filme será mais ou menos sensibilizado de acordo com as características bloqueadoras do objeto.

No caso da projeção poligonal, a distância extraída em cada projeção pode ser comparada a sensibilização de um filme. Níveis de cinza mais altos não permitirão a passagem da projeção, obtendo-se distâncias menores.

Uma adaptação do método criado para o uso em detectores de faces é fazer a inversão da imagem antes da projeção. Com isso, os valores originalmente mais escuros ganharão níveis mais altos dentro da matriz da imagem, conseguindo bloquear a projeção, enquanto os valores mais baixos, neste caso brancos, deixarão seguir a projeção.

A motivação para fazer esta projeção pode ser verificada na figura 5. Os olhos presentes em qualquer face são freqüentemente mais escuros do que o resto da face, o que também ocorre com a boca e menos freqüentemente com o nariz [2].

Uma outra adaptação ao método de projeção é elevar ao quadrado todos os valores da figura invertida antes de ser projetada. Com isso as seqüências de valores baixos na matriz de pixels, relativos aos pontos mais claros devido à inversão, não reduzirão em muito a energia da projeção, possibilitando que ela seja atenuada apenas pelos valores bem acima de 0,5.

Esta compressão nos valores mais baixos causada pela aplicação de uma função quadrática é semelhante e inspirada na forma com que são atenuados os raios X nos objetos a terem suas projeções examinadas no filme sensibilizado [4].

Assim como em [7], a escolha do polígono de projeção define as direções utilizadas. Neste traba-

lho, adaptou-se a projeção usando um quadrado como base, ou seja, serão extraídas as distâncias a partir de todo o perímetro do quadrado com direção ortogonal ao lado, como mostra a figura 7:

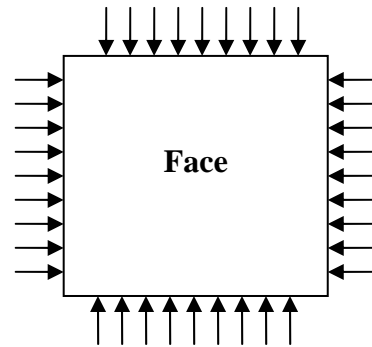


Figura 7 – Direções de projeção poligonal usando um quadrado como base.

Com o intuito de não perder a informação central da figura, muito possivelmente relevante à classificação, parte-se a figura em quatro pedaços e extraem-se novamente as distâncias de projeção, conforme mostra a figura 8.

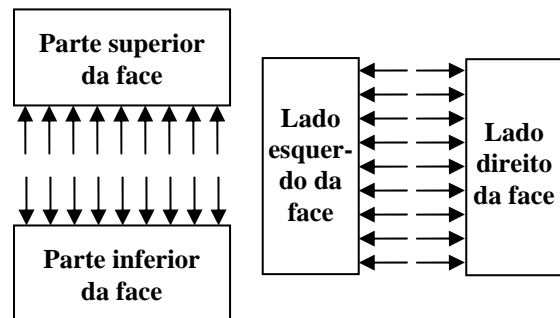


Figura 8 – Direções de projeção poligonal partindo-se a figura em quatro pedaços.

Como são extraídas as distâncias linha a linha e coluna a coluna em toda a extensão da figura, uma figura de tamanho $H \times V$ terá número de projeções igual a:

$$N_{\text{projeções}} = 4 \cdot H + 4 \cdot V$$

O número de projeções define o tamanho do espaço de dados, devendo ser tanto menor quanto possível para possibilitar treinamentos e processamentos mais rápidos da rede neural. Em uma figura quadrada onde $H = V$, a projeção poligonal proposta gerará menos características do que a simples vetorização com figuras maiores do que 8×8 . O banco de dados utilizado é de 19×19 , onde obtém-se um espaço de dados de 361 valores com a simples vetorização e igual a 152 com a projeção poligonal.

Embora o número de características seja maior com a simples vetorização dos pontos da imagem, o método de projeção poligonal extrai informações mais relevantes da imagem. Para provar esta característica pode-se usar a técnica de componentes principais (PCA) [3].

O uso de PCA permite projetar o espaço de dados em um novo espaço onde as variáveis são estatisticamente decorrelacionadas [6] [3]. Com isso, pode-se retirar variáveis de menor variância mantendo o máximo de informação possível. O tamanho do banco de dados será, bem possivelmente, menor, sem que se perca muita informação relevante [3].

A tabela 1 mostra a aplicação de PCA no banco de dados usando-se a simples vetorização e a projeção poligonal como forma de extração de características.

Vale ressaltar que os bancos de dados, após a aplicação dos dois métodos, foram normalizados antes da aplicação de PCA para possuírem média zero e variância unitária.

Pode-se notar, examinando a tabela 1, que se retirando apenas 1% da variância do banco de dados, usando simples vetorização como codificação, o tamanho do espaço de dados decai para apenas 58 características mais relevantes, mostrando que esta forma de representação não carrega muita informação.

Em contrapartida, a aplicação de PCA no mesmo banco de dados, mas agora usando a extração de características feita com projeção poligonal, não permite que espaço de dados diminua muito. Isso significa que as características extraídas com este método carregam mais informação, possibilitando treinamentos mais rápidos e possivelmente uma melhor generalização [3].

Tabela 1 – Aplicação de PCA no banco de dados

Testes / Extração	Simples vetorização	Projeção poligonal
Tamanho do banco de dados	361	152
Tamanho do espaço de dados com PCA a 1%	58	126
Redução em %	83,9%	19,2%

As figuras 9 e 10 mostram um exemplo de projeção poligonal de uma face. Os pontos brancos da imagem, que não está mostrada elevada ao quadrado

ou invertida, permitem a passagem da projeção, enquanto com os pontos pretos isso não ocorre.

Na figura 9 as quatro primeiras projeções são extraídas. Cada gráfico possui 19 pontos que indicam a distância obtida pela projeção. Na figura 11 são mostradas as projeções conseguidas partindo a figura em duas partes na horizontal e depois na vertical.

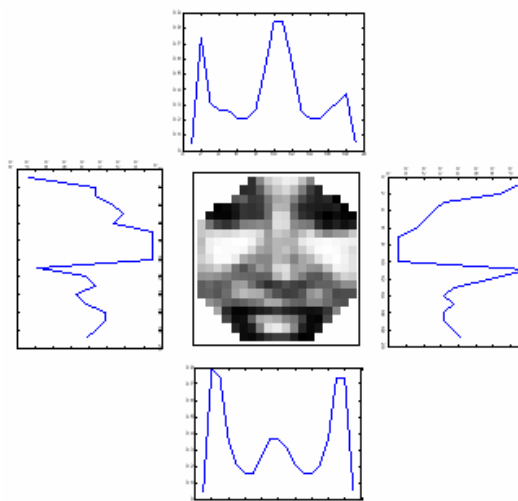


Figura 9 – Exemplo de projeção poligonal, quatro primeiras projeções.

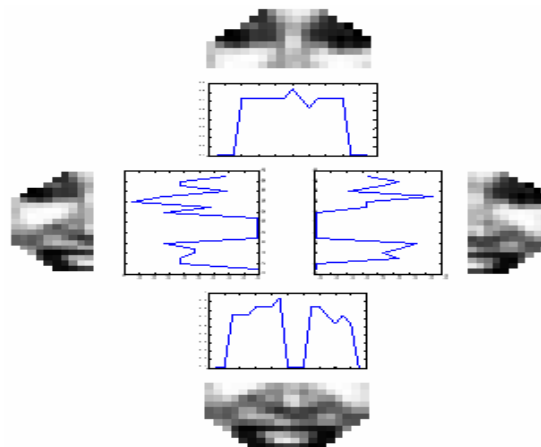


Figura 10 – Exemplo de projeção poligonal, projeções obtidas partindo-se a figura em duas partes na horizontal e na vertical.

4. Conclusões e Resultados

Para testar o método de projeção poligonal de faces proposto neste trabalho, foram criados 32 modelos de rede neural variando-se o número de camadas escondidas, a porcentagem de retirada de PCA, e as funções de ativação das mesmas [3]. As redes foram treinadas com taxa de aprendizado adaptativa e momento [3].

As 32 redes foram inicialmente treinadas usando o método proposto em [1] e [2], onde a extração de características é a simples vetorização.

Além disso, usando o mesmo banco de dados e as mesmas configurações de rede neural, um outro conjunto de 32 redes foi treinado usando como extração de características a projeção poligonal mostrada neste trabalho.

A tabela 2 mostra a média dos resultados obtidos pelas 32 redes. Os resultados mostrados são relativos ao banco de dados de testes, que não é utilizado para ajustar os pesos da rede neural durante a fase de treinamento [1].

Tabela 2 – Resultados obtidos, são mostradas as médias da taxa de reconhecimento de 32 diferentes redes usando os métodos indicados (banco de dados de testes).

Resultados	Classificações corretas (%), Banco de dados de faces.	Classificações corretas (%), Banco de dados de não faces.
Simple Vetorização	47,97%	96,83%
Projeção poligonal	65,84%	92,07%
Varição	Ganho de 17,87%	Perda de 4,76%

Como se pode perceber, de acordo com a tabela 2, houve um grande avanço na taxa de classificação correta no banco de dados de faces, da ordem de aproximadamente 17%, mostrando que o método consegue extrair características mais relevantes à classificação, tornando possível um aumento na generalização da rede neural.

No entanto, houve uma pequena queda na capacidade de classificação correta do banco de dados de não faces, cerca de 5%, com o uso do novo método. Esta queda não é tão relevante, pois é bastante pequena e o sistema localizador de faces tem a capacidade de eliminar elementos erradamente definidos como faces nos ajustes finos, descritos na seção 2.4 e em [2] [8].

Dessa forma, usando o método proposto, é possível melhorar a capacidade do sistema em reconhecer faces, obtendo uma pequena redução da capacidade de rejeitar padrões não-faces.

Como o banco de dados de teste, usado nos resultados da tabela 2, possui muitas faces rotacionadas, não totalmente frontais, e com posição dos olhos não totalmente padronizada, pode-se concluir que a extração de características permitiu uma certa imuniza-

ção contra estas variações, diminuindo o trabalho de generalização da rede neural.

Como trabalho futuro será proposto o exaustivo teste desse novo método nos sistemas de localização de face em figuras maiores, bem como o estudo das técnicas de janelamento para otimização do processo de localização de faces.

Referências:

- [1] Rowley, A. Henry; Kanade, Takeo. Neural network-based face detection. ISBN 0-599-52020-5, 1999.
- [2] Moutinho, Adriano M.; Thomé, Antonio Carlos Gay; Biondi, Luiz B. Coelho, Pedro H. G.; Meza Lúcia ângulo. Localização de faces usando redes neurais artificiais. Engevista, revista da engenharia da UFF, 2004.
- [3] Haykin, Simon. Neural Networks: A Comprehensive Foundation, Prentice Hall PTR, ISBN 0023527617 1994.
- [4] Jain. A. K. Fundamentals of digital image processing, Prentice-Hall Inc. ISBN 0-13-336165-9, 1989.
- [5] Gonzalez, Rafael C.; Woods, Richard E.; Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc. ISBN 0201508036, 1992.
- [6] Zurada, Jacek M. Introduction to Artificial Neural Systems. PWS Publishing Co., ISBN 053495460X 1999.
- [7] Vianna K. Gizelle; Rodrigues J. Roberto; Thomé, c. g. Antonio. Extração de Características para o Reconhecimento de Dígitos. SBRN, 2000.
- [8] Moutinho, Adriano M.; Thomé, Antonio Carlos Gay; Biondi, Luiz B. Coelho, Pedro H. G.; Meza Lúcia ângulo. Localização de faces usando redes neurais artificiais. International Conference Enterprise Information Systems (ICEIS), 2004.