

Detecção de Faces em Imagens Coloridas Usando Redes Neurais Artificiais

Adriano Martins Moutinho e Antonio Carlos Gay Thomé
Universidade Federal do Rio de Janeiro, Núcleo de Computação Eletrônica
adrianomm@gmail.com thome@nce.ufrj.br

Resumo

Encontrar faces humanas em imagens coloridas é normalmente uma das tarefas mais complexas em um sistema automático de reconhecimento. A razão de tal complexidade vem do fato de que a imagem de um rosto, mesmo frontal, pode possuir muitas variações de forma, cor, tamanho, proporção e iluminação.

Neste trabalho se descreve e propõe um sistema para localizar faces segundo uma abordagem relativamente simples e que tem mostrado resultados bastante promissores. O esquema idealizado faz uso das informações de cor, luminância e bordas extraídas da imagem, em conjunto com o uso de redes neurais artificiais [1] (RNA), que são a chave do processo de decisão.

Embora a maior motivação para o desenvolvimento deste trabalho seja os sistemas automáticos de segurança, pois todos devem possuir algum meio de encontrar a face antes de reconhecer a identidade do indivíduo em questão, também existem outras aplicações interessantes como os sistemas de compressão, que podem utilizar níveis de qualidade mais altos na região da face, e motores de busca que podem catalogar e indexar imagens que contêm faces.

1. Introdução

Encontrar a região que contém um rosto humano em uma fotografia em cores é mais difícil se não existem restrições quanto ao tamanho ou posição da face na imagem. Para isso, o sistema deveria procurar por diversas variações faciais, o que tornaria a busca ineficiente e possivelmente reduziria a chance de sucesso.

Neste trabalho, considera-se como face o menor retângulo que contém a área dos olhos, nariz e boca de um indivíduo [2] [3]. O sistema localizador proposto está preparado para encontrar faces frontais, como as da figura 1. Leves rotações ou translações na face serão toleradas, pois a rede neural está adaptada para generalizar algumas dessas diferenças.

O sistema detector de faces proposto é subdividido em vários módulos, como mostra a figura 2. O primeiro módulo do sistema é uma rede neural especializada em encontrar regiões com a cromaticidade típica da pele humana. O detector neural de pele, em conjunto com o

módulo detector de bordas, seleciona na imagem as áreas onde há maior probabilidade de conter faces, conforme descrito na seção 2. O segundo módulo do sistema é constituído por um segmentador, cuja tarefa é extrair sub-imagens da figura original nas áreas sugeridas pelo módulo descrito na seção 2, além de tratar os dados e apresentá-los a outra rede neural especializada em reconhecer padrões faciais, que constitui o terceiro módulo do sistema. O segmentador é descrito na seção 3 e o módulo neural na seção 4.

Por último, todas as regiões de face que tenham sido confirmadas pelo módulo neural são encaminhadas a um quarto módulo, de ajuste fino, que busca reduzir o número de regiões encontradas através de uma análise de redundância e melhor delimitar a face na figura. Este método está descrito na seção 5.



Figura 1 – Exemplos de faces

2. Detectores de indícios da presença de faces.

A rede neural descrita na seção 4, é projetada para reconhecer padrões faciais, bastando para tanto aplicar na imagem original o tratamento descrito em 4.1.

Para que o módulo neural funcione corretamente, a imagem original precisa ser subdividida em várias sub-imagens distintas, de forma a que a rede possa arbitrar quais destas correspondem a faces.

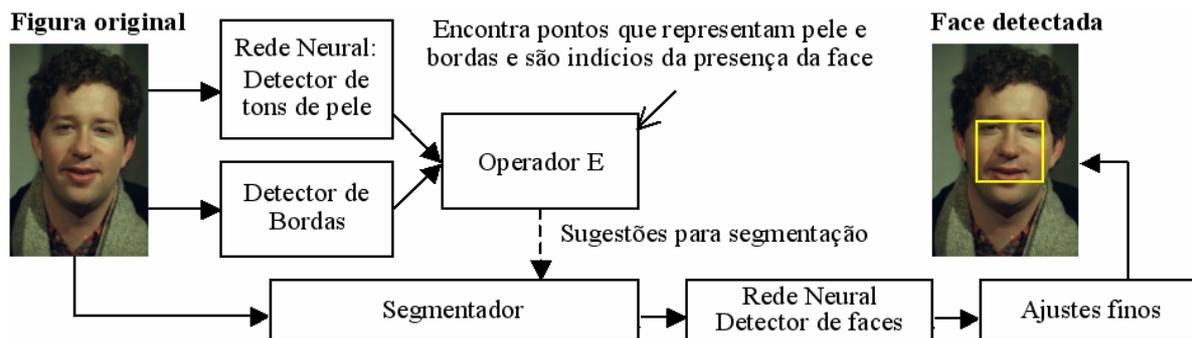


Figura 2 – Diagrama em bloco do sistema localizador de faces

A principal restrição a esta abordagem está no fato de que é sempre possível extrair muitas sub-imagens distintas, com sobreposição, mesmo de uma figura pequena. Partindo do princípio de que a face pode ter qualquer tamanho e proporção, um algoritmo de segmentação que seguisse esta abordagem se tornaria computacionalmente lento.

Propõe-se então que sejam consideradas apenas as regiões da imagem onde possam existir indícios da presença de uma face, diminuindo assim, a área de busca e tornando o algoritmo mais eficiente. Uma indicação de que pode haver uma face em uma figura é a presença de áreas contendo tons de pele e bordas, como mostrado nas seções 2.1 e 2.2.

2.1 Sistema detector de tons de pele.

A despeito das diferentes raças, os conjuntos de cores que representam possíveis tons de pele humana são finitos e podem ser razoavelmente delimitados. Partindo do pressuposto de que todo rosto humano certamente tem uma coloração de pele, pode-se tomar como premissa de que grandes áreas de pele indicam uma maior probabilidade de conter faces.

Detectores de tom de pele foram propostos por diversos trabalhos [5] [6]. Neste trabalho propõe-se uma implementação usando redes neurais artificiais (RNA) [1].

Para compor um banco de dados de informações sobre cor de pele, uma série de imagens contendo faces humanas foi obtida através de motores de busca de imagens na Internet [7]. Partes destas imagens foram extraídas fazendo-se uso de um esquema manual de segmentação, onde o usuário seleciona um ponto (pixel) contido em uma área onde se identifique a presença de pele.

A partir de cada ponto pré-selecionado, o sistema manual de segmentação busca por uma região conexa [8] composta por pontos cujos componentes vermelho, verde e azul (RGB) estejam dentro de uma região do espaço de cores definida por um limiar também especificado pelo usuário.

Cada região conexa extraída da imagem é então mostrada ao usuário para verificação final, e, uma vez confirmada, é adicionada a um banco de dados de tons de pele.

Este processo é então repetido para adicionar tons de pele e tons que não são de pele em um banco de dados, formando duas classes. Uma rede neural artificial [1] pode ser então configurada e treinada para discriminar regiões de pele e não-pele.

A rede neural estará então apta a decidir quais os pontos de uma nova figura têm cor de pele ou não. Para tanto, basta alimentar a rede com os valores RGB de toda a imagem, sendo que a decisão será tomada ponto a ponto.

Testar cada ponto da figura pode se tornar computacionalmente lento, dependendo do tamanho da matriz de imagem. Observou-se, no entanto, que o desempenho do algoritmo é muito pouco afetado quando se reduz a resolução da imagem antes de efetuar a busca pelos pontos de pele. Isto ocorre pois normalmente os pontos de pele são encontrados em grandes agrupamentos, o que permite que ainda sejam localizados mesmo com uma redução na resolução.

A figura 3 mostra o filtro de pele aplicado em uma figura de 728x512 do banco de dados de faces (FERET) [10]. As partes brancas da figura 3 (C,D e E) indicam áreas em que a rede neural classificou como pele enquanto nas partes pretas, a classificação é como não-pele.

Na figura 3 três detecções de pele são mostradas, com resolução total (C), ou seja, aplicado em todos os pixels, metade (D) e um quarto (25%) de resolução (E).

Para facilitar a detecção de pele em imagens com pouco contraste ou iluminação inadequada, é aplicado uma equalização de histograma, mostrado na figura 3-B. Como a imagem é colorida, deve-se tomar a precaução de aplicar a equalização de histograma apenas na componente de iluminação, o que pode ser feito convertendo-se o espaço de cores da imagem, originalmente em RGB, para o formato CIE L*ab [9].



Figura 3 – Teste de detecção de pele - Figura original (A), aplicação de correção de histograma (B), teste de pele em resolução total (C), meia (D) e um quarto de resolução (E).



Figura 4 - Aplicação do filtro de bordas (F) com detalhe da face. Filtro de pele (G). Pontos borda-pele (H).

Embora o resultado do filtro de pele (figura 3-A) possa não ter sido excelente, pois uma grande área foi definida incorretamente como pele, ainda é bastante útil porque certamente proporciona uma diminuição significativa na área de busca pela face.

O fato é que a detecção de pontos de pele é uma tarefa complexa, pois dependendo da iluminação e de objetos presentes na figura, pode-se ter a mesma cor típica de pele presente em áreas que não representem pele propriamente dita. Isso faz com que a total separação destes conjuntos seja muito difícil na prática [5].

2.2 Sistema detector de borda-pele.

Outra informação que indica maior probabilidade de encontrar faces é a presença de bordas na imagem [2] [3]. Os detalhes formados pelos olhos, boca e nariz são características presentes na face e geram bordas que podem ser detectadas com o uso de um filtro convencional de detecção de bordas do tipo *Canny* ou *Prewitt* [11]. A figura 4-F mostra a mesma imagem da figura 3-A após a aplicação de um filtro de bordas *Canny*.

Se as informações do filtro de pele forem cruzadas com as geradas pelo filtro de borda, pode-se conseguir um indício melhor das prováveis posições de faces, como mostrado na figura 4-H. Os pontos que tenham sido detectados por ambos os filtros, e que podem ser chamados de pontos de borda-pele, podem ser obtidos a partir de um

operador lógico **E (AND)** aplicado sobre as duas imagens.

Vale ressaltar que a informação dos pontos de borda-pele não é usada diretamente para detectar faces, mas para diminuir significativamente o espaço de busca como mostrado na figura 4-H. O módulo segmentador vai considerar apenas sub-imagens que contenham estes pontos, o que representa um número muito menor de regiões a serem avaliadas do que se a imagem inteira tivesse que ser considerada.

3. Segmentador

Este módulo é responsável por fragmentar a imagem original em diversas sub-imagens, e extrair de cada uma delas os descritores a serem apresentados ao classificador neural especializado em reconhecer padrões faciais.

Aplicando-se um tratamento estatístico sobre as imagens contidas no banco de dados de faces FERET [10], conseguiu-se obter um dado importante sobre a forma geométrica mais adequada para a segmentação de faces. A figura 5 mostra um histograma obtido após a marcação manual da posição da face em 2135 figuras do banco FERET de faces. A variável do histograma é a razão da largura pela altura da face. O resultado obtido mostra uma média próxima da unidade, ou seja, existe uma tendência das faces apresentarem uma geometria que se aproxima de um quadrado, o

que pode simplificar bastante o processo de segmentação.

O módulo segmentador, a partir dos indícios da posição da face obtidos conforme a seção 2.2, recorta quadrados de diferentes dimensões, centrados em nos pontos de borda-pele. Todos os quadrados são então redimensionados para um mesmo tamanho, definido aqui como 19x19, para serem então tratados e entregues à rede neural.

Para detectar faces de diferentes dimensões, o usuário do sistema deve especificar os tamanhos máximo e mínimo, em número de pixels, do quadrado a ser usado como máscara para segmentação. O módulo segmentador então recortará todos as sub-imagens com centro em todos os pontos de borda-pele, com os tamanhos de máscara desde o mínimo até o máximo, com incrementos também definidos pelo usuário.

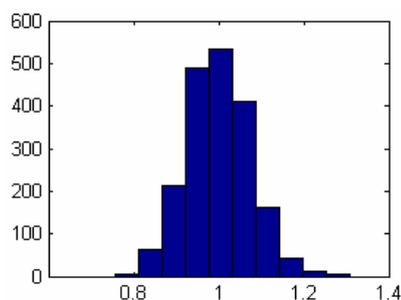


Figura 5 – Histograma da razão largura por altura em faces do banco de dados FERET [10]

Como o número de pontos de borda-pele ainda pode ser muito elevado, prejudicando o tempo de resposta do algoritmo, adota-se uma estratégia de agrupamento com o propósito de reduzir o número de vizinhos próximos, que tendem a gerar máscaras muito semelhantes.

4. Detector Neural de Faces

O uso de redes neurais para detectar faces foi originalmente proposto em [4]. O conceito básico da detecção neural de faces é utilizar os métodos de pré-processamento descritos na seção 4.1, com o intuito de fazer com que um padrão facial fique bastante diferente de um padrão não-facial. Para possibilitar a detecção, utiliza-se uma rede neural MLP [1] treinada previamente com os padrões mostrados na figura 6 (faces e não-faces) [4].



Figura 6 – Padrões de face e não face [4]

4.1 Pré-processamento das sub-imagens

Com a finalidade de obter o melhor desempenho do módulo neural de reconhecimento de padrões faciais, aplica-se um tratamento prévio em cada máscara da imagem extraída pelo módulo segmentador, descrito na seção 3.

O tratamento aplicado consiste em uma seqüência de transformações que inclui: equalização de histograma [11] [4], a aplicação de máscaras ovais [4] [2] e uma correção de iluminação [2].

Inicialmente é aplicada a correção de iluminação. Este método, melhor descrito em [2], calcula e gera um mapa com o gradiente dos tons de cinza que compõem a imagem original, e que se supõe ser representativo da parcela de iluminação a que a imagem foi submetida.

A seguir, o mapa de gradiente obtido é subtraído da imagem original, gerando uma nova imagem com brilho ajustado em toda a sua extensão.

O passo seguinte, visando aumentar o contraste da imagem, consiste em aplicar uma equalização de histograma, feita seguindo os métodos clássicos de equalização [11] [8].

Finalmente, é aplicada a técnica de mascaramento oval [2], utilizada para retirar os pontos das bordas da imagem, uma vez que os mesmos não são relevantes à classificação [4]. A figura 7 ilustra o processo de mascaramento oval.



Figura 7 - Processo de mascaramento oval

A figura 8 mostra o resultado obtido com a aplicação da seqüência de filtros. A imagem final está pronta para ser codificada e apresentada ao modelo neural de reconhecimento. Originalmente [2] [4], a codificação usada tinha como base os níveis de cinza da matriz de pixels da imagem final da figura 8.

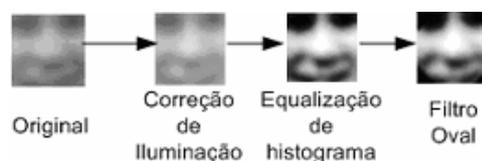


Figura 8 – Seqüência de filtros aplicados para obter maior generalização

No entanto, trabalhos posteriores [3] mostraram que pode-se obter uma melhor capacidade de detecção da RNA, com o uso de uma técnica de extração de características chamada projeção poligonal de faces. Esta técnica toma por base a medição da distância entre pontos de referência na

figura e a borda da imagem. Maiores detalhes sobre projeção poligonal de faces são encontrados em [3].

5. Ajustes finos

O módulo segmentador, descrito na seção 3, tende a extrair da figura original diversas sub-imagens muito próximas. Quando uma destas sub-imagens é reconhecida como face pela rede neural descrita na seção 4, as extrações adjacentes tendem também a serem classificadas como face, causando múltiplas detecções, como mostra a figura 9-A.

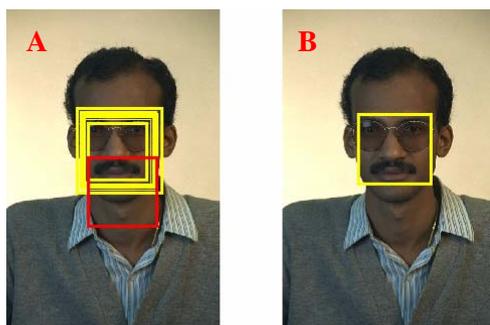


Figura 9 - Redução por centróide

Para obter uma única posição final mais provável da face, aplica-se uma técnica chamada redução por centróide [2] [3]. Esta técnica define como “vizinho de centróide” todo quadrado que tem seu centro de massa situado dentro da área definida por outro e vice-versa. A partir da definição, todos os quadrados “vizinhos de centróide” são selecionados e reduzidos a apenas um com a dimensão média de seus vértices.

As múltiplas detecções são comuns e podem ser usadas como meio de filtrar erros do sistema. Como as faces são detectadas múltiplas vezes, um quadrado que não possua “vizinhos de centróide” pode ser considerado um erro e ignorado. A figura 9-B mostra um caso onde um quadrado, marcado em vermelho, foi ignorado por este motivo.

6. Testes e desempenho do sistema

O sistema detector de faces proposto neste trabalho foi avaliado com o uso do banco de dados FERET [10]. Como este banco de dados de faces possui poses em perfil, uma separação foi feita para selecionar apenas as imagens contendo poses frontais, formando um banco de dados de 2135 imagens.

Com o intuito de criar uma ferramenta para calcular de forma automática o desempenho do sistema proposto, todas as 2135 figuras tiveram suas faces marcadas manualmente, ou seja, delimitou-se manualmente o menor retângulo que

contém a área dos olhos, nariz e boca de cada imagem.

Para comparar automaticamente a face detectada pelo sistema com a obtida manualmente, consideram-se os critérios mostrados na equação 1.

O primeiro critério considera que a interseção da área da face marcada manualmente com a face encontrada deve ser maior que 80%, ou seja, apenas 20% da área da face pode estar fora da área detectada. O segundo critério é para impedir que a área detectada seja muito maior que a área real da face, limitando a área adicional a no máximo a mesma área da face. Considera-se então, como acerto, os casos em que ambos os critérios forem atingidos.

$$\begin{aligned} \text{Area}_{\text{detect}} \cap \text{Area}_{\text{face}} &\geq 0,8 \cdot \text{Area}_{\text{face}} \\ \text{Area}_{\text{adicional}} &\leq \text{Area}_{\text{face}} \end{aligned}$$

Equação 1 - Critério de decisão para avaliação do sistema de detecção de faces

A tabela 1 mostra os resultados considerando o critério adotado, enquanto a figura 10 mostra alguns resultados do sistema.

Tabela 1 - Resultados com o banco de dados FERET

	Acertos	Erros	Total
Em número de figuras	1654	481	2135
Em %	77.47%	22.53%	Figuras

7. Conclusões

O sistema vem se mostrando eficiente na detecção de faces frontais em imagens coloridas, podendo ser estendido a imagens com mais de uma face. A figura 10 mostra algumas fotografias cujas faces foram encontradas pelo sistema.

Apesar de ainda possuir falhas, o sistema obteve um aumento bastante significativo na capacidade de reconhecimento com o uso de detectores de borda-pele, em relação ao que vinha sendo alcançado em [2] e [3]. Estes resultados promissores indicam que vale a pena investir em pesquisas futuras nesta área.

8. Referências Bibliográficas

- [1] Haykin, Simon. Neural Networks: A Comprehensive Foundation, Prentice Hall PTR, ISBN 0023527617 1994.
- [2] Moutinho, Adriano M.; Thomé, Antonio Carlos Gay; Biondi, Luiz N. Coelho, Pedro H. G.; Localização de faces usando redes neurais

artificiais. Engevista, revista da engenharia da UFF, 2004.

- [3] Moutinho, Adriano M.; Thomé, Antonio Carlos Gay. Polygonal Projection to help Face Detection. International Conference Enterprise Information Systems (ICEIS), 2005.
- [4] Rowley, A. Henry; Kanade, Takeo. Neural network-based face detection. ISBN 0-599-52020-5, 1999.
- [5] Jones, M. J. e Rehg, J. M.: 1998, Statistical Color Models with Applications to Skin Detection, Technical Report CRL-98-11, Hewlett Packard Laboratories
- [6] Brand, J. e Mason, J.: 2001, Proceedings of the 2001 IEEE International Conference On Image Processing.

- [7] Sistema de busca de imagens em www.google.com
- [8] Jain, A. K.: 1989, Fundamentals of digital image processing, Prentice-Hall, Inc.
- [9] Easy RGB homepage em www.easyrgb.com
- [10] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, October 2000.
- [11] Gonzalez, Rafael C.; Woods, Richard E.; Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc. ISBN 0201508036, 1992.

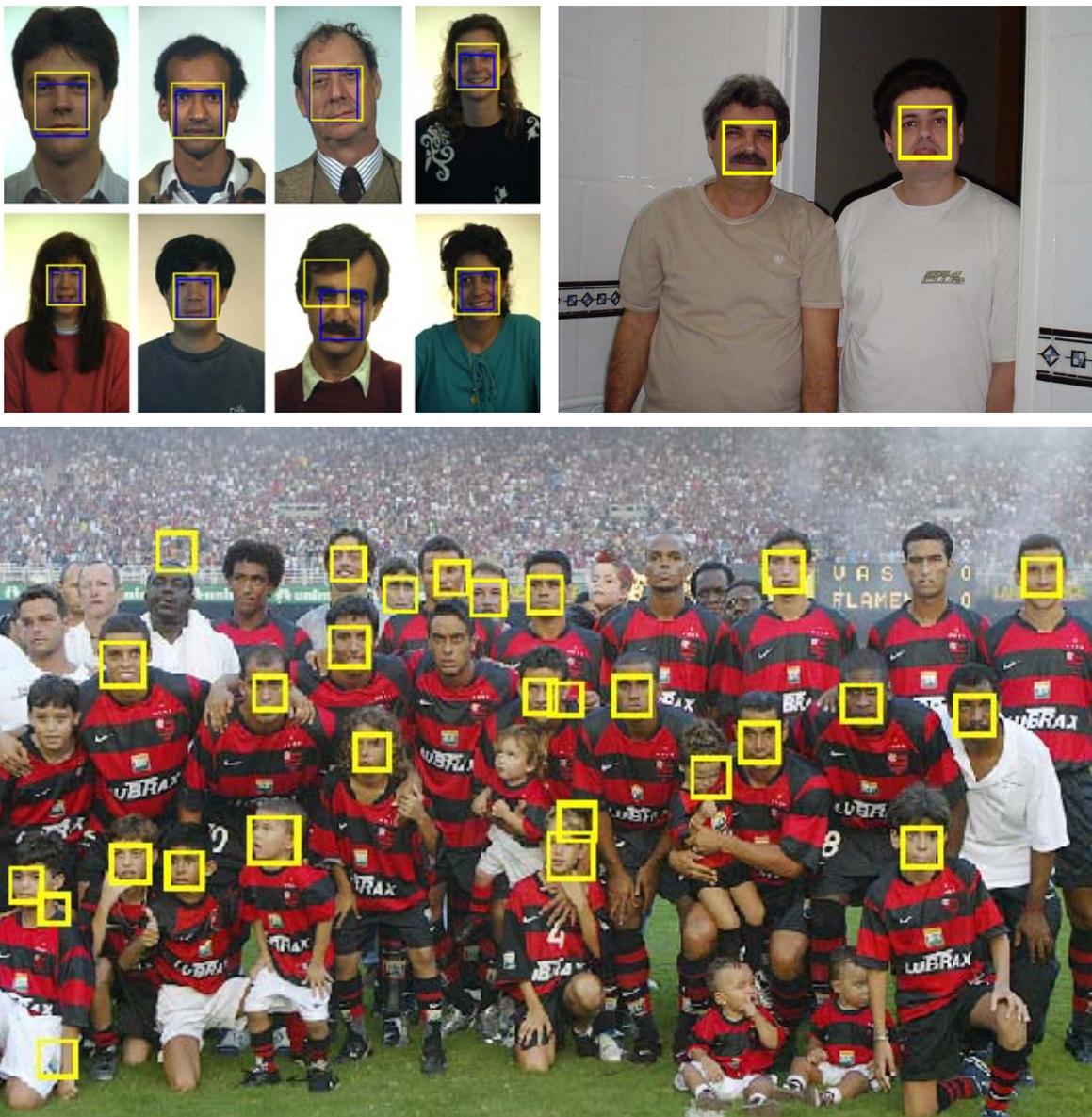


Figura 10 - Testes usando o sistema de detecção de faces. Bancos de dados FERET (esquerda superior, marcações manuais em azul e detectadas pelo sistema em amarelo). Fotografias contendo mais de uma face (direita e abaixo)