Um Sistema de Autenticação por Faces Usando Filtros de Correlação em Vídeos

José F. L. de Oliveira, Eduardo A. B. da Silva, Manuel A. P. Cardoso e Axel G. Hollanda

***Resumo*— Este trabalho é o resultado do início do desenvolvimento de um sistema para auxiliar deficientes visuais a reconhecer faces e objetos. Para se fazer um sistema deste tipo, o problema de reconhecimento de faces deve ser tratado e isto é feito empregando-se algoritmos de reconhecimento, tais como CFA – *Class-dependence Feature Analysis*, e uma *webcam*. O objetivo deste trabalho é combinar, aprimorar e desenvolver algoritmos de detecção e de reconhecimento de faces para elaborar um sistema em *software* capaz de detectar e reconhecer faces, cadastradas previamente em um banco de dados, obtidas a partir das imagens captadas por uma *webcam*. Para o caso específico do CFA, é proposto um algoritmo de seleção das imagens das faces que irão compor o conjunto de treinamento, que reduz o tempo de treinamento, eliminando a ocorrência de imagens redundantes. A redução no tempo de treinamento é de cerca de 80%, sem impactar a identificação, sendo bastante significativa. Além disto, algumas técnicas para aumentar a confiabilidade e a estabilidade da identificação do CFA também são propostas, as quais empregam a seqüência de vídeo capturada da *webcam* de forma muito simples.**

***Palavras-Chave*— Reconhecimento de Faces, Reconhecimento de Objetos, Filtros de Correlação**

***Abstract*— This work is the result of the begining of the development of a system for helping visually impaired people to recognize faces and objects. In order to make such a system, the problem of face recognion must be addressed and this is done by employing recognition algorithms, such as CFA – *Class-dependence Feature Analysis*, and a webcam. The objective of this work is to combine, improve, and develop algorithms for face detection and recognition so as to create a software-based system which is able to detect and recognize faces, previously enrolled in a databasis, obtained from images captured from a webcam. Specifically for the case of CFA, an algorithm for selecting the images that will compose the training set is proposed which reduces the training time by removing redundant images. The training time reduction is about 80%, without impacting identification performance, which is quite significant. Moreover, some techniques for increasing verification reability and stability are also proposed, that employ the video sequence captured from the webcam in a very simple way.**

***Keywords*— Face Recognition, Object Recognition, Correlation Filters**

I. INTRODUÇÃO

A demanda por algoritmos capazes de detectar e reconhecer faces e/ou objetos vem crescendo nos últimos anos. Tais algoritmos, além de serem usados em aplicações como controle de acesso e verificação de conformidade em linhas de produção, estão começando a ser empregados para auxiliar pessoas com deficiência visual a interagirem melhor com o ambiente ao seu redor, dando mais independência e segurança ao portador de deficiência visual. Um dispositivo, que será decorrente da evolução deste trabalho, capaz de atender a esta última classe de aplicações é a LDV – *Lanterna para Deficientes Visuais*. Em geral, o deficiente visual consegue identificar outra pessoa quando esta toma a iniciativa de se comunicar. Com a LDV, o portador de deficiência visual, além de ser capaz de se tornar ciente da presença de pessoas ao seu redor, poderá "reconhecê-las".

O objetivo deste trabalho é combinar, aprimorar e desenvolver algoritmos de detecção, tais como o de Viola-Jones [1], [2], [3], [4], [5], e de reconhecimento de faces, tais como o CFA (*Class-dependence Feature Analysis*) [6], para elaborar um sistema em *software* capaz de detectar e reconhecer faces, cadastradas previamente em um banco de dados, a partir das imagens captadas por uma *webcam* de baixa resolução, tal como 320×240 *pixels*. Apesar de haver diversos trabalhos publicados tanto na parte de detecção quanto na de reconhecimento de faces e objetos, são desconhecidos aqueles que detalham a implementação de um sistema como o que será descrito aqui. Optou-se por utilizar, inicialmente, o algoritmo Viola-Jones na parte de detecção por se tratar de um algoritmo estado-da-arte muito rápido e eficiente. Já na parte de reconhecimento decidiu-se desenvolver um algoritmo baseado no CFA. Evidentemente, outras técnicas reconhecidamente estado-da-arte, como o EBGM (*Elastic Bunch Graph Matching*) [7], [8] ou até mesmo o KCFA (*Kernel Class-dependence Feature Analysis*) [9], foram consideradas. Entretanto, a menor complexidade computacional do CFA favoreceu a sua escolha. O algoritmo SIFT (*Scale-Invariant Feature Transform*) [10], [11], [12] também é promissor em sistemas como este devido a sua generalidade, eficiência e relativa baixa complexidade. Um novo sistema de reconhecimento de faces e objetos baseado no SIFT já está sendo implementado, mas sua descrição foge ao escopo deste trabalho.

A seguir, na secão II, são apresentados conceitos básicos sobre filtros de correlação e, particularmente, sobre CFA. Na seção III, o pré-processamento das imagens das faces, que tem por objetivo reduzir o efeito das variações de iluminação e pose, é tratado. A seção IV é dedicada à apresentação dos métodos desenvolvidos para reduzir o tempo de treinamento do CFA e aumentar a estabilidade e a confiabilidade da identificação. Os resultados são apresentados na seção V e as conclusões, na seção VI.

II. CONCEITOS BÁSICOS

A tecnologia dos filtros de correlação é uma ferramenta básica para o casamento de imagens no domínio da freqüência [13]. Em métodos de filtros de correlação, variações normais em imagens de treinamento autênticas podem ser acomodadas pelo projeto de um arranjo no domínio da freqüência, chamado de filtro de correlação, que captura a parte consistente das imagens de treinamento, desenfatizando as partes inconsistentes, isto é, as freqüências inconsistentes. O reconhecimento de objetos é feito pela correlação cruzada de uma imagem de entrada com um filtro projetado, usando a FFT (*Fast Fourrier Transform*). Filtros de correlação também oferecem invariância ao deslocamento incorporada. Se a imagem de entrada é transladada com relação às de treinamento, o pico de saída será deslocado do mesmo valor. Este deslocamento pode ser utilizado pela correlação de saída para alinhar imagens. Outra vantagem de se utilizar filtros de correlação é

José Fernando Leite de Oliveira, Manuel Augusto Pinto Cardoso e Axel Guimarães Hollanda, Instituto de Tecnologia José Rocha Sérgio Cardoso, Distrito Industrial, Manaus, AM, CEP: 69.075-210, Brasil. E-mail: jleite@lps.ufrj.br; manuel@internext.com.br; aghi@lps.ufrj.br

Eduardo Antônio Barros da Silva, COPPE/PEE/LPS, Universidade Federal do Rio de Janeiro, Caixa Postal 68.504, Rio de Janeiro, RJ, CEP: 21.945-970, Brasil. E-mail: eduardo@lps.ufrj.br.

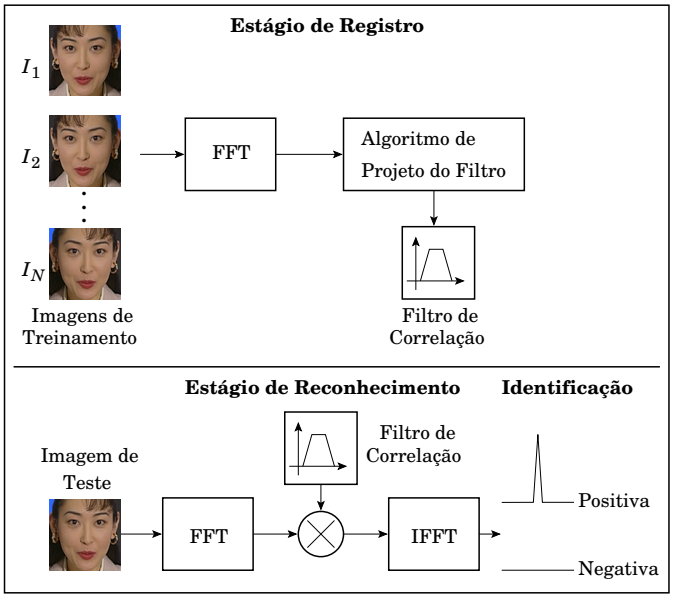

Fig. 1. O Conceito de reconhecimento de faces estáticas utilizando filtros de correlação.

que eles oferecem soluções de forma fechada as quais são computacionalmente efetivas [6].

A. Reconhecimento de Faces com Filtros de Correlação

O conceito básico do reconhecimento de faces estáticas usando filtros de correlação é mostrado na figura 1, adaptada de [6]. Há dois estágios: o de registro e o de reconhecimento. No estágio de registro, uma ou mais imagens da face de um indivíduo são obtidas. Estas imagens devem refletir a variabilidade esperada da imagem da face, devido à rotação, mudança de escala, mudança de iluminação, etc. As transformadas de Fourier 2-D (2-D FT, 2-D *Fourier Transform*) destas imagens de treinamento são usadas por um algoritmo de projeto de filtros de correlação para determinar um único arranjo no domínio da freqüência, chamado de filtro de correlação, que é armazenado. No estágio de reconhecimento, o usuário apresenta a imagem de uma face e a sua 2-D FT é multiplicada pelo filtro de correlação armazenado. A 2-D FT inversa deste produto resulta na saída correlacionada. Se o filtro for bem projetado, deve-se observar um grande pico na saída de correlação se a face foi reconhecida e nenhum pico discernível, caso contrário. A localização do pico indica a posição da imagem de entrada e, portanto, provê invariância automática ao deslocamento, sendo possível dispensar um estágio de centralização da mesma [6].

B. Filtro de Compromisso Ótimo

Uma forma de projetar o filtro de correlação é otimizar um ou mais critérios de correlação, sob as restrições do pico de saída de correlação c_j, o qual é o produto interno da imagem de treinamento e do filtro a ser determinado

$$c_j = \mathbf{x}_j^T \cdot \mathbf{h}, \tag{1}$$

onde $\mathbf{x}_j$ denota a j-ésima imagem de treinamento e $\mathbf{h}$, o filtro. O símbolo "T" denota o complexo conjugado transposto. Normalmente, faz-se $c_j = 1$ para imagens de treinamento da classe "verdadeiro" e $c_j = 0$, para as da classe "falso" [6].

Critérios diferentes levam a filtros com propriedades diferentes. O filtro *função discriminante sintética de variância mínima* (MVSDF *filter - Minimum-Variance Synthetic Discriminant Function* filter) minimiza a variância do ruído da saída de correlação $\mathbf{h}^T\mathbf{C}\mathbf{h}$, onde $\mathbf{C}$ é uma matriz diagonal cujos elementos C_{ii} representam a densidade espectral de potência do ruído na freqüência f_i. O filtro *energia de correlação média mínima* (MACE *filter - Minimum Average Correlation Energy filter*) minimiza a energia média da saída de correlação $\mathbf{h}^T\mathbf{D}\mathbf{h}$, onde $\mathbf{D}$ é a média de $\mathbf{D}_j$ que é a potência espectral da j-ésima imagem. $\mathbf{D}_j$ é uma matriz diagonal cujos elementos D_{ii}^j representam a potência espectral da j-ésima imagem de treinamento na freqüência f_i.

O filtro MACE enfatiza freqüências espaciais altas para produzir grandes picos de correlação, enquanto o filtro MVSDF, em geral, suprime as altas freqüências para obter tolerância ao ruído. Embora seja desejável atender a ambos os critérios, eles não podem ser minimizados simultaneamente. O filtro de compromisso ótimo (OTF - *Optimal Tradeoff Filter*) é projetado para balancear estes dois critérios, minimizando $\mathbf{h}^T\mathbf{T}\mathbf{h}$, onde $\mathbf{T} = \alpha\mathbf{D} + \beta\mathbf{C}$, $\beta = \sqrt{1-\alpha^2}$ e $0 \leq \alpha \leq 1$. O OTF é, então, dado por

$$\mathbf{h}_{\text{OTF}} = \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{c}, \tag{2}$$

onde $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}]$ é uma matriz $M \times N$ e cada $\mathbf{x}_j$ é a transformada 2-D de Fourier da j-ésima imagem de treinamento na forma de um vetor M-dimensional [6].

C. Derivação da Solução do Filtro Ótimo

A solução dada pela equação 2 é obtida encontrando-se o filtro $\mathbf{f}$ que minimiza o valor da função $\mathbf{\Phi}(\mathbf{f}) = \mathbf{e}^T\mathbf{e}$, onde $\mathbf{e} = \mathbf{c} - \mathbf{Y}^T\mathbf{f}$. Desta forma,

$$\begin{aligned}\mathbf{\Phi}(\mathbf{f}) &= \left(\mathbf{c}^T - \mathbf{f}^T\mathbf{Y}\right)\left(\mathbf{c} - \mathbf{Y}^T\mathbf{f}\right) \\ &= \mathbf{c}^T\mathbf{c} - \mathbf{c}^T\mathbf{Y}^T\mathbf{f} - \mathbf{f}^T\mathbf{Y}\mathbf{c} + \mathbf{f}^T\mathbf{Y}\mathbf{Y}^T\mathbf{f}\end{aligned} \tag{3}$$

e, portanto,

$$\frac{\partial\mathbf{\Phi}}{\partial\mathbf{f}} = -2\mathbf{c}^T\mathbf{Y}^T + 2\mathbf{f}^T\mathbf{Y}\mathbf{Y}^T, \tag{4}$$

usando o fato de que $\partial\mathbf{A}\mathbf{x}/\partial\mathbf{x} = \partial\mathbf{x}^T\mathbf{A}^T/\partial\mathbf{x} = \mathbf{A}$ e $\partial\mathbf{x}^T\mathbf{A}\mathbf{x}/\partial\mathbf{x} = \mathbf{x}^T\left(\mathbf{A}^T + \mathbf{A}\right)$. Logo, $\partial\mathbf{\Phi}/\partial\mathbf{f} = 0$ implica $\mathbf{Y}^T\mathbf{f} = \mathbf{c}$. Supondo que as colunas de $\mathbf{Y}$ sejam linearmente independentes quando $M \geq N$, a matriz $\mathbf{Y}^T\mathbf{Y}$ de dimensão $N \times N$ tem inversa. Assim sendo,

$$\mathbf{Y}^T\mathbf{f} = \mathbf{c} = \mathbf{Y}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{c}, \tag{5}$$

o que implica $\mathbf{f} = \mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{c}$.

Fazendo-se $\mathbf{Y} = \mathbf{T}^{-1/2}\mathbf{X}$ e substituindo na equação 5 obtém-se a equação 2, pois $\mathbf{T}^{-1/2}\mathbf{f} = \mathbf{h}$. Note que, sendo a matriz diagonal $\mathbf{T}$ inversível, $\mathbf{X}^T\mathbf{T}^{-1}\mathbf{X}$ também possui inversa quando $M \geq N$. Da forma como foi definida, a matriz $\mathbf{T}$ é responsável pelo "branqueamento" do espectro das imagens que compõem as colunas da matriz $\mathbf{X}$.

D. Análise de Características Dependente da Classe

O método Análise de Características Dependente da Classe (CFA) treina um banco de filtros de correlação baseado nos dados de um conjunto de treinamento genérico, onde se tem múltiplas imagens genuínas de cada classe. O conjunto de banco de filtros é então usado em experimentos de validação para extrair características discriminantes dependentes da classe para reconhecimento.

Entretanto, quanto se utiliza este algoritmo num dispositivo de reconhecimento de faces, como a LDV, o conjunto de treinamento estaria inicialmente vazio e, posteriormente, à

medida que novos indivíduos (classes) fossem cadastrados, o número de classes aumentaria progressivamente. Ocorre que, para haver um bom desempenho do algoritmo CFA, o número inicial de classes deve ser grande o suficiente para permitir um treinamento adequado [6], [9] dos filtros de correlação. A solução proposta neste trabalho para resolver este problema é a de dividir o conjunto de treinamento em duas partes: uma composta por indivíduos que deverão ser efetivamente reconhecidos, denominados *indivíduos primários* ou das *classes primárias* e outra composta por indivíduos que servirão somente para compor um número de classes inicial suficiente para o treinamento adequado dos filtros de correlação, denominados *indivíduos secundários* ou das *classes secundárias*. Se houver um pico de correlação significativo num indivíduo secundário, o sistema simplesmente retorna uma identificação negativa.

Um problema que surge com este tipo de solução é o de indivíduos das classes primárias poderem ser semelhantes a indivíduos das classes secundárias. Neste caso, poderia haver picos elevados nestas duas classes, dificultando o reconhecimento. Este problema é parcialmente resolvido com o algoritmo desenvolvido na subseção IV-A.

Outro problema, cuja solução será discutida na seção IV-B, a ser resolvido é o da interpretação do vetor de correlação, produzido pelo algoritmo de reconhecimento CFA, que é o produto escalar do vetor da DFT da imagem pela matriz do banco de filtros treinado, pois é preciso definir quando um pico de correlação é significativo ou não.

III. Pré-processamento das Imagens das Faces

Um dos primeiros problemas a ser resolvido para a implementação do sistema de reconhecimento de faces com uma *webcam* decorre do fato de que, de forma geral, os algoritmos usados para o reconhecimento precisam, como é o caso do CFA, normalizar a posição e a iluminação das faces, procurando reduzir ao máximo a influência, normalmente negativa, de suas variações durante reconhecimento. A seguir, dois métodos, um para a normalização da posição e outro para a normalização da iluminação, são apresentados, sendo que o método para a normalização da posição é uma contribuição original deste trabalho.

A. Normalização da Posição

Normalmente, a normalização da posição de uma face emprega as coordenadas dos olhos por serem pontos de fácil localização. Uma vez que se tenha suas coordenadas, uma translação seguida de uma transformação de escala e rotação é o suficiente para normalizar a posição. Para um sistema de reconhecimento de imagens estáticas, a localização não automática, ou seja, com intervenção humana, destes pontos não representa um problema sério. Já no caso da identificação dinâmica de imagens das faces geradas pela *webcam*, este procedimento não pode ser adotado na prática. Para contornar este problema, o detector de Viola-Jones é utilizado para obter as coordenadas automaticamente. Na verdade, o detector determina as coordenadas aproximadas dos olhos, mas visto que as coordenadas dos olhos das imagens do banco de dados de treinamento são obtidas pelo mesmo processo, o erro da aproximação é compensado.

B. Normalização da Iluminação

O método de normalização da iluminação descrito a seguir foi proposto em [14] e consiste basicamente de três etapas: correção gama, filtragem por diferença de gaussianas (*Difference of Gaussian Filtering* – DoG *Filtering*) e equalização do contraste.

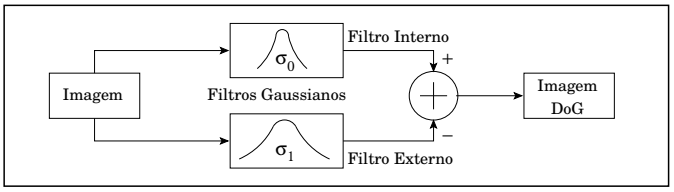

Fig. 2. Filtragem por diferença de gaussianas – DoG *Filter*.

1) Correção Gama: É uma transformação não linear dos níveis de cinza que substitui o valor $I(x,y)$ por $I^{\gamma}(x,y)$ ou por $\log\left(I(x,y)\right)$ quando $\gamma = 0$, onde $\gamma \in [0,1]$ é um parâmetro definido pelo usuário. Tem o efeito de ampliar a faixa dinâmica local da imagem nas áreas escuras ou sombreadas, reduzindo-a nas áreas claras e muito iluminadas. Um expoente γ na faixa $[0;0{,}5]$ é um bom compromisso. O valor $\gamma = 0{,}2$, sugerido em [14], é usado como valor padrão.

2) Filtragem por Diferença de Gaussianas: A correção gama não remove a influência de todos os gradientes de intensidade tais como efeitos de sombreamento. Sombreamento induzido pela estrutura da superfície é potencialmente um indicador visual útil mas é predominantemente informação de baixa freqüência espacial que é difícil de separar dos efeitos causados pelos gradientes de iluminação. Filtragem passa-altas remove ambas as informações útil e incidental, simplificando, portanto, o problema de reconhecimento e, em muitos casos, aumentando o desempenho total do sistema. De forma similar, suprimindo-se as freqüências espaciais mais altas reduzem-se *aliasing* e ruído. Na prática, isto é feito sem destruir demasiadamente a parte do sinal na qual o reconhecimento precisa se basear.

A filtragem DoG, ver figura 2, é uma forma conveniente de se obter o comportamento passa-banda resultante. Detalhes espaciais finos são criticamente importantes para o reconhecimento, portanto o filtro gaussiano interno é tipicamente bem estreito com $\sigma_0 \leq 1$ *pixel*, enquanto que o externo pode ter σ_1 de 2 a 4 *pixels* ou mais, dependendo da freqüência espacial em que a informação de baixa freqüência torna-se mais enganosa do que informativa. Para conjuntos de dados com grande variação de iluminação, [14] recomenda $\sigma_1 \approx 2$. Para variações menos extremas de iluminação, pode-se usar um valor de até 4 *pixels*.

3) Equalização de Contraste: Esta etapa reescala globalmente as intensidades da imagem para padronizar uma medida robusta de toda a variação do contraste. É importante utilizar um estimador robusto porque o sinal tipicamente ainda contém uma pequena mistura de valores extremos produzidos pelas áreas mais iluminadas da imagem, lixo nas bordas da imagem e regiões escuras tais como as narinas. O processo de equalização é feito em dois passos como mostrado nas equações 6 e 7

$$\begin{aligned} I_0(x,y) &= |I(x,y)|^a \\ I_1(x,y) &= \frac{I(x,y)}{\bar{I}_0^{1/a}} \end{aligned} \tag{6}$$

$$\begin{aligned} I_2(x,y) &= [\min\{\tau, |I_1(x,y)|\}]^a \\ I_3(x,y) &= \frac{I_1(x,y)}{\bar{I}_2^{1/a}}, \end{aligned} \tag{7}$$

onde a é um expoente fortemente compressivo que reduz a influência dos valores elevados e τ é um limiar usado para truncar os valores elevados após a primeira fase de normalização. $\bar{I}_0$ e $\bar{I}_2$ denotam os valores médios de $I_0(x,y)$

e $I_2(x,y)$, respectivamente. Como valores padrão usam-se $a = 0{,}1$ e $\tau = 10$.

A imagem resultante está agora bem escalada mais ainda contém valores extremos. Para reduzir a sua influência, aplica-se finalmente uma função não linear para comprimi-los. Para este fim, é empregada a tangente hiperbólica

$$I_4(x,y) = \tau \tanh\left[\frac{I_3(x,y)}{\tau}\right], \tag{8}$$

limitando $I_3(x,y)$ à faixa de valores no intervalo $(-\tau, \tau)$.

IV. Sistema de Reconhecimento de Faces

Nesta seção, os algoritmos e técnicas desenvolvidos para aprimorar o reconhecimento de faces com CFA são descritos. Apresenta-se, inicialmente, o algoritmo de seleção das imagens de treinamento, que atua eficazmente na redução do tempo total de treinamento. Em seguida, descreve-se a heurística criada para decidir se um pico de correlação é significativo ou não. Esta heurística permite definir o nível da confiabilidade da identificação e da rapidez com que a mesma é executada. Por fim, apresenta-se o método desenvolvido para estabilizar o vetor de picos de correlação, que atua na estabilização da identificação.

A. Seleção das Imagens de Treinamento

Para calcular os filtros utilizados no CFA, deve-se dispor de N_C indivíduos (classes) com N_{T_c} imagens de treinamento cada (amostras), perfazendo um total de $N_I = \sum_{c=0}^{N_C-1} N_{T_c}$ imagens, onde $c = 0, 1, \ldots, N_C - 1$. Em geral, para se obter filtros que permitam um reconhecimento satisfatório, as imagens devem ser o mais representativas possível para cada classe, ou seja, que as imagens correspondentes à cada classe estejam, se possível, em regiões do $\mathbb{R}^N$ sem interseção. Como, na prática, isto pode não ocorrer, como no caso de sósias e gêmeos, o que se pode fazer é tentar reduzir estas regiões de interseção. Outro problema que ocorre é o de imagens muito semelhantes dentro de uma mesma classe. Embora isto não represente um problema para o desempenho do reconhecimento, acarreta um aumento desnecessário do tempo de treinamento. Um método bem simples para reduzir estes problemas utiliza o ângulo entre duas imagens definido por

$$\cos(\theta) = \frac{\mathbf{U} \cdot \mathbf{V}}{|\mathbf{U}||\mathbf{V}|}. \tag{9}$$

Se as imagens tiverem média zero e varância unitária, tem-se que

$$\cos(\theta) = \frac{\mathbf{U} \cdot \mathbf{V}}{HW - 1}, \tag{10}$$

considerando-se uma estimativa não polarizada da variância, onde H é a altura e W é a largura das imagens $\mathbf{U}$ e $\mathbf{V}$. Seja θ_s o ângulo selecionado de separação das imagens. Então,

$$\begin{cases} \text{se } \theta \leq \theta_s & \Rightarrow \quad \text{imagens são semelhantes} \\ \text{se } \theta > \theta_s & \Rightarrow \quad \text{imagens não são semelhantes.} \end{cases} \tag{11}$$

Note que, para imagens com média zero e variância unitária, as classes ficam distribuídas na superfície de uma hiperesfera de raio $\sqrt{HW-1}$.

Empregando este critério, ao varrer a lista de imagens candidatas ao treinamento, uma dada imagem é incluída efetivamente no treinamento se *não* for semelhante a nenhuma outra que já tenha sido previamente selecionada para treinamento. A figura 3 ilustra em duas dimensões a atuação do método proposto. Note que se uma imagem primária for semelhante a uma secundária, ver seção II-D, esta será removida. Se uma classe secundária ficar vazia ela é simplesmente removida. Se uma classe primária ficar vazia, uma solução é reduzir o ângulo de separação θ_s e aplicar novamente o algoritmo.

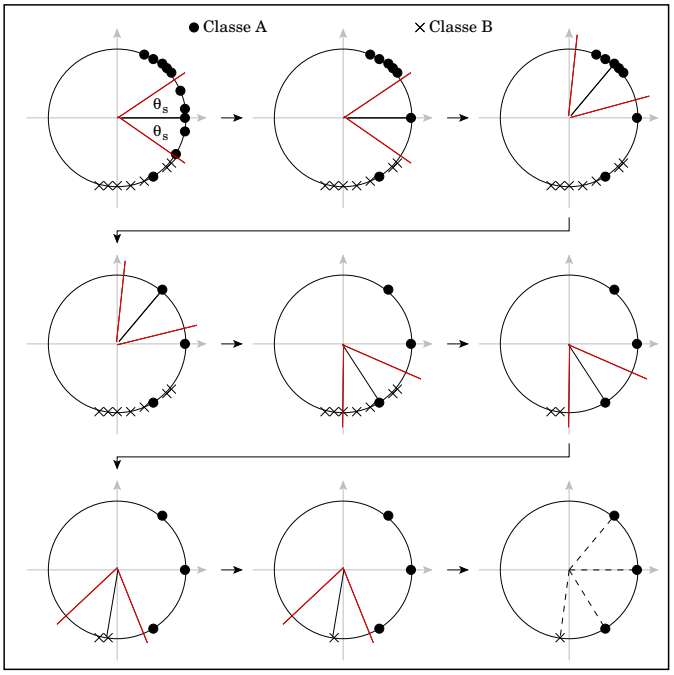

Fig. 3. Reduzindo a redundância da lista de imagens de treinamento.

B. Interpretação do Vetor de Picos de Correlação

É preciso definir o que será considerado um pico de correlação significativo. Como o vetor de correlação obtido pelo algoritmo CFA é real, a coordenada de maior valor, normalizado para 1, corresponde, em princípio, à face identificada. Como sempre haverá um máximo, mesmo que a face não esteja cadastrada, é preciso fazer com que o mesmo corresponda a uma identificação positiva estável e confiável. Para este fim, um contador, C, armazena quantas vezes seguidas o máximo principal, m_p (ID 7, figura 4), recai sobre um determinado indivíduo. Se o indivíduo muda ou se o máximo secundário m_s (ID 6, na figura 4) é maior que b, o contador é zerado. Considera-se que há uma mudança de indivíduo quando a detecção da face é interrompida, como a que ocorre se o indivíduo sai da frente da câmera. Então, se $m_s \leq a$, considera-se que o indivíduo foi identificado com um único quadro. Se $a < m_s \leq b$ e $C \geq N_F$ considera-se que o indivíduo foi identificado após N_F ou mais quadros consecutivos. Ou seja, quando $m_s \leq a$, um único quadro é suficiente para se considerar a identificação como sendo confiável. Quando $a < m_s \leq b$, pelo menos N_F quadros consecutivos são necessários para se ter uma identificação confiável. Os valores adotados para a, b e N_F serão discutidos na seção V.

C. Estabilização do Vetor de Picos de Correlação

Durante o desenvolvimento do sistema de reconhecimento de faces com CFA, notou-se que os máximos secundários, com valores acima do limiar b, do vetor de picos de correlação ocorriam aleatoriamente entre as classes, enquanto que o máximo principal permanecia estável. Isto fazia com que o contador C fosse zerado freqüentemente, dificultando a identificação do indivíduo. Para reduzir os efeitos deste problema foi desenvolvido o método descrito a seguir.

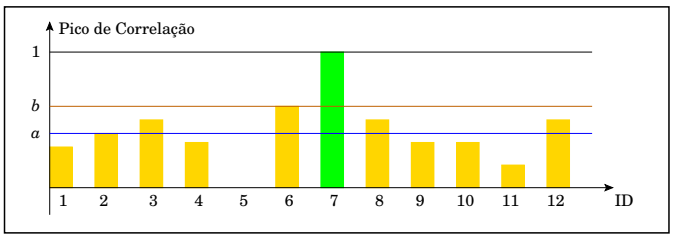

Fig. 4. Heurística para a identificação de faces com CFA.

Seja $\mathbf{c}[n] = [c_0[n], c_1[n], \ldots, c_j[n], \ldots, c_{N_C-1}[n]]$ o vetor de picos de correlação para o quadro n, onde N_C é o número de classes. Como a câmera fornece uma seqüência de vídeo, o vetor de picos de correlação obtido para os indivíduos cadastrados, ver figura 4, pode ser usado para obter o pico de correlação, $\bar{c}_j[n]$, ao longo do tempo para cada indivíduo do banco. À medida em que os vetores de correlação dos quadros passados, usados para obter $\bar{\mathbf{c}}$, são armazenados, são também multiplicados por um fator λ positivo menor que um. Isto faz com que, dependendo do valor de λ, os quadros passados influenciem mais (ou menos) o valor de $\bar{\mathbf{c}}[n]$ dado por

$$\bar{\mathbf{c}}[n] = \frac{\sum_{m=0}^{N_W-1} \lambda^m \mathbf{c}[n-m]}{\max_j \left\{ \sum_{m=0}^{N_W-1} \lambda^m \mathbf{c}[n-m] \right\}}, \qquad (12)$$

onde N_W é o número de quadros na janela usada para calcular $\bar{\mathbf{c}}[n]$ e $\mathbf{c}[n-m] = 0$ quando $n < m$. Os valores adotados para N_W e λ serão discutidos também na seção V.

V. Resultados

Nos experimentos efetuados foram utilizados um computador com processador Intel® Pentium® Core 2 Duo 3 GHz, com 2 MB de *cache* e 2 GB de RAM, uma *webcam* Logitech® QuickCam Chat e sistema operacional Linux.

A. Seleção das Imagens de Treinamento

Inicialmente, foi verificado o impacto no tempo de treinamento e na capacidade de reconhecimento produzido pelo algoritmo desenvolvido na seção IV-A, que visa selecionar somente um subconjunto representativo das imagens de treinamento usado para obter o banco de filtros de correlação. Na tabela I, são mostrados os tempos relativos de pré-processamento, T_p, e de treinamento, T_t, em função do ângulo de separação θ_s. O conjunto de treinamento contém 46 classes, sendo 42 classes secundárias (ver definição na seção II-D) com 60 imagens cada, 3 classes primárias com 600 imagens cada e 1 classe primária com 360 imagens, perfazendo um total de 4.680 imagens de 80×92 *pixels* já com a posição e iluminação normalizadas. Os indivíduos das classes primárias também foram usados para compor as classes secundárias, com o propósito de testar a funcionalidade do algoritmo da seção IV-A na eliminação dos efeitos de interferência que uma classe secundária pode ter num indivíduo da classe primária.

Quando $\theta_s = 0°$, o tempo de pré-processamento inclui somente o de normalização da posição e da iluminação das imagens do conjunto de treinamento já que, em princípio, devido ao ruído de captura introduzido pela câmera e pela própria variação da iluminação ambiente, não haveria duas imagens exatamente iguais neste conjunto. Desta forma, não se utilizou o algoritmo de seleção das imagens de treinamento quando $\theta_s = 0°$. O valor de $\theta_s > 0°$ mínimo, adotado na tabela I para aplicação efetiva do algoritmo de seleção, foi escolhido com base no cálculo do valor mínimo do ângulo entre as diversas classes. As classes secundárias correspondentes às primarias foram excluídas para efetuar este cálculo.

TABELA I
Tempo de treinamento T_t e de pré-processamento T_p em função do ângulo de separação θ_s utilizado na otimização da lista de imagens de treinamento.

$\theta_s(°)$	$\frac{T_p}{T_p^*}$	$\frac{T_t}{T_t^*}$	$\frac{T_p+T_t}{(T_p+T_t)^*}$	$\frac{T_t}{T_p}$	$N_I(\%)$
0	0,14	1,00	1,00	39,19	100
45	1,00	0,21	0,39	1,13	57
50	0,81	0,13	0,27	0,84	45
55	0,62	0,07	0,18	0,61	35
60	0,45	0,03	0,11	0,56	24
65	0,31	0,01	0,07	0,20	13

O símbolo "*" denota o valor máximo assumido pela variável.

Apesar de T_p para $\theta_s > 0°$ ser maior do que para $\theta_s = 0°$, a remoção das imagens redundantes reduz drasticamente o tempo de treinamento T_t, acarretando uma redução total ($T_p + T_t$) significativa. Quando há indivíduos que compõem tanto a classe primária quanto a secundária, observa-se uma melhora significativa na capacidade de reconhecimento dos mesmos até $\theta_s = 55°$. Quando os indivíduos são todos distintos, não se observa degradação na capacidade de reconhecimento. Em ambos os casos, o tempo total ($T_p + T_t$) para o cálculo do filtro $\mathbf{h}$ é reduzido significativamente em cerca de 80% para $\theta_s = 55°$, como se pode ver na tabela I. Para este conjunto de treinamento, $\theta_s = 55°$ foi a maior separação que não acarretou degradação observável do reconhecimento.

B. Estabilização e Interpretação do Vetor de Picos de Correlação

Os valores a, b e N_F, que determinam se um pico de correlação é significativo ou não, foram determinados experimentalmente. Antes, porém, foram determinadas faixas de valores de λ e N_W para a estabilização dos picos de correlação. Observou-se que para condições de iluminação favoráveis os valores $\lambda = 0{,}8$ e $N_W = 15$ eram suficientes. Já para ambientes com iluminação desfavorável, os valores $\lambda = 1$ e $N_W = 50$ eram mais adequados. Em geral, valores de λ na faixa $[0{,}8; 1]$ e de N_W na faixa entre 15 e 50 se mostraram efetivos na estabilização dos picos de correlação.

Com os picos de correlação estabilizados, passou-se para a determinação experimental de a, b e N_F. Inicialmente, desativou-se o teste do limiar b e vários indivíduos, cadastrados ou não, foram usados para testar o reconhecimento. O limiar a foi posto inicialmente em 0,5 e teve que ser reduzido para cerca de 0,18 para evitar que houvesse falsos positivos com a utilização de apenas um quadro da seqüência. Para este valor de a, os indivíduos cadastrados têm identificação positiva somente com condições de iluminação excelentes. O passo seguinte foi o de determinar valores para b e N_F. Como a câmera gerava vídeo a uma taxa de 15 quadros por segundo, incluindo detecção, pré-processamento e reconhecimento, decidiu-se adotar $N_F = 15$, ou seja, pelos menos um segundo seria necessário para fazer uma identificação positiva do indivíduo quando $a < m_s \leq b$. Em seguida, fez-se $b = 1$ inicialmente e, gradativamente, foi-se reduzindo seu o valor até que não fossem observados falsos positivos. O valor encontrado foi de cerca de 0,31.

C. Avaliação de Desempenho do Sistema

Em geral, testes de desempenho envolvendo algoritmos de reconhecimento de faces seguem os procedimentos descritos em [15], para obter a taxa de verificação positiva (TAR – *True Accept Rate*) e a taxa de falsos positivos (FAR – *False Accept Rate*) correspondente. Para que se pudesse fazer uma análise estatística avançada, foi estimado que um banco de dados com cerca de 50.000 imagens seria necessário. Este banco foi de fato criado tirando-se fotos de 200 indivíduos por semana durante um ano letivo, gerando o FRGC Ver2.0 (*Face Recognition Grand Challenge Ver2.0*) [15].

Para medir de forma confiável a TAR e a FAR do sistema de reconhecimento proposto neste trabalho, algo semelhante deveria ser feito, porém empregando seqüências de vídeo. Um banco com 50.000 seqüências de vídeo, com resolução de 320×240 *pixels* (25 vezes menos *pixels* que a do FRGC Ver2.0), com 10s de duração e 15 quadros por segundo, teria cerca de 2 Tbytes. Embora um espaço de 2 Tbytes não seja um grande problema hoje em dia, montar um banco desta magnitude requer um tempo considerável. Portanto, como um banco como este não está disponível (pelo menos não se tem conhecimento de que esteja), optou-se por adotar outra estratégia para avaliar, ainda que de forma imprecisa, a TAR e a FAR.

Após os ajustes dos parâmetros do sistema descritos nas seções V-A e V-B terem sido feitos, 20 indivíduos externos ao conjunto de treinamento e que não participaram dos ajustes dos parâmetros foram utilizados para "validar" a configuração e estimar a FAR. Todos os 20 indivíduos são pesquisadores em processamento de sinais e imagens. Eles foram informados previamente como o sistema operava para reconhecer faces. O gráfico de picos de correlação foi exibido em tempo real para cada um dos indivíduos que, com auxílio do primeiro autor, tentaram provocar a ocorrência de falsos positivos, alterando a pose, a expressão facial e o afastamento da câmera. Todos os indivíduos tiveram pelo menos cinco minutos para tentar realizar esta tarefa.

Somente um dos indivíduos conseguiu provocar a ocorrência de falsos positivos, mas somente quando se afastou da câmera, que não possui controle automático de foco. Neste caso, além das imagens analisadas pelo algoritmo ficarem fora de foco, a normalização da posição também produz uma suavização da imagem através da transformação de escalamento. Detalhes são perdidos e aumentam as chances do sistema cometer erros. Além disto, três dos indivíduos das classes primárias foram instruídos a proceder da mesma forma que os externos ao conjunto de treinamento para induzir erros de identificação, mas todos foram identificados corretamente.

D. Considerações Finais

O sistema de detecção e reconhecimento de faces foi concebido para faces frontais e com condições de iluminação controladas, mas os testes mostram que mesmo com a variação da posição das faces (10 a 20° de variação na horizontal e vertical) e de pequenas variações na intensidade da iluminação, o reconhecimento ainda é possível. O quanto a posição das faces pode variar depende tanto das condições de iluminação quanto da diversidade das imagens das faces que foram fornecidas para o treinamento.

Entretanto, mudanças um pouco mais acentuadas na iluminação, tais como alteração da posição da fonte de iluminação, podem dificultar o reconhecimento. É sempre possível retreinar os filtros de correlação para acomodar as novas variações, às custas de novos períodos de aquisição de imagens nas movas condições e de treinamento, o que pode ser considerado uma desvantagem do uso do algoritmo CFA.

Trocar a câmera que fez a aquisição das imagens de treinamento por outra de características diferentes é um exprimento que também precisa ser feito, para checar se os filtros de correlação calculados com o conjunto de treinamento obtido por uma câmera podem ser usados com outra diferente.

VI. Conclusões

Neste trabalho, foram desenvolvidos métodos que permitiram utilizar o algoritmo de reconhecimento de faces CFA, num sistema em *software* de reconhecimento de faces que obtém imagens de uma *webcam*, de forma mais eficiente. Os métodos desenvolvidos propiciaram: (a) a utilização do detector de faces de Viola-Jones para obtenção rápida e automática das coordenadas aproximadas dos olhos, permitindo a normalização automática da posição (seção III-A); (b) a redução da redundância do conjunto de treinamento, levando a uma redução significativa do tempo de treinamento de cerca de 80%, sem impactar o reconhecimento (seção IV-A); (c) a determinação da significância de um pico de correlação, aumentando a confiabilidade da identificação (seção IV-B); (d) a estabilização do vetor de picos de correlação, permitindo, com isto, que a identificação ficasse mais estável (seção IV-C). Os resultados experimentais também mostraram que o sistema permanece funcional para posições de face não frontais e com pequenas variações na intensidade da iluminação.

Referências

[1] P. Viola e M. J. Jones, *Robust Real-time Object Detection*, Technical Report Series CRL 2001/01, Cambridge Research Laboratory/Compaq Computer Corporation, Cambridge, Massachusetts 02142, USA, Fevereiro 2001.

[2] P. Viola e M. J. Jones, "Robust Real-time Object Detection". *Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling*, Vancouver, Canada, 13 de Julho 2001.

[3] P. Viola e M. J. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade". In: *Advances in Neural Information Processing System 14*, pp. 1311–1318, 2001.

[4] P. Viola e M. J. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 1, pp. I.511–I.518, Kauai, HI, USA, 8-14 de Dezembro 2001.

[5] P. Viola e M. J. Jones, "Robust Real-time Face Detection", *International Journal of Computer Vision*, v. 57, n. 2, pp. 137–154, 2004.

[6] C. Xie, M. Savvides e B. V. K. V. Kumar, "Redundant Class-Dependence Feature Analysis Based on Correlation Filters Using FRGC 2.0 Data". In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 3, pp. 153–158 (6), San Diego, California, USA, 20-25 de Junho 2005.

[7] L. Wiskott, J.-M. Fellous, N. Krüger e *et al.*, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, pp. 775–779, Julho 1997.

[8] L. Wiskott, J.-M. Fellous, N. Krüger e *et al.*, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*. Isbn: 0-8493-2055-0 ed. L.C. Jain et al. (Editores) CRC Press, 1999.

[9] R. Abiantum, M. Savvides e B. V. K. V. Kumar, "How Low Can You Go? Low Resolution Face Recognition Study Using Kernel Correlation Feature Analysis on the FRGCv2 dataset". In: *IEEE Biometrics Symposium*, Baltimore, Maryland, USA, 19-21 de Setembro 2006.

[10] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features". *Proceedings of the Seventh International Conference on Computer Vision*, Kerkyra, Grécia, 20–27 de Setembro 1999.

[11] D. G. Lowe, "Local Feature View Clustering for 3D Object Recognition". *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, Dezembro 2001.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, v. 60, pp. 91–110, Novembro 2004.

[13] M. Savvides, B. V. K. V. Kumar e P. Khosla, "Face Verification Using Correlation Filters". In: *Proceedings of Third IEEE Automatic Identification Advanced Technologies*, pp. 56–61, Terrytown, New York, USA, 14-15 de Março 2002.

[14] X. Tan e B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions". In: *Proceedings of the 2007 Analysis and Modeling of Faces and Gestures*, pp. 168–182, Rio de Janeiro, Brasil, 20 de Outubro 2007.

[15] P. J. Philips, P. J. Flynn, T. Scruggs e *et al.*, "Overview of The Face Recognition Grand Challenge". *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, USA, 20-25 de Junho 2005.