

Filtros de Correlação e Características Invariantes à Escala para o Reconhecimento de Faces

Rodrigo L. Prates, Marcelo B. Larcher, José F. L. de Oliveira e Eduardo A. B. da Silva

Resumo—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, a partir de imagens obtidas de dispositivos de captura como uma *webcam*. Com este intuito, dois algoritmos estado-da-arte foram selecionados: o CFA – *Class-dependence Feature Analysis* e o SURF – *Speed Up Robust Features*, sendo este conceitualmente similar ao SIFT – *Scale Invariant Feature Transform*. Para o CFA, é proposto o uso da DCT – *Discrete Cosine Transform* e KLT – *Karhunen-Loève Transform* como alternativas à DFT – *Discrete Fourier Transform* para reduzir tempo do cálculo dos filtros de correlação e/ou obter uma ROC melhor. Para o SURF, dois parâmetros de configuração são testados, tendo também como objetivo determinar a melhor ROC.

Palavras-Chave—Reconhecimento de Faces, Filtros de Correlação, Características Invariantes à Escala

Abstract—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 capture devices such as a webcam. In order to implement such a system, two state-of-the-art algorithms were selected: CFA – *Class-dependence Feature Analysis* and SURF – *Speed Up Robust Features*, the last one being conceptually similar to SIFT – *Scale Invariant Feature Transform*. For CFA, it is proposed to use of DCT – *Discrete Cosine Transform* and KLT – *Karhunen-Loève Transform* as alternatives to the DFT, so as to reduce the time to compute the correlation filters and/or improve the ROC. For SURF, two configuration parameters are tested, having also as objective the determination of the best ROC.

Keywords—Face Recognition, Correlation Filters, Scale Invariant Features

I. INTRODUÇÃO

O reconhecimento de faces e/ou objetos é uma área de pesquisa atualmente muito ativa, com uma demanda crescente por algoritmos cada vez mais robustos. Tais algoritmos apresentam diversas aplicações, como por exemplo em controle de acesso e verificação de conformidade em linhas de produção, e também no auxílio de pessoas com deficiência visual.

Dada a crescente importância e utilidade destes métodos de detecção e reconhecimento, este trabalho tem por objetivo 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*. Um sistema de reconhecimento facial deve ser também capaz de detectar e reconhecer imagens em diversas condições, garantido assim a robustez necessária para o seu funcionamento.

Para implementar o sistema descrito no parágrafo anterior, optou-se por utilizar, inicialmente, o algoritmo de Viola-Jones [1], [2], [3], [4], [5] 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 algoritmos baseados no CFA – *Class-dependence Feature Analysis* [6] e no SURF – *Speed Up Robust Features* [7] cujos conceitos são bastante similares aos do SIFT – *Scale Invariant Feature Transform* [8], [9], [10]. Evidentemente, outras técnicas reconhecidamente estado-da-arte, como o EBGM (*Elastic Bunch Graph*

Rodrigo Leite Prates, Marcelo Barbosa Larcher, José Fernando Leite de Oliveira e 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: rodrigolp@lps.ufrj.br; marcelo.larcher@lps.ufrj.br; jleite@lps.ufrj.br; eduardo@lps.ufrj.br.

Matching) [11], [12] ou até mesmo o KCFA (*Kernel Classdependence Feature Analysis*) [13], foram consideradas. Entretanto, a menor complexidade computacional do CFA e a comprovada capacidade de reconhecimento de objetos, aliada à generalidade, eficiência e relativa baixa complexidade do SURF favoreceram as suas escolhas. Como SURF e SIFT trabalham com conceitos semelhantes, este também será testado em trabalhos futuros.

Em trabalho anterior, já havia sido proposto para o CFA um algoritmo de seleção das imagens das faces que compõem o conjunto de treinamento e que reduz o tempo de treinamento em cerca de 80%, eliminando a ocorrência de imagens redundantes [14]. Além disso, algumas técnicas para aumentar a confiabilidade e a estabilidade da identificação de faces do CFA também foram propostas, as quais empregam quadros passados da sequência de vídeo capturada pela *webcam* [14].

Neste trabalho, com o intuito de aumentar a taxa de verdadeiros positivos e de reduzir a taxa de falsos positivos e/ou de reduzir o tempo de treinamento, decidiu-se testar outras formas de implementação do CFA, alterando-se a forma tradicional de extração de características no domínio da frequência através da DFT – *Discrete Fourier Transform* [15]. A DFT foi substituída pela DCT – *Discrete Cosine Transform* e pela KLT – *Karhunen-Loève Transform* [15].

A seguir, na seção II, são apresentados conceitos básicos relacionados ao CFA e ao SURF, sendo o CFA apresentado na subseção II-A e o SURF, na II-B. Na seção III, o pré-processamento das imagens das faces, que tem por objetivo reduzir o efeito das variações de iluminação e pose, é abordado rapidamente. A seção IV é dedicada à apresentação dos sistemas de reconhecimento de faces com CFA e SURF. A subseção IV-A é dedicada à apresentação dos métodos desenvolvidos para reduzir o tempo de treinamento do CFA e aumentar a estabilidade e a confiabilidade da identificação. A subseção IV-B apresenta os procedimentos utilizados para incorporar o SURF a um sistema de reconhecimento de faces. Já na seção V, serão apresentadas as modificações introduzidas no CFA pela utilização da DCT e da KLT no cálculo dos filtros de correlação. A seção VI é dedicada aos resultados preliminares obtidos com o CFA e o SURF. Na subseção VI-A, além dos resultados relativos ao CFA, é também descrita a metodologia seguida para a avaliação de ambas as modificações introduzidas pela KLT e DCT. Por fim, são apresentadas as conclusões na seção VII.

II. CONCEITOS BÁSICOS

A. O Algoritmo CFA

A tecnologia dos filtros de correlação é uma ferramenta básica para o casamento de imagens no domínio da frequência. Em métodos de filtros de correlação como o CFA, variações normais em imagens de treinamento autênticas podem ser acomodadas pelo projeto de um arranjo no domínio da frequência, chamado de filtro de correlação, que captura a parte consistente das imagens de treinamento, desenfocando as partes inconsistentes, isto é, as frequências inconsistentes. O reconhecimento de objetos é feito pela correlação cruzada de uma imagem de entrada com um filtro projetado, usando a DFT. 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

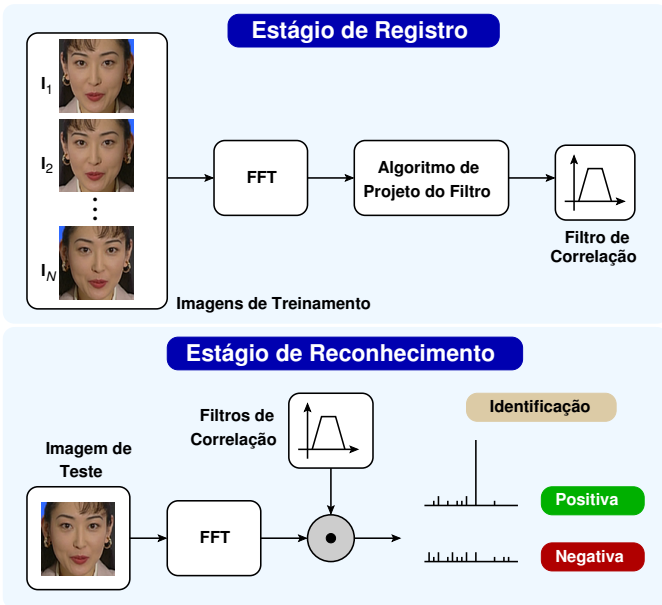


Fig. 1

O CONCEITO DE RECONHECIMENTO DE FACES ESTÁTICAS UTILIZANDO FILTROS DE CORRELAÇÃO.

imagens. Outra vantagem de se utilizar filtros de correlação é que eles oferecem soluções de forma fechada as quais são computacionalmente efetivas.

1) *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 frequência, 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].

2) *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

$$c_j = \mathbf{x}_j^T \cdot \mathbf{h}, \quad (1)$$

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 frequência 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_{jj}^j representam a potência espectral da j -ésima imagem de treinamento na frequência f_i .

O filtro MACE enfatiza frequências espaciais altas para produzir grandes picos de correlação, enquanto o filtro MVSDF, em geral, suprime as altas frequências 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

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

onde $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}]$ 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]. A dedução da equação 2 pode ser encontrada em [14].

B. O Algoritmo SURF

Algoritmos como SIFT e SURF, que extraem diferentes características invariantes de imagens para obter uma correspondência confiável entre imagens diferentes de um mesmo objeto ou cena, podem ser adaptados e utilizados no reconhecimento facial. Estas características extraídas das imagens das faces são invariantes a escalamentos e rotações e parcialmente invariantes a mudanças na iluminação. O objetivo desta parte do trabalho é realizar o reconhecimento de faces humanas através do algoritmo SURF, com ajustes de parâmetros a fim de otimizar o desempenho do sistema. Também poderão ser desenvolvidas ferramentas de pré-processamento de imagens para auxiliar neste reconhecimento, como as descritas na seção III.

1) *Casamento de Pontos-Chave*: Na definição dos pontos-chave do algoritmo SURF, apenas as características mais “fortes”, cujo valor do hessiano é maior que um dado limiar L , são extraídas. Portanto, quanto menor for este limiar mais pontos-chave são identificados em uma dada imagem.

O casamento dos pontos-chave é realizado na etapa de correspondência, onde cada um dos pontos-chave de uma imagem é comparado com todos os pontos-chave de outra imagem através do cálculo da distância euclidiana entre os vetores de seus descritores. O par de correspondência com a menor distância euclidiana, definido como vizinho mais próximo (*nearest neighbor*), é detectado se sua distância for menor que a distância do segundo vizinho mais próximo multiplicada por uma constante k , ou seja:

$$d_{N_1} < k d_{N_2}, \quad (3)$$

onde d_{N_1} e d_{N_2} são, respectivamente, as distâncias ao vizinho mais próximo e ao segundo vizinho mais próximo.

2) *O Ponto de Corte*: Em cada casamento, é gerado um certo número de correspondências entre os pontos-chave. Estes casamentos produzem dois grupos de resultados: os de correspondência entre faces de pessoas diferentes e da mesma pessoa. Em geral, o número de correspondências entre pessoas diferentes é menor que o da mesma pessoa, por isto, estabelece-se um ponto de corte entre estes grupos a fim de se obter os melhores resultados. A seleção do ponto de corte é realizada com base na eficiência, que é a média aritmética da sensibilidade e especificidade.

A sensibilidade (*SENS*), ou taxa de verdadeiros positivos (V_P), é a proporção do número de indivíduos reconhecidos corretamente (V_P) entre todos os resultados de mesma pessoa ($V_P + F_N$), ou seja:

$$SENS = \frac{V_P}{V_P + F_N}, \quad (4)$$

onde F_N é o número de falsos negativos. A especificidade (*ESPEC*), ou taxa de verdadeiros negativos (V_N), é a proporção do número

de indivíduos que, corretamente, não foram identificados (V_N) nas correspondências entre pessoas diferentes ($V_N + F_P$). Sua equação é definida por:

$$ESPEC = \frac{V_N}{V_N + F_P}, \quad (5)$$

onde F_P é o número de falsos negativos.

III. PRÉ-PROCESSAMENTO DAS IMAGENS DAS FACES

Um dos primeiros problemas a ser resolvido para a implementação do sistema de reconhecimento de faces 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, serão apresentados os dois métodos usados, um para a normalização da posição e outro para a normalização da iluminação, conforme [14].

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. Enquanto que para um sistema de reconhecimento de imagens estáticas, a localização não automática desses pontos não é um problema sério, para um sistema dinâmico de faces geradas por *webcam*, passa ser um procedimento que não pode ser adotado na prática. Para contornar este problema, o detector de Viola-Jones é utilizado para obter as coordenadas automaticamente.

B. Normalização da Iluminação

O método de normalização da iluminação descrito a seguir foi proposto em [16] e consiste basicamente de três etapas: correção gama, filtragem por diferença de gaussianas (*Difference of Gaussian Filtering - DoG Filtering*) e equalização de contraste. A correção de gama é uma transformação não linear dos níveis de cinza da imagem. Já a filtragem por diferença de gaussianas é obtida através do uso de dois filtros gaussianos de desvios padrões diferentes, sendo que a imagem final é a diferença entre duas imagens, cada uma resultante da imagem original filtrada por cada um dos filtros. Isso é feito para remover a influência dos gradientes de intensidade, tais como efeitos de sombreamento, não removidos pela correção de gama. Por fim, a equalização de contraste reescala globalmente as intensidades da imagem para padronizar uma medida robusta de toda a variação do contraste. Mais detalhes sobre este tipo de normalização podem ser encontrados em [16], [14].

IV. SISTEMAS DE RECONHECIMENTO DE FACES

A. Sistema Utilizando o CFA

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.

1) *Seleção das Imagens de Treinamento*: Para calcular os filtros utilizados no CFA, deve-se dispor de uma certa quantidade de indivíduos (classes), com uma certa quantidade de imagens de treinamento em cada classe. 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 a cada classe estejam, se possível, em regiões do \mathbb{R}^N sem interseção. Como, na prática, isto pode não ocorrer, como no caso de sócias

e gêmeos, o que se pode fazer é tentar reduzir estas regiões de interseção. Um método bem simples para reduzir este problema, conforme [14], utiliza o ângulo entre duas imagens definido por

$$\cos(\theta) = \frac{\mathbf{U} \cdot \mathbf{V}}{|\mathbf{U}||\mathbf{V}|}. \quad (6)$$

Seja θ_s o ângulo selecionado de separação das imagens. Então,

$$\begin{cases} \text{se } \theta \leq \theta_s & \Rightarrow \text{imagens são semelhantes} \\ \text{se } \theta > \theta_s & \Rightarrow \text{imagens não são semelhantes.} \end{cases} \quad (7)$$

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.

2) *Interpretação do Vetor de Picos de Correlação*: 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. De forma a tratar os casos onde mais de um pico de correção tem valor unitário ou próximo desse valor, um limiar de aceitação foi criado de maneira que ao existir mais de um pico de correção acima desse limiar, considera-se que o indivíduo de pico de correlação unitário não foi reconhecido. Esse limiar pode variar, numa faixa normalizada, de 0 a 1.

B. Sistema Utilizando o SURF

Da mesma forma que o CFA, em um sistema de reconhecimento de faces utilizando o SURF, as informações de interesse concentram-se obviamente na face humana, sendo necessária a eliminação da imagem de fundo para que os pontos-chave do registro de um indivíduo sejam somente aqueles correspondentes à sua face. Com este propósito, o algoritmo de Viola-Jones é novamente usado para a detecção automática de faces humanas. Então, com a devida normalização geométrica, é possível obter os pontos-chaves somente da face. Já no reconhecimento o detector pode ser completamente dispensado visto que, provavelmente, só haverá uma quantidade significativa de casamentos de pontos-chave com a imagem fornecida para reconhecimento se o indivíduo, registrado previamente, estiver presente na cena. É claro que, se se usar o detector de Viola-Jones para limitar a área de reconhecimento, o número de pontos-chave calculados tenderá a ser bem menor.

O registro de cada indivíduo é feito armazenado-se uma ou mais de suas imagens em formato monocromático de 8 bpp com tamanho fixo de 256×256 pixels e, em seguida, calculam-se os pontos-chave correspondentes que são também armazenados. Após os pontos-chave de registro terem sido calculados as imagens de registro podem ser descartadas.

V. MODIFICAÇÕES IMPLEMENTADAS NO CFA

Foram propostas modificações que visam a melhorar o desempenho do reconhecimento e o custo computacional do sistema como um todo. A utilização da DCT na representação das imagens foi implementada com o objetivo de tornar o treinamento mais rápido e melhorar o índice de acerto do reconhecimento. A KLT foi introduzida na tentativa de obter o melhor desempenho possível, uma vez que é a transformada ótima. Tais motivos serão detalhados mais adiante nesta seção, assim como os métodos empregados e a metodologia usada para a avaliação.

A. DCT no Algoritmo de Reconhecimento

As características da DCT que podem ser úteis na tentativa de aperfeiçoar o algoritmo de reconhecimento são:

- maior capacidade de compressão em relação à DFT;
- possuir apenas coeficientes reais, enquanto os coeficientes da DFT têm componente imaginária.

Esperava-se obter com a substituição da DFT pela DCT uma redução significativa no tempo de treinamento devido ao fato de a DCT produzir como saída uma sequência real, o que permite realizar os cálculos do CFA apenas com operações utilizando números reais. Além disso, como a DCT concentra a energia do sinal em um número menor de coeficientes, reduz-se a quantidade de informação desnecessária ao reconhecimento e, com isso, espera-se uma melhora no desempenho do sistema.

B. KLT no Algoritmo de Reconhecimento

No cálculo da matriz da KLT, foi utilizado um extenso conjunto de imagens para obter uma matriz de covariância que pudesse simular grande parte das variações estatísticas das faces. Para o cálculo da matriz de covariância foi utilizada a base FERET [17]. As imagens foram pré-processadas, passando por uma normalização da posição, normalização da iluminação, correção gama, filtragem por diferença de gaussianas e equalização de contraste. Por conveniência, optou-se por utilizar um subconjunto da FERET composto apenas por imagens de indivíduos sem óculos e sem barba. Deste subconjunto foram extraídas as coordenadas de olhos utilizadas na normalização geométrica.

C. Metodologia para a Avaliação das Modificações

A fim de avaliar as várias formas de implementação do algoritmo de CFA, originalmente por DFT, e também por DCT e KLT, uma metodologia de testes foi criada. A ideia é se obter, ao final, um entendimento melhor sobre quais as vantagens e desvantagens de cada implementação, e em quais situações cada implementação apresenta maior taxa de verdadeiros positivos e menor taxa de falsos positivos. As etapas do processo de avaliação seguem:

- 1) separar um conjunto de imagens, advinda de duas bases de dados, para compor os conjuntos de treinamento e teste;
- 2) utilizar a validação cruzada sobre os dados acima;
- 3) variar os parâmetros envolvidos na implementação do CFA e do sistema proposto por [14];
- 4) gerar os filtros do algoritmo CFA;
- 5) apresentar os resultados, para os três métodos, na forma de curvas ROC, ou seja, taxa de verdadeiros positivos por taxa de falsos positivos;
- 6) concluir qual o método com os maiores valores de taxa de verdadeiros positivos para a menor taxa de falsos positivos, e em qual configuração de parâmetros.

Foram escolhidas as bases BioID [18] e FRD_ITJRSC-1.7. A primeira é uma base disponibilizada pela empresa de segurança digital BioID, composta por 1521 imagens frontais, em tons de cinza, com resolução de 384×286 pixels, de 23 pessoas diferentes. A quantidade de imagens por pessoa variando, indo de 6 imagens por pessoa a 15 imagens por pessoa. A base FRD-ITJRSC-1.7 é uma base montada na sua maior parte em Manaus, no Instituto de Tecnologia José Rocha Sérgio Cardoso (ITJRSC), e uma pequena parte no Rio de Janeiro. Ela é composta por 4920 imagens coloridas de 45 indivíduos, 60 imagens por indivíduo, com resolução de 320×240 pixels, montada sob diferentes níveis de iluminação e posicionamento do rosto. Essa base de dados não foi publicada. Para a base BioID foram selecionados 5 indivíduos, 98 imagens cada, para formar os conjuntos de treinamento e teste, na forma de validação cruzada com 7 *Folds*. Além disso, foi adicionado para cada *Fold*, no teste, um total de 525 imagens de 15 indivíduos, que não fizeram parte do treinamento, para se garantir a obtenção de um número mínimo de falsos positivos.

Já para a base FRD_ITJRSC-1.7 foram selecionados 5 indivíduos, 300 imagens cada, para formar os conjuntos de treinamento e teste, na forma de validação cruzada com 10 *Folds*. Assim como foi feito na base anterior, um certo número de imagens foram adicionadas ao conjunto de teste, num total de 480 imagens de 8 indivíduos. Com os *Folds* montados, estes dados devem ser usados para se gerar os filtros

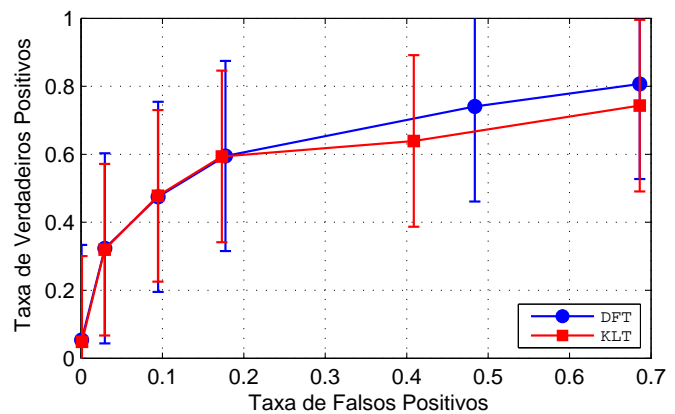


Fig. 2
CURVAS ROC PARA O MÉTODO CFA COM AS TRANSFORMADAS DFT E KLT UTILIZANDO A BASE BIOID.

do algoritmo CFA, para cada uma das 3 implementações possíveis. Ao gerar os filtros, algumas alterações podem ser feitas de maneira a permitir um melhor desempenho das implementações. Os seguintes parâmetros foram considerados relevantes:

- limiar de aceitação (ver IV-A.2);
- α e β do projeto do filtro do CFA (ver II-A.2);
- nível de ruído incidente sobre as imagens de treinamento.

Com a configuração de parâmetros acima definida, o treinamento é realizado para se gerar os filtros do algoritmo CFA para cada método.

VI. RESULTADOS

A. Resultados do Algoritmo CFA

Com base no que foi exposto na subseção V-C, são apresentados alguns resultados preliminares, visto que o projeto está em desenvolvimento. A figura 2 é a curva ROC para os métodos por KLT e DFT, utilizando a base BioID e variando-se o limiar de aceitação. Para a KLT, foram empregados os parâmetros $\alpha = 1$ e $\beta = 0$ e para a DFT os testes foram feitos com $\alpha = 0$ e $\beta = 1$. É verdade que, para uma comparação justa e precisa entre KLT e DFT os parâmetros α e β deveriam ser aqueles obtidos para as melhores curvas ROC. A figura 2 apresenta somente dois casos particulares de curvas ROC para KLT e DFT. Assim, ainda não há resultados estatísticos confiáveis para tirar conclusões definitivas a respeito das vantagens e desvantagens das diferentes transformadas.

Foram definidos 6 valores dentro de uma faixa de 0,125 a 0,875 para a variação desse limiar. Esses são os 6 pontos de cada uma das curvas ROC, onde cada ponto foi obtido pela média para os 7 *Folds*. Como se pode observar, para valores de limiar próximos de 0, a taxa de verdadeiros positivos é baixa assim como a de falsos positivos. Isso é esperado, já que com o limiar baixo, temos quase todos os picos de correlação acima desse limiar, e assim o pico de correlação unitário não é reconhecido. Por outro lado, aumentando-se o limiar de aceitação, as taxas de verdadeiros positivos são naturalmente maiores. Por fim, as barras de erro de cada ponto foram criadas considerando-se uma distribuição *t-student*, já que o número de pontos $N < 30$ [19].

Observando a figura VI-A, não se consegue perceber nenhuma vantagem da KLT em relação à DCT no desempenho do reconhecimento de faces da CFA. Experimentos com precisão estatística maior precisam ser conduzidos para permitir a corroboração deste resultado preliminar. Além disto, ainda não se sabe se as curvas ROC são as melhores que se podem obter, pois os valores ótimos de α e β ainda precisam ser investigados em cada caso. Com relação ao uso da DCT, não foi observado, nestes primeiros testes, um potencial

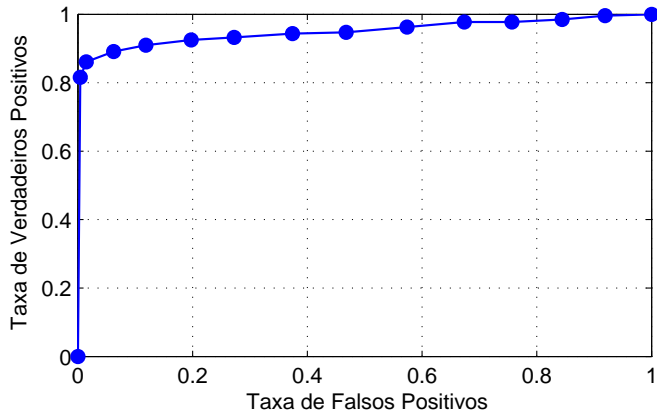


Fig. 3

ROC DA MELHOR CONFIGURAÇÃO OBTIDA PARA O MÉTODO DE RECONHECIMENTO DE FACES COM SURF.

para melhorar o desempenho da CFA em termos de ROC, valendo as mesmas observações feitas acima acerca da KLT. Entretanto, com a DCT, o tempo de treinamento dos filtros de correlação é reduzido sensivelmente.

B. Resultados do Algoritmo SURF

Para este experimento são utilizadas 575 imagens coloridas, das quais 540 imagens são do banco de dados FRGC 2.0 e outras 35 imagens adquiridas através de uma câmera digital **Canon PowerShot SX220 HS**. Todo o conjunto é composto por 5 imagens frontais de 115 pessoas distintas, formando um banco de imagens com diferentes condições de iluminação, variações de expressões faciais e posicionamento da face. As imagens selecionadas contêm o rosto de cada indivíduo e o cenário no qual ele se encontra inserido.

Os dois parâmetros L e k do algoritmo SURF, mencionados na subseção II-B.1, são variados com o propósito de determinar o quanto eles influenciam o seu desempenho. Neste experimento, são utilizados limiares de valores $L = \{0, 100, 300, 500\}$ e para a constante k são adotados os valores $k = \{0, 35; 0, 40; 0, 45; 0, 50; 0, 55; 0, 60\}$. Com isto, são obtidas 24 curvas ROC, onde o parâmetro variado é o limiar de corte mencionado na seção II-B.2. Visando a selecionar a melhor curva ROC, para cada valor de k foi feito um gráfico contendo as curvas para todos os limiares L testados. De cada um destes gráficos, foi selecionada a melhor curva ROC. Verificou-se, então, que $L = 0$ correspondia a melhor curva ROC para todos os valores de k testados. Finalmente as curvas ROC para todos os valores de k e $L = 0$ foram postas em um mesmo gráfico. Feito isto, verificou-se que não havia uma diferença significativa entre elas e a melhor curva ROC encontrada para o SURF neste experimento é a mostrada na figura 3.

Portanto, este experimento demonstra uma capacidade promissora do algoritmo SURF no reconhecimento de faces com taxa de verdadeiros positivos de cerca de 86% para uma taxa de falsos positivos de cerca de 1,5%. Para 10% de falsos positivos tem-se 90% de verdadeiros positivos.

VII. CONCLUSÕES

Neste artigo, foram propostos e testados dois sistemas de reconhecimento faces: um utilizando CFA – *Class-dependence Feature Analysis* e outro utilizando SURF – *Speed Up Robust Features*, conceitualmente semelhante ao SIFT – *Scale Invariant Feature Transform*. Para o primeiro sistema, além das propostas de [14] foram introduzidas as transformadas DCT e KLT em substituição à DFT no cálculo dos filtros de correlação da CFA. Os primeiros testes

não mostram nenhuma melhora obtida pela KLT, como pode ser observado por meio da curva ROC. Entretanto, experimentos com precisão estatística maior deverão ser realizados para confirmar este resultado preliminar. Com relação ao uso da DCT, não foi observado, nestes primeiros testes, uma melhora de desempenho em termos de ROC. Entretanto, o tempo de treinamento dos filtros de correlação é reduzido sensivelmente. Para o segundo sistema, utilizando o SURF, os primeiros testes demonstram uma capacidade promissora no reconhecimento de faces com taxa de verdadeiros positivos de cerca de 86% para uma taxa de falsos positivos de cerca de 1,5%. Para 10% de falsos positivos, têm-se cerca de 90% de verdadeiros positivos.

REFERÊNCIAS

- [1] VIOLA, P., JONES, M. J., *Robust Real-time Object Detection*, Technical Report Series CRL 2001/01, Cambridge Research Laboratory/Compaq Computer Corporation, Cambridge, Massachusetts 02142, USA, Fevereiro 2001. 1
- [2] VIOLA, P., JONES, M. J., “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. 1
- [3] VIOLA, P., JONES, M. J., “Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade”. In: *Advances in Neural Information Processing System 14*, pp. 1311–1318, 2001. 1
- [4] VIOLA, P., JONES, M. J., “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. 1511–1518, Kauai, HI, USA, 8-14 de Dezembro 2001. 1
- [5] VIOLA, P., JONES, M. J., “Robust Real-time Face Detection”, *International Journal of Computer Vision*, v. 57, n. 2, pp. 137–154, 2004. 1
- [6] XIE, C., SAVVIDES, M., KUMAR, B. V. K. V., “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. 1, 2
- [7] BAY, H., TUYTELAARS, T., GOOL, L. V., “SURF: Speeded Up Robust Features”. *European Conference on Computer Vision - ECCV 2006*, Europe, 2006. 1
- [8] LOWE, D. G., “Object Recognition from Local Scale-Invariant Features”. *Proceedings of the Seventh International Conference on Computer Vision*, Kerkyra, Grécia, 20–27 de Setembro 1999. 1
- [9] LOWE, D. G., “Local Feature View Clustering for 3D Object Recognition”. *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, Dezembro 2001. 1
- [10] LOWE, D. G., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, v. 60, pp. 91–110, Novembro 2004. 1
- [11] WISKOTT, L., FELLOUS, J.-M., KRÜGER, N., *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. 1
- [12] WISKOTT, L., FELLOUS, J.-M., KRÜGER, N., *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. 1
- [13] ABIANTUM, R., SAVVIDES, M., KUMAR, B. V. K. V., “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. 1
- [14] DE OLIVEIRA, J. F. L., DA SILVA, E. A. B., CARDOSO, M. A. P., *et al.*, “Um Sistema de Autenticação por Faces Usando Filtros de Correlação em Vídeo”. *XXVII Simpósio Brasileiro de Telecomunicações - SBRT 2009*, Blumenau, Santa Catarina, Brasil, 29 de Setembro a 2 de Outubro 2009. 1, 2, 3, 4, 5
- [15] JAIN, A. K., *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey 07632, Prentice Hall, 1989. 1
- [16] TAN, X., TRIGGS, B., “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. 3
- [17] FERET, “The Color FERET Database”, 2003. Disponível em <http://face.nist.gov/colorferet>. 4
- [18] BIOID, “BioID Face Database”, 2011. Disponível em <http://www.bioid.com/support/download/software/bioid-face-database.html>. 4
- [19] PAPAULIS, A., *Probability, Random Variables, and Stochastic Processes*. 3 ed. Singapore, McGraw-Hill, 1991. 4