UNIVERSIDADE FEDERAL DA PARAÍBA

CENTRO DE INFORMÁTICA

DEPARTAMENTO DE INFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

ADRIANO DA SILVA MARINHO

UMA NOVA VERSÃO DE UM SISTEMA DE DETECÇÃO E RECONHECIMENTO DE FACE UTILIZANDO A TRANSFORMADA COSSENO DISCRETA

JOÃO PESSOA – PB

JULHO DE 2012

ADRIANO DA SILVA MARINHO

UMA NOVA VERSÃO DE UM SISTEMA DE DETECÇÃO E RECONHECIMENTO DE FACE UTILIZANDO A TRANSFORMADA COSSENO DISCRETA

Dissertação submetida ao Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba como parte dos requisitos necessários para obtenção do título de Mestre em Informática.

Orientador: Prof. Dr. Ed Porto Bezerra.

Co-orientador: Prof. Dr. Leonardo Vidal Batista.

JOÃO PESSOA – PB

JULHO DE 2012

M338u Marinho, Adriano da Silva.

Uma nova versão de um sistema de detecção e reconhecimento de face utilizando a Transformada Cosseno Discreta / Adriano da Silva Marinho.-- João Pessoa, 2012.

80f.

Orientador: Ed Porto Bezerra

Co-orientador: Leonardo Vidal Batista

Dissertação (Mestrado) – UFPB/CCEN

1. Informática. 2. Autenticação biométrica. 3. Transformada Cosseno Discreta (DCT). 4. Reconhecimento facial.

UFPB/BC CDU: 004(043)

Este trabalho é dedicado a JanKees van der Poel e a Silvana Ferreira Fernandes Galvão (in memoriam).

AGRADECIMENTOS

A Deus.

À minha família, por ter me apoiado e tolerado todos os momentos em que estive ausente me dedicando ao mestrado.

A JanKees van der Poel, que apesar de já não estar mais entre nós, foi quem me fez seguir o rumo acadêmico com seus valorosos conselhos. Sempre serei grato por suas orientações e é uma pena que ele não possa ter ciência disto.

Aos meus orientadores, Ed Porto e Leonardo Vidal, por terem estendido a mão e me aceitado de bom grado.

Aos meus amigos e companheiros de trabalho da Vsoft, pela paciência, preocupação e dicas. Em especial, ao meu amigo Raphael Marques, que por várias vezes leu a dissertação em busca de erros ou formas de melhorar o texto sem ter a mínima obrigação de fazer isto. Obrigado mesmo.

Realizar um curso de mestrado não é apenas pesquisar e escrever uma dissertação, nós passamos por uma experiência de docência também. A todos aqueles que tiveram aulas comigo e guardaram minhas dicas, obrigado. Em especial, a dois deles, Iron Araújo e Igor Malheiros, que além de alunos se tornaram amigos sempre dispostos a ouvir minhas opiniões, mesmo quando nós divergíamos.

Aos meus companheiros de graduação (e agora de mestrado), Amanda Cavalcanti, Eduardo Freire, José Rogério, Pollyane Carvalho, Ricardo Mendes e Yuri Malheiros, pelo suporte mútuo ao longo do trabalho, obrigado. A dois deles em especial: Amanda, por toda a paciência e momentos de descontração comigo mesmo estando do outro lado do oceano e a Rogério por nunca me deixar fraquejar na fase final do trabalho. Sem vocês eu não teria conseguido.

RESUMO

Sistemas de identificação confiáveis tornaram-se componentes chaves de várias aplicações que disponibilizam serviços para usuários autenticados. Uma vez que métodos de autenticação tradicionais (como os que utilizam senhas ou *smartcards*) podem ser manipulados com o objetivo de burlar os sistemas, métodos de autenticação biométrica vêm recebendo mais atenção nos últimos anos. Um dos traços biométricos é a face. O problema do reconhecimento de faces em vídeo e foto é objeto de pesquisa, uma vez que existem muitos fatores que influenciam na detecção e no reconhecimento, tais como: iluminação, posição da face, imagem ao fundo, diferentes expressões faciais, etc. É possível realizar reconhecimento facial utilizando a Transformada Cosseno Discreta (DCT). Com o intuito de adequar um Sistema de Detecção e Reconhecimento de Faces a ambientes não controlados, neste trabalho foram desenvolvidas duas melhorias para ele: um módulo normalizador de imagens em relação à rotação e à escala e uma modificação na etapa de seleção de atributos, por meio da inserção de um filtro passa-baixas não ideal. O sistema e suas modificações foram testados nos bancos de faces UFPB, ORL, Yale, GTAV e Vsoft, desenvolvido especialmente para o trabalho. Os testes mostraram a eficácia do módulo de normalização da imagem, mas ainda assim o sistema não é adequado para qualquer ambiente.

Palavras-chave: Autenticação biométrica. Reconhecimento Facial. DCT.

ABSTRACT

Reliable identification systems have become key components in many applications that provide services to authenticated users. Since traditional authentication methods (such as using passwords or smartcards) can be manipulated in order to bypass the systems, biometric authentication methods have been receiving more attention in recent years. One of the biometric traits is the face. The problem of recognizing faces in video and photo still is an object of research, since there are many factors that influence the detection and recognition, such as lighting, position of the face, the background image, different facial expressions, etc. One can perform face recognition using Discrete Cosine Transform (DCT). In order to adjust a face recognition system to uncontrolled environments, two improvements for it were developed in this work: a image normalization module with respect to rotation and scale, and a change in the feature extraction module through the insertion of a non-ideal low-pass filter. The system and its modifications were tested on the following face databases: UFPB, ORL, Yale, and VSoft GTAV, developed specially for the job. Tests showed the efficiency of the image normalization module, but the system still is not adequate for every environment.

Keywords: Biometric authentication. Face recognition. DCT.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO

A proliferação de serviços que necessitam de autenticação gerou uma demanda por novos métodos para estabelecer a identidade usuários (ROSS *et al.*, 2006). O problema de estabelecer a identidade de um indivíduo pode ser dividido em dois: autenticação e identificação. No primeiro, um indivíduo alega uma identidade e o sistema deve confirmar ou negar. No segundo, o sistema deve estabelecer uma identidade para um indivíduo desconhecido (THIAN, 2001) (JAIN *et al.*, 2011). Métodos tradicionais para estabelecer a identidade de um usuário incluem mecanismos baseados em conhecimento (por exemplo, senhas) e mecanismos baseados em *tokens* (por exemplo, cartões de identidade). Porém, tais mecanismos podem ser perdidos, roubados ou até mesmo manipulados com o objetivo de burlar o sistema. Neste contexto, a autenticação e a identificação por Biometria (ROSS *et al.*, 2006) surgem como alternativas.

A autenticação biométrica oferece um mecanismo de autenticação mais confiável utilizando traços físicos (como a íris ou impressão digital) ou comportamentais (como o modo de escrever) que permitem identificar usuários baseados em suas características naturais. Assim, é possível estabelecer a identidade de um usuário baseado em quem ele é ao invés de em o que ele possui ou de que ele lembra. A face é um dos traços biométricos físicos menos intrusivos.

A área de análise de faces pode ser dividida em diversas subáreas, como reconhecimento de face, detecção/localização de face, reconhecimento de expressões faciais e análise de poses (ZHAO *et al.*, 2003). É importante diferenciar detecção e reconhecimento. O reconhecimento de faces consiste em identificar um indivíduo por intermédio da análise de sua face, comparando-a com outras faces pré-rotuladas. A detecção ou localização de faces é a determinação da presença e posição espacial de cada face existente em uma imagem. A detecção de face frequentemente é utilizada como uma etapa inicial para o reconhecimento de faces.

O problema do reconhecimento de faces em vídeo e foto não é trivial, pois existem muitos fatores que influenciam na detecção e no reconhecimento, tais como: iluminação, posição da face, imagem ao fundo, diferentes expressões faciais. Assim, é comum encontrar na literatura sistemas de reconhecimento de faces em ambientes controlados (ZHAO *et al.*, 2003) (SELLAHEWA e JASSIM, 2010) (SINGH *et al.*, 2010).

Existem muitos métodos para o reconhecimento de faces propostos na literatura. Basicamente eles se dividem em métodos holísticos, métodos analíticos (não holísticos) e métodos híbridos; cada um destes métodos possui vantagens e desvantagens (ZHAO e CHELLAPPA, 2002) (CHAUDHARI e KALE, 2010). Os métodos holísticos são aqueles que analisam a face como um todo, normalmente

gerando modelos de faces e realizando comparações entre esses modelos para efetuar o reconhecimento facial. O Sistema de Detecção e Reconhecimento Facial (SDRF) desenvolvido em (OMAIA, 2009), doravante chamado apenas de SDRF, é um exemplo de sistema holístico. Os métodos analíticos objetivam construir um espaço de características no qual a manipulação da face se torna mais simples. O método de (BRUNELLI e POGGIO, 1993), que utiliza as características geométricas da face (posições de olhos, boca, nariz, queixo, etc., e suas relações) é um exemplo de método analítico.

É possível utilizar a Transformada Cosseno Discreta (DCT) (KHAYAM, 2003) para realizar reconhecimento facial holístico, como visto em (HAFED e LEVINE, 2001), (MATOS, 2008) e (OMAIA, 2009). Estes sistemas são passíveis de melhoramentos, assim como outros presentes na literatura.

Por exemplo, o SDRF de (OMAIA, 2009), possui resultados em bancos de faces controlados, porém, é necessário investigar se ele é um sistema adequado para qualquer tipo de ambiente e por que a sua taxa de acertos do reconhecimento facial foi tão baixa quando testado em vídeo coletivo (com mais de uma pessoa presente no vídeo), de modo que ele possa ser utilizado comercialmente. Ele foi escolhido, em detrimento de outros sistemas, para ser melhorado no presente trabalho porque:

a) É um sistema eficaz, que chega a atingir mais de 90% em bancos de faces presentes na literatura. Estes resultados são detalhados na seção 3.6.
b) É fácil investigar soluções para os casos de falha do SDRF, uma vez que estes são conhecidos e estão bem documentados em (OMAIA, 2009). Os casos de erros também são mencionados na seção 3.6.
c) O código do SDRF é de fácil acesso.

Assim, o objetivo geral deste trabalho é adequar o SDRF a ambientes não controlados. Para tanto, foram definidos os seguintes objetivos específicos:

1. Desenvolver um módulo de normalização de imagens em relação à rotação e escala. Este módulo deverá retirar as inclinações das imagens de face, de modo que o ângulo entre os olhos seja de 0º. Ele também corrigirá as diferenças de escala entre as imagens, de modo que as imagens de face fiquem com a mesma distância em relação à câmera.
2. Modificar a etapa de seleção de atributos do SDRF por meio da inserção de um filtro passa baixas não ideal.

Estrutura da Dissertação

Esta dissertação está organizada em seis capítulos. No Capítulo 1 é apresentada a Fundamentação Teórica necessária para o entendimento do trabalho,

na qual se incluem definições de processamento digital de imagens e reconhecimento facial. No Capítulo 2 são descritos os bancos de faces utilizados no trabalho e nos trabalhos correlatos. No Capítulo 3 são elencados trabalhos correlatos. No Capítulo 4 é apresentado o desenvolvimento do trabalho. No Capítulo 5 são apresentados e discutidos os resultados alcançados. No Capítulo 6 são apresentadas as considerações finais bem como ponderações sobre trabalhos futuros.

1. FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo, são abordados temas necessários à compreensão das melhorias propostas. Tais temas são: o processamento digital de imagens (Seção 1.1), a DCT e o domínio da frequência (Seção 1.2), o reconhecimento de padrões (Seção 1.3) e a detecção e o reconhecimento de faces (Seção 1.4).

1.1 Processamento Digital de Imagens

Imagens são sinais, ou seja, funções que conduzem alguma informação a respeito de algo com um interesse. Uma imagem monocromática é uma função $f(x, y)$ em que x e y representam as suas coordenadas espaciais e o valor de $f(x, y)$ representa um valor de intensidade luminosa na posição x e y da imagem, geralmente chamada nível de cinza. Para serem representadas no computador, é necessário digitalizar as imagens, o que poder ser feito por meio dos processos de amostragem e quantização (BATISTA, 2005). Amostragem refere-se a capturar em intervalos de tempo bem definidos o sinal analógico que representa a imagem, enquanto que quantização refere-se a definir os valores que o sinal pode assumir no meio digital (LORDÃO, 2009). O resultado de tais processos é uma *imagem digital monocromática*, representada da seguinte forma:

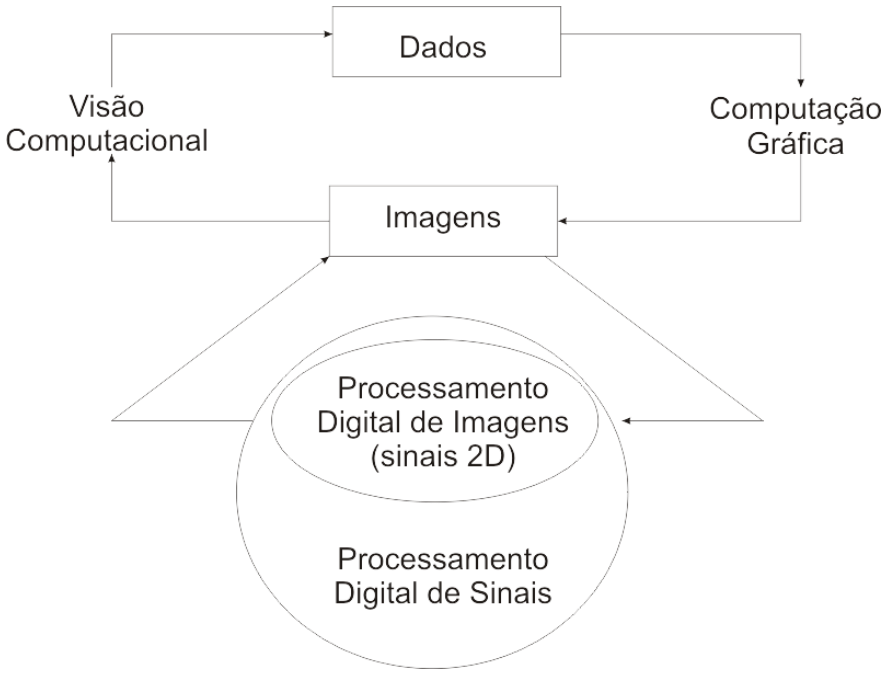
$$\begin{pmatrix} f(0,0) & \cdots & f(0, C-1) \\ \vdots & \ddots & \vdots \\ f(L-1, 0) & \cdots & f(L-1, C-1) \end{pmatrix}.$$

Os valores L e C representam, respectivamente, o número máximo de linhas e colunas da imagem digitalizada, e são determinados no processo de digitalização. Cada elemento da matriz $f(x,\ y)$ é conhecido como pixel, e, como dito, representa o valor do nível de cinza naquele ponto. O número total de níveis de cinza N também é determinado no processo de digitalização. O menor valor é 0, e representa o preto, enquanto o maior valor, $N-1$, representa o branco.

O Processamento Digital de Imagens (PDI) é o campo da Ciência da Computação que se dedica ao processamento de imagens digitais em um computador digital (GONZALEZ *et al.*, 2003). Em síntese, pode-se dizer que o PDI consiste em qualquer forma de processamento de dados no qual a entrada e saída são imagens tais como fotografias ou quadros de vídeo. Ao contrário do tratamento de imagens, que se preocupa somente com a manipulação de figuras para sua representação final, o PDI é um estágio para novos processamentos de dados tais como aprendizagem de máquina ou reconhecimento de padrões (BATISTA, 2005).

O Processamento Digital de Imagens é multidisciplinar (BATISTA, 2005), e tem relações muito próximas com as áreas de Computação Gráfica (a qual consiste em obter imagens a partir de dados) e Visão Computacional (a qual consiste em obter dados a partir de imagens), como mostra a Figura 1. Muitas vezes um sistema de PDI exerce tarefas dessas duas áreas ao mesmo tempo. Outros exemplos de áreas de aplicação de PDI são Inteligência Artificial, Animação, Reconhecimento de Padrões e a Indústria do Entretenimento.

Figura 1 - Processamento digital de imagens e áreas correlatas.



Fonte: (BATISTA, 2005).

1.1.1 Transformações Geométricas em Imagens

As transformações geométricas são operações de processamento digital de imagens que permitem modificar as coordenadas dos pixels de uma imagem, preservando a sua forma. As coordenadas (i, j) de uma imagem de entrada são modificadas para as coordenadas (i', j') de uma imagem de saída (BATISTA, 2005) (MARQUES FILHO e VIEIRA NETO, 1999) (SOLOMON e BRECKON, 2011). Nesta seção são detalhadas as operações de rotação e redimensionamento de imagens, pois elas são utilizadas no presente trabalho.

A operação de rotação de um ângulo θ em torno de um ponto (i_R, j_R) qualquer da imagem é dada pela Equação 1.1.

$$i' = (i - i_R)\cos\theta - (j - i_R)\sin\theta + i_R$$
$$j' = (i - i_R)\sin\theta + (j - i_R)\cos\theta + i_R \quad (1.1)$$

Sendo:

(i, j) as coordenadas do pixel original.

(i', j') as coordenadas do pixel rotacionado.

(i_R, j_R) eixo de rotação.

θ o ângulo de rotação.

Na Figura 2 é apresentado um exemplo de rotação onde a imagem original foi rotacionada em 45º em torno do centro da imagem.

Figura 2 - Exemplo de rotação.

(a) Imagem original. **(b) Imagem rotacionada.**

A operação de redimensionamento consiste em reduzir ou ampliar a imagem por um fator, que pode ser igual para as dimensões horizontal e vertical. A operação pode ser feita nas linhas de acordo com a Equação 1.2 e pode ser adaptada para ser aplicada nas colunas da imagem.

$$i' = fatorDeEscala \times (i - i_R) + i_R \quad (1.2)$$

Sendo:

$fatorDeEscala$ o fator de redimensionamento da imagem. Se $fatorDeEscala > 1$ a imagem será expandida, se $0 < fatorDeEscala < 1$ a imagem será contraída.

i a coordenada do pixel original.

i' a coordenada do pixel redimensionado.

i_R centro de redimensionamento, isto é, após a operação, todos os pontos das linha se afastam ou se aproximam de i_R.

1.1.2 Métodos para a Normalização de Iluminação de Imagens

Na seção que trata do Reconhecimento Facial (1.3), será mostrado que no processo de reconhecimento, as imagens podem passar por uma etapa de normalização, a qual é aplicada logo após a detecção da face.

Dentre as formas de normalização, existe a normalização fotométrica, que consiste em consiste em melhorar as condições da imagem de modo que ruídos, sombras e variações de contraste ou de brilho não afetem o processo de reconhecimento.

Algumas técnicas de normalização de iluminação de imagens são: Equalização de Histograma, Transformação Não-Linear e Log About. Estas técnicas são apresentadas no presente trabalho porque são utilizadas nos testes da alteração da etapa de seleção de atributos.

Equalização de Histograma

A Equalização de Histograma tem como objetivo redistribuir os valores de tons de cinza dos pixels em uma imagem, de modo a obter um histograma mais uniforme. A forma mais comum de se equalizar um histograma é utilizar a Função de Distribuição Acumulada como visto em (MARQUES FILHO e VIEIRA NETO, 1999) e que está expressa na Equação (1.3).

$$S_k = T(r_k) = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} P_r\,(r_j). \qquad (1.3)$$

Sendo:

n o número total de pixels na imagem.

$0 < r_k < 1$.

$K = 0,1,\ldots,L-1$, em que L é o número de níveis de cinza da imagem digitalizada;

$P_r(r_j)$ é a .probabilidade do $j-ésimo$ nível de cinza.

Transformações Não Lineares (Log)

Transformações não-lineares são muito utilizadas na área de visão computacional.

A função logarítmica pode estender os níveis de cinza mais baixos e comprimir os níveis de cinza mais altos, o que pode melhorar a iluminação deficiente na sua essência. É muito útil para sombras e imagens não-uniformes (LIU *et al.*, 2002).

A Equação (1.4) define a função logarítmica utilizada em (LIU *et al.*, 2002).

$$g(x,y) = a + \frac{\ln(f(x,y)+1)}{b\ \ln c} \tag{1.4}$$

Sendo:

$f(x,y)$ a imagem de entrada.

$g(x,y)$ a imagem de saída.

a constante e igual a 10.

b constante e igual a 0,017.

c constante e igual a 10.

O Método *Log About*

O método *Log About,* proposto em (LIU *et al.*, 2002), tem o objetivo de compensar problemas de iluminação em sistemas de detecção de face. Este método consiste na aplicação de um filtro passa-alta na imagem original, seguida de uma transformação logarítmica como a da Equação (1.4)

Na Figura 3 é apresentado um exemplo de imagem que foi submetida ao método *Log About*. O filtro passa alta é definido pela seguinte máscara de convolução:

$$A = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Esta máscara é a utilizada em (LIU *et al.*, 2002). Na Figura 3 é apresentada uma comparação entre os resultados dos métodos Log e *Log About*.

Figura 3 – Resultados dos métodos Log e *Log About*.

(a) Imagem original.

(b) Resultado após Log.

(c) Resultado após *Log About.*

1.2 O Domínio da Frequência e a Transformada Cosseno Discreta

Converter uma imagem do domínio do espaço para o domínio da frequência pode ser necessário para diversos fins. Usam-se operações matemáticas chamadas transformadas (ou transformações) para efetuar a conversão (PARKER, 1996).

Uma transformação consiste em mapear (transformar) um conjunto de coordenadas para outro. No caso de funções, isto nada mais é que alterar o domínio.

Existem diversas transformadas na literatura e a escolha entre elas depende da natureza do problema. A Transformada Cosseno Discreta é uma das transformadas capazes de converter um sinal do domínio do espaço para o domínio da frequência. Assim como outras transformadas, a DCT tenta descorrelacionar dados. Ela parte da ideia de que qualquer sinal discreto $x[n]$ pode ser decomposto em um somatório de n funções cosseno, cada função cosseno com sua amplitude, frequência e fase. A DCT, então, transforma um sinal discreto x em um conjunto de coeficientes X em que cada coeficiente $X[k]$ expressa a importância de uma onda cosseno para a formação do sinal original (KHAYAM, 2003).

A DCT possui algumas variações. Seja $f[m,n]$ um sinal discreto dado, a transformação para o domínio da frequência pela DCT-2D produz uma sequência, $F[k,l]$, dada pela Equação 1.5:

$$F[k,l] = c_k c_l \sum_{m=0}^{L-1} \sum_{n=0}^{C-1} f[m,n] \cos\frac{(2m+1)k\pi}{2L} \cos\frac{(2n+1)l\pi}{2C}. \quad (1.5)$$

Sendo:

L o número de linhas da imagem.

C o número de colunas da imagem.

$0 \leq k \leq L-1$

$0 \leq l \leq C-1$

$$c_k = \begin{cases} \frac{1}{\sqrt{L}}, se\ k = 0 \\ \sqrt{\frac{2}{L}}, se\ 1 \leq k \leq L-1 \end{cases}$$

$$c_l = \begin{cases} \frac{1}{\sqrt{C}}, se\ l = 0 \\ \sqrt{\frac{2}{C}}, se\ 1 \leq l \leq C-1 \end{cases}$$

O sinal original, $f[m,n]$ pode ser obtido a partir da transformação inversa IDCT, dada pela Equação 1.6:

$$f[m,n] = c_m c_n \sum_{k=0}^{L-1} \sum_{l=0}^{C-1} F[k,l] \cos\frac{(2k+1)m\pi}{2L} \cos\frac{(2l+1)n\pi}{2C}. \quad (1.6)$$

Sendo:

L o número de linhas da imagem.

C o número de colunas da imagem.

$0 \leq m \leq L-1$

$0 \leq n \leq C-1$

$$c_m = \begin{cases} \frac{1}{\sqrt{L}}, se\ m = 0 \\ \sqrt{\frac{2}{L}}, se\ 1 \leq m \leq L-1 \end{cases}$$

$$c_n = \begin{cases} \frac{1}{\sqrt{C}}, se\ n = 0 \\ \sqrt{\frac{2}{C}}, se\ 1 \leq n \leq C-1 \end{cases}$$

Os coeficientes mais próximos da origem gerados pela DCT-2D referem-se às frequências mais baixas do sinal, que representam as características gerais, normalmente as mais representativas do sinal original. Os últimos coeficientes referem-se às frequências mais altas do sinal, que geralmente representam os detalhes, as bordas ou o ruído presente no sinal. Assim, se o nível de cinza de uma imagem muda vagarosamente ao longo de suas colunas, então essa imagem seria representada no domínio DCT como uma imagem contendo predominantemente cossenos de baixas frequências, como mostrado na Figura 4 Algo que muda de nível de cinza abruptamente, como uma borda, será representado por cossenos de alta frequência (Figura 5) (KHAYAM, 2003).

Figura 4 - Mudança abrupta de nível de cinza.

(a) Imagem original **(b) DCT**

Algo que muda de nível de cinza abruptamente, como uma borda, será representado por cossenos de alta frequência, como mostrado na Figura 5 (KHAYAM, 2003).

Figura 5 - Mudança suave de nível de cinza.

(a) Imagem original **(b) DCT**

1.3 Reconhecimento de Padrões

Reconhecimento de padrões consiste em observar dados e tomar ações baseadas na categoria destes (DUDA *et al.*, 2001) (AHMAD *et al.*, 2011). Embora natural para os seres humanos, construir um sistema automatizado de reconhecimento de padrões é uma tarefa complexa.

Um padrão é algo que segue alguma regra, ou conjunto de regras, de forma que seja possível distingui-lo de outros padrões. Uma impressão digital ou uma face humana são exemplos de padrões (JAIN *et al.*, 2000).

As seguintes etapas de processamento geralmente estão presentes em sistemas que efetuam reconhecimento de padrões: modelagem, pré-processamento da entrada, extração de características e classificação (DUDA *et al.*, 2001). No pré-processamento, operações para retirada de ruído e redução de informação são feitas com o objetivo de simplificar as etapas posteriores. A extração de características (ou seleção de atributos) é a etapa em que são extraídas informações úteis para o processo de reconhecimento. A etapa final é a classificação, em que as características extraídas dos dados de entrada são analisadas para que o objeto em análise seja colocado em uma determinada categoria.

Duas estratégias podem ser utilizadas ao classificar padrões: classificação supervisionada e classificação não supervisionada (WEBB, 2002). Na classificação supervisionada, o padrão de entrada é identificado como um membro de uma classe pré-definida pelo projetista do sistema. Na classificação não supervisionada, o padrão é determinado por uma fronteira de classe desconhecida, ou seja, as classes são aprendidas baseadas nas similaridades dos padrões.

Existem vários métodos para efetuar o reconhecimento de padrões, tais como: casamento sintático ou estrutural, redes neurais, classificação estatística e *template matching* (JAIN *et al.*, 2000). Dois métodos, a classificação por vizinho mais próximo e

o *template matching*, serão detalhados nas próximas seções, pois eles são utilizados no presente trabalho.

1.3.1 Classificador do Vizinho Mais Próximo

A Classificação por vizinhos mais próximos (do inglês, *k-nn, k-nearest neighbors*) é um método de classificação baseado nos objetos mais próximos em um espaço de atributos. O algoritmo é simples: um objeto de consulta é classificado de acordo com a sua distância em relação a outros objetos chamados objetos de treinamento (seus vizinhos). O Objeto de consulta é atribuído à classe dos seus k vizinhos mais próximos. O classificador do vizinho mais próximo (NN) é um caso particular do classificador por vizinhos mais próximos no qual $k = 1$; nele o objeto de consulta é atribuído à classe do vizinho mais próximo (DUDA *et al.*, 2001). A distância é alguma métrica de comparação, como a distância Euclidiana.

1.3.2 Template Matching

Em reconhecimento de padrões, *matching* (casamento) é uma operação utilizada para determinar a similaridade entre dois objetos do mesmo tipo. No *template matching* um objeto de consulta é comparado a padrões previamente armazenados (*templates*), utilizando alguma função de similaridade, como correlação. O resultado dessa função é um valor que é comparado a algum limiar. Sendo superior ao limiar, o objeto pertence à classe; se inferior, ele não pertence (JAIN *et al.*, 2000). *Template Matching* é eficiente para aplicações em que os objetos a serem classificados têm poucas distorções ou ruídos.

No caso de imagens, o *template matching* pode ser implementado como um algoritmo de janela deslizante. Seja $F(x, y)$ a imagem de consulta, em que (x, y) representa as coordenadas de cada pixel da imagem, e $T(x_t, y_t)$ a imagem de *template*, em que (x_t, y_t) são as coordenadas do *template*. O centro do template é movido sobre cada ponto (x, y) da imagem de consulta, e uma métrica de comparação, é calculada, como, por exemplo, a correlação entre o centro do *template* e os pixels que compõem a sua vizinhança. Seguindo no exemplo da correlação, após o template ter deslizado por toda a imagem de entrada, a melhor posição será aquela que apresentar o maior valor de correlação.?

1.3.3 Métricas de Comparação

Nas seções anteriores foi mencionado o fato de que métricas de comparação são utilizadas durante a fase de classificação de um sistema reconhecedor de padrões. Estas métricas podem ser de distância (ou seja, quanto menor, mais

parecido) ou similaridade (quanto maior, mais parecido). Nesta seção, são detalhadas as métricas de comparação utilizadas no presente trabalho: correlação e as distâncias Euclidiana e de Minkowski.

Correlação

A correlação é uma medida de similaridade entre dois sinais geralmente utilizada para procurar um sinal menor (*template*) dentro de um sinal maior. A correlação pode ser aplicada entre imagens digitais de acordo com a equação (1.7).

$$C_{fg}[i,j] = \sum_{k=1}^{m}\sum_{l=1}^{n}(g[k+i,l+j] - \bar{g})(f[k,l] - \overline{f[k,l]}) \tag{1.7}$$

Sendo:

g a imagem de *template*;

$\bar{g}$ a média da imagem do *template*;

f a imagem em que o olho será procurado;

$\overline{f[k,l]}$ a média na vizinhança; e

C_{fg} a correlação entre f e g..

A correlação pode ainda ser feita de maneira normalizada nos casos que o brilho das imagens varie muito, de acordo com a equação (1.8).

$$M[i,j] = \frac{C_{fg}\,[i,j]}{\sqrt{\sum_{k=1}^{m}\sum_{l=1}^{n}(f[k,l] - \overline{f[k,l]})^2 \sum_{k=1}^{m}\sum_{l=1}^{n}(g[k+i,l+j] - \bar{g})^2}} \tag{1.8}$$

Sendo:

$M[i,j]$ a correlação normalizada entre f e g.

Distância de Minkowski

Dados dois pontos $P_1\ (x_1, y_1)$ e $P_2\ (x_2, y_2)$, a distância de Minkowski de ordem p entre esses pontos é obtida a partir da equação (1.9)

$$D = \sqrt[p]{|x_1 - x_2|^p + |y_1 - y_2|^p} \tag{1.9}$$

Distância Euclidiana

A distância euclidiana entre dois pontos é um caso especial da distância de Minkowski onde $p = 2$. A operação é mostrada na Equação (1.10), considerando os mesmos dois pontos $P_1\ (x_1, y_1)$ e $P_2\ (x_2, y_2)$.

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1.10)$$

1.4 Detecção e Reconhecimento de Faces

Embora o reconhecimento facial seja uma tarefa que os seres humanos fazem constantemente sem esforços, ele ainda consiste em uma tarefa complexa para o computador.

Por muitas razões o reconhecimento facial é objeto de pesquisa, sendo algumas dessas razões a preocupação crescente por segurança pública, a necessidade de verificar a identidade de um indivíduo para fins de acesso (físico ou virtual) e a análise de dados para fins entretenimento multimídia. O primeiro sistema de reconhecimento facial data da década de 1960 e desde então vários outros sistemas foram desenvolvidos, mesmo assim, ainda há obstáculos a serem vencidos no Reconhecimento Facial, especialmente no que diz respeito a ambientes não controlados .

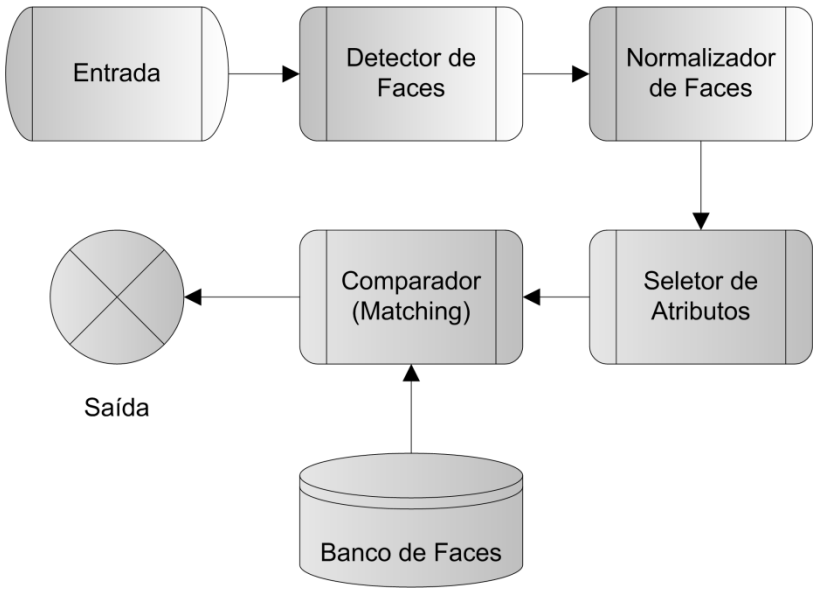
A pesquisa nesta área é motivada pelos desafios fundamentais que o problema possui e também pela grande quantidade de aplicações em que a identificação de um ser humano é necessária. O reconhecimento facial tem se tornado cada vez mais importante e tem se beneficiado da crescente demanda por segurança em diversas atividades humanas e também dos rápidos avanços em tecnologia como câmeras digitais, Internet e dispositivos móveis. (JAIN e LI, 2011).

O reconhecimento facial possui vantagens em relação aos outros traços biométricos. Além de ser natural e não intrusivo, uma face pode ser capturada à distância e de uma forma que o usuário não perceba.

Computacionalmente, o reconhecimento facial consiste em identificar um indivíduo por intermédio da análise de sua face, comparando-a com outras faces pré-rotuladas/cadastradas em um banco de dados.

Em geral, a arquitetura de um sistema de reconhecimento facial consiste na mostrada na Figura 6 (JAIN e LI, 2011). Ela é detalhada nos parágrafos seguintes.

Figura 6 - Arquitetura geral de um sistema de reconhecimento facial.



Fonte: Adaptado de (JAIN e LI, 2011)

O sistema de reconhecimento facial recebe como entrada uma imagem ou vídeo contendo uma ou mais pessoas. Nem sempre apenas a imagem facial é fornecida, isto faz com seja necessário um módulo detector de faces. A detecção ou localização de faces é a determinação da presença e posição espacial de cada face existente em uma imagem. A detecção de face frequentemente é utilizada como uma etapa inicial para o reconhecimento.

O módulo detector de faces retorna uma imagem contendo a face detectada. Esta imagem serve como entrada do módulo normalizador, que pode operar padronizando as dimensões, a inclinação e a iluminação das imagens.

Uma vez normalizada, a imagem é passada para o módulo extrator de características, o funcionamento do extrator de características depende do método usado para o reconhecimento facial, como será mostrado mais adiante. O extrator produz um vetor de características, que é utilizado no módulo comparador para produzir a saída do sistema. Essa saída depende do tipo de reconhecimento que se está fazendo, pois sistemas de reconhecimento facial podem operar em dois modos: identificação ou verificação. No processo de identificação facial, uma comparação 1:N é feita, de modo que a entrada para o sistema é uma face desconhecida e ela é associada a uma das faces previamente cadastradas no banco de faces. No processo de verificação, uma comparação 1:1 é feita, de modo que a entrada para o sistema é uma face desconhecida e uma identidade alegada pelo usuário, e a saída é uma valor de verdadeiro ou falso que aceita ou rejeita tal alegação (ZHAO *et al.*, 2003).

O desempenho dos sistemas de reconhecimento facial depende de uma variedade de fatores, tais como iluminação, pose, expressões faciais, idade dos sujeitos, cabelo e acessórios. Baseado nesses fatores, os sistemas podem ser divididos em duas categorias: *cooperative user scenarios* e *noncooperative user scenarios*. Na primeira, o usuário coopera com o sistema apresentando sua face de uma maneira apropriada (por exemplo, de frente para a câmera, com uma expressão neutra e de olhos abertos). Na segunda, o usuário não tem conhecimento de que está sendo identificado. Um exemplo dessa situação é a procura de um bandido em uma estação de metrô (LI e JAIN, 2011).

Por sofrer a influência de fatores como os citados anteriormente, é comum encontrar na literatura sistemas de reconhecimento de faces em ambientes controlados (ZHAO *et al.*, 2003) (SELLAHEWA e JASSIM, 2010) (SINGH *et al.*, 2010).

Existem muitas técnicas para o reconhecimento de faces propostas na literatura. Basicamente, elas se dividem em duas linhas: analíticos e holísticos.

O objetivo dos métodos analíticos (ou não holísticos) é construir um espaço de características no qual a manipulação da face se torna mais simples. Podem-se citar como exemplo as características geométricas (BRUNELLI e POGGIO, 1993). Neste método, características da face (como olhos, nariz, boca e queixo) são detectadas e propriedades e relações existentes entre elas são usadas como descritores para que o reconhecimento facial seja efetuado.

Os métodos holísticos analisam a face como um todo, sem identificar características individuais, normalmente gerando modelos e utilizando classificadores para treinar o algoritmo.

Além disso, podem-se combinar os dois tipos de métodos para formar métodos híbridos, uma vez que os classificadores dos métodos holísticos podem ser usados em conjunto com as características obtidas pelos métodos analíticos (ZHAO *et al.*, 2003).

O primeiro sistema de reconhecimento facial automático foi desenvolvido em (KANADE, 1973). Em seguida, o método conhecido como *eigenfaces* (TURK e PENTLAND, 1991) revigorou a pesquisa nessa área. Podem ser citados como outros marcos na área foram o método *fisherfaces* (BELHUMEUR *et al.*, 1997), o uso de filtros locais como *Gabor Jets* (WISKOTT *et al.*, 1997) para obter melhores características faciais e o desenvolvimento do método *AdaBoost* (VIOLA e JONES, 2001) para fazer detecção em tempo real.

As técnicas de reconhecimento facial evoluíram de modo que atualmente a detecção e o rastreamento (*tracking*) de faces é considerado um problema bem resolvido para *cooperative user scenarios*. Ainda em se tratando de *cooperative user scenarios*, a verificação 1:1 funciona bem com imagens frontais desde que as faces

sejam capturadas com boa resolução, embora isto não possa ser dito para verificação 1:N. No caso de *non-cooperative scenarios*, o reconhecimento facial continua sendo desafiador. Técnicas mais atuais incluem o reconhecimento de faces utilizando luz infravermelha (LI e YI, 2011) e o reconhecimento de faces utilizando imagens 3D (KAKADIARIS *et al.*, 2011).

1.5 Considerações Finais do Capítulo

Neste capítulo foram mostrados conceitos relacionados ao Processamento Digital de Imagens, Reconhecimento de Padrões e Reconhecimento Facial. As operações utilizadas no presente trabalho, como as operações de redimensionamento e rotação de imagens e as operações de normalização de iluminação de imagens, receberam um maior destaque. A transformada cosseno discreta (DCT) e o domínio da frequência foram descritos porque são utilizados no processo de reconhecimento facial do sistema. O Reconhecimento de Padrões foi mencionado brevemente para um melhor entendimento do Reconhecimento Facial, já que este é um caso particular daquele. O Capítulo 2 detalhará todos os bancos de faces utilizados no trabalho.

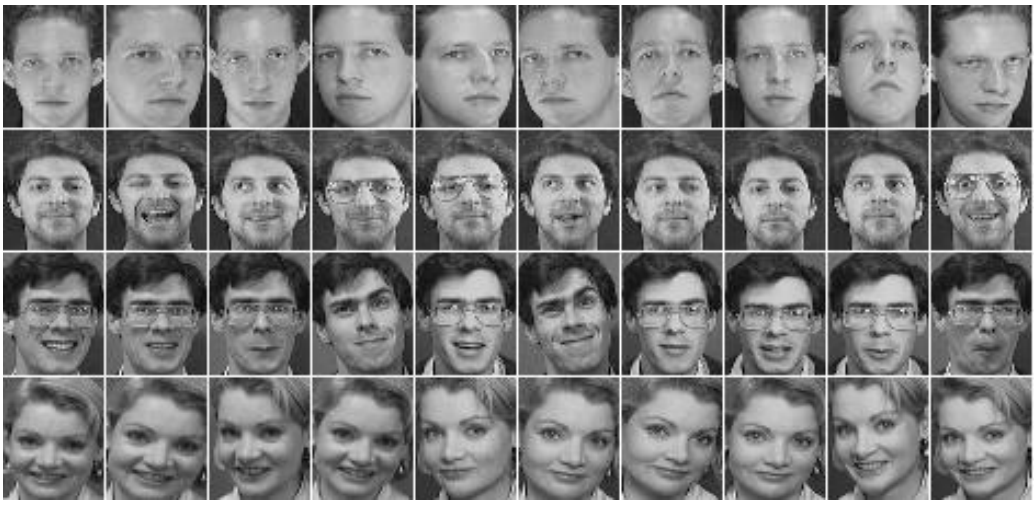
2. BANCOS DE FACES

Neste Capítulo são descritos os bancos de faces utilizados neste trabalho e nos trabalhos correlatos, a saber: ORL (AT&T LABORATORIES, 2002), UFPB (OMAIA, 2009), *Yale Faces Database* (GEORGHIADES, 1997) e *GTAV Face Database* (TARRÉS e RAMA, 2011). Os bancos Yale e GTAV foram escolhidos para testar a modificação da Seção 4.2 em bancos de iluminação não uniforme. Também é relatado o desenvolvimento de um banco de faces próprio, o Vsoft.

2.1 Banco de Faces ORL

O banco de faces ORL (do inglês, *Olivetti Research Lab*) foi desenvolvido nos laboratórios da Olivetti em Cambridge, Inglaterra. Ele é composto de 400 fotos de faces de 40 pessoas diferentes, em que cada pessoa possui 10 fotos em diferentes poses, com pequenas variações de iluminação, expressões faciais, e acessórios. Todas as imagens foram fotografadas contra um fundo homogêneo, e as faces estão todas verticais e frontais, com pequenas variações de angulação. Todas as imagens possuem tamanho 92 x 112 pixels e estão em escala de cinza com 256 níveis de cinza. Na Figura 7 são apresentados exemplos de faces do banco ORL.

Figura 7 - Amostras do banco de faces ORL.



Fonte: (OMAIA, 2009).

2.2 Banco de Faces UFPB

O banco de faces UFPB foi criado para validar o SDRF desenvolvido em (OMAIA, 2009). Ele possui duas versões, sendo uma em foto e outra em vídeo. No banco de faces UFPB-Fotos, 20 imagens de 40 pessoas diferentes foram selecionadas, totalizando 800 imagens de faces. Todas foram capturadas

frontalmente, com pouca rotação. Todas as imagens possuem tamanho 128 x 128 pixels e estão em escala de cinza com 256 níveis de cinza. Na Figura 8 são apresentados exemplos de faces do banco UFPB-Fotos.

Figura 8 - Amostras do banco de faces UFPB-Fotos.

Fonte: (OMAIA, 2009).

O banco de faces UFPB-Vídeo possui um vídeo de aproximadamente 10 segundos para cada pessoa. Os vídeos são coloridos no formato YUV, possuem resolução máxima de 1920x1080 pixels, e foram capturados a 30 quadros por segundo, a partir de uma câmera de vídeo de alta definição (*full HD*). Nos vídeos, cada pessoa executa pequenos movimentos circulares com a cabeça e varia um pouco as expressões faciais. Este banco de dados ainda possui um vídeo coletivo que exibe 22 pessoas, das 40 pessoas do banco completo, executando suas tarefas cotidianas nos laboratórios. No presente trabalho os vídeos individuais não são avaliados uma vez que o banco UFPB-Fotos foi gerado a partir de capturas desses vídeos individuais. Isto gera um viés no reconhecimento facial desses vídeos, pois o banco UFPB-Fotos é utilizado na fase de treinamento para reconhecer os vídeos do banco UFPB-Vídeos.

2.3 Banco de Faces Yale

O banco de faces Yale contém 165 imagens distribuídas entre 15 pessoas com 11 poses que incluem variações de iluminação e expressões faciais. As imagens também estão em escala de cinza com 256 níveis. O tamanho das imagens é 320 x 243 pixels. Na Figura 9 são apresentados exemplos de faces do banco Yale.

Figura 9 - Amostras do banco de faces Yale.

Fonte: (GEORGHIADES, 1997).

2.4 Banco de Faces GTAV

O banco de faces GTAV foi desenvolvido pelo *Audio Visual Technologies Group* da *Technical University of Catalonia*. Ele contém 1188 imagens, distribuídas entre 27 pessoas com 48 poses, com variações de iluminação e posição e também em escala de cinza com 256 níveis. O tamanho das imagens é 320 x 240 pixels. Na Figura 10, são apresentados exemplos de faces do banco GTAV.

Figura 10 - Amostras do banco de faces GTAV.

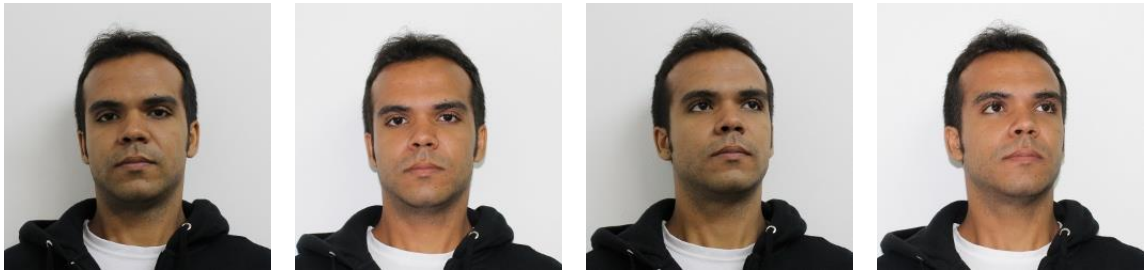
Fonte: (TARRÉS e RAMA, 2011).

2.5 Banco de Faces VSOFT

O vídeo coletivo produzido por (OMAIA, 2009) apresenta os seguintes problemas: movimento brusco da câmera, imagens borradas e *background* não uniforme. Diante deste fato, e da impossibilidade de reunir os sujeitos do banco UFPB para gravar novos vídeos coletivos, resolveu-se desenvolver um banco de faces próprio, intitulado Vsoft-Faces, para que pudessem ser gerados vídeos coletivos em ambientes mais controlados e assim fazer uma melhor avaliação da nova versão do SDRF.

O banco Vsoft-Faces é composto por duas partes: Vsoft-Fotos e Vsoft-Vídeos. O banco Vsoft-Fotos possui imagens de 14 sujeitos em 6 poses, sob duas fontes de iluminação e contra um fundo branco uniforme. As fotos originais possuem 1500 x 1500 pixels. Na Figura 11, são apresentadas algumas imagens que fazem parte das fotos originais do banco.

Figura 11 - Amostras do banco de faces Vsoft-Fotos.



O banco Vsoft-Fotos foi testado no sistema com uma variação do banco onde as imagens possuem dimensões de 250 x 250 pixels com 256 níveis de cinza e com o mínimo de *background* possível. Estas imagens foram obtidas a partir de um recorte das imagens originais. Na Figura 12 são apresentados exemplos das imagens recortadas.

Figura 12 - Exemplos de imagens recortadas.

O banco Vsoft-Vídeos possui dois vídeos coletivos, filmados com 30 quadros por segundo em uma câmera HD. Na Figura 13, é apresentado um quadro de um dos vídeos coletivos desse banco.

Figura 13 - Um quadro de vídeo coletivo do banco Vsoft

2.6 Considerações Finais do Capítulo

Neste Capítulo foram apresentados todos os bancos de faces necessários ao entendimento deste trabalho e dos trabalhos correlatos. Os bancos apresentados são: ORL, UFPB, Yale, GTAV e Vsoft.

Os bancos ORL e UFPB foram apresentados por serem utilizados em alguns dos trabalhos correlatos. Os bancos Yale e GTAV foram apresentados porque foram escolhidos para serem testados no presente trabalho devido ao fato de eles apresentarem variação de iluminação. O banco Vsoft foi apresentado por ter sido desenvolvido para o presente trabalho. Ele leva esse nome porque foi desenvolvido na empresa Vsoft Tecnologia.

No Capítulo 3 serão apresentados alguns trabalhos correlatos.

3. TRABALHOS CORRELATOS

Neste capítulo tanto são descritos trabalhos relacionados a localização de olhos em imagens de faces quanto trabalhos que utilizam a DCT para efetuar reconhecimento facial.

3.1 O método de Ke e Kang

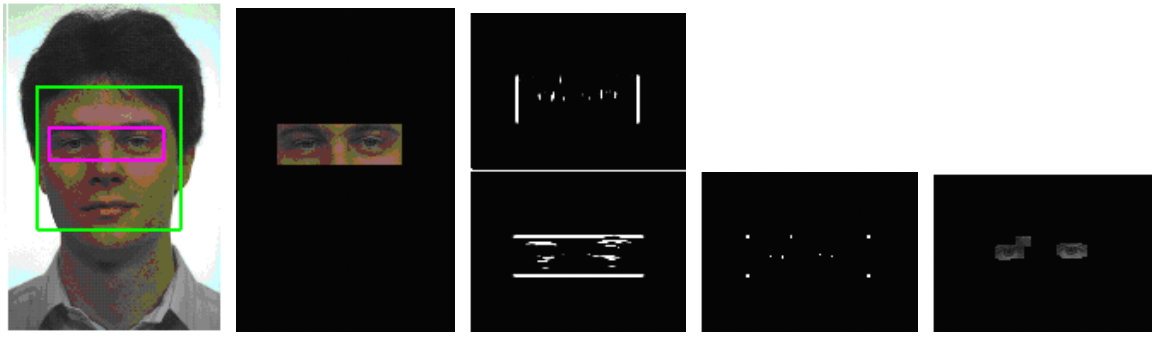
Neste trabalho foi desenvolvido um novo método para localizar os olhos em imagens de face; ele funciona em duas etapas: a localização da face e a posterior localização dos olhos na face detectada (KE e KANG, 2010).

Primeiramente, o método usa um detector baseado nos atributos de Haar (VIOLA e JONES, 2004) para detectar as faces presentes numa imagem e a região de olhos (região que engloba o par de olhos) presente em cada face detectada.

Em seguida, ele utiliza filtragem pelo gradiente de Gabor (combinado nas direções vertical e horizontal) (KE e HUANG, 2009) na região de olhos para delimitar regiões candidatas a serem olhos dentro da região de olhos previamente detectada. Depois, ele utiliza filtros morfológicos para restringir ainda mais as regiões candidatas.

Por fim, é efetuado *template matching* entre as regiões candidatas e templates de olhos representando os olhos direito e esquerdo. Na Figura 14 são apresentados os passos do método.

Figura 14 - Passos do método.



(a) Detecção da face e da região de olhos. | **(b) Extração da região de olhos.** | **(c) Delimitação das regiões candidatas.** | **(d) Após os filtros morfológicos.** | **(e) Região de olhos para *template matching*.**

Fonte: Adaptado de (KE e KANG, 2010).

O método foi testado com 1741 imagens de tamanho 256 x 384 pixels escolhidas aleatoriamente do banco de faces FERET (PHILLIPS *et al.*, 1997), divididas

em um grupo de 264 imagens (grupo 1) e outro de 1477 (grupo 2), e também foi testado em imagens próprias de vídeo de tamanho 320 x 240 pixels. No grupo 1 foi atingido 96,6% de acertos e no grupo 2, 87,34%.

Os resultados apontam para robustez em relação à pose, iluminação e expressão facial, porém o ângulo de tirar a foto e as formas dos olhos influenciam no processo de localização.

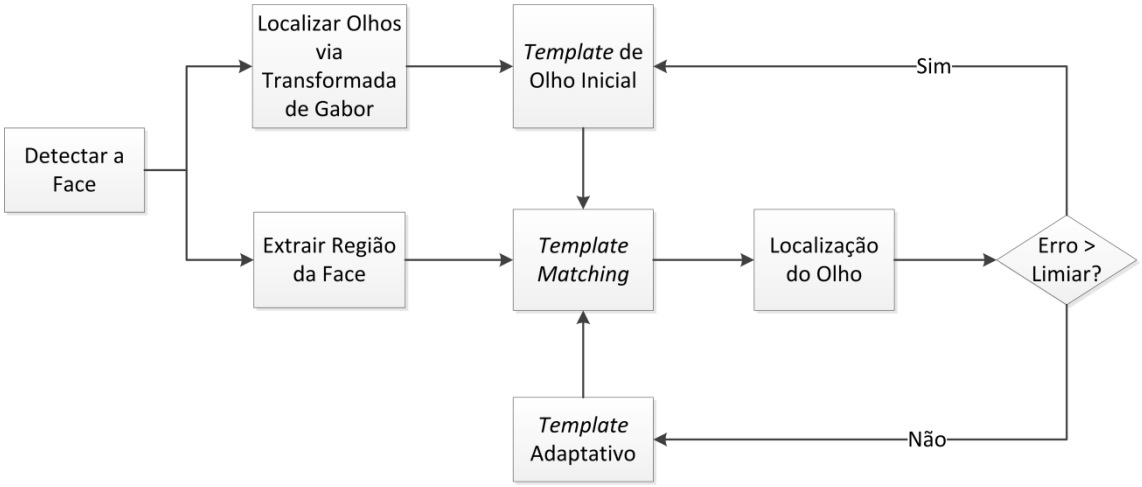
3.2 O método de Jing

Neste trabalho, é proposto um algoritmo para localizar os olhos utilizando uma webcam (JING *et al.*, 2010).

Primeiramente, a face é detectada também com um detector baseado nos atributos de Haar (VIOLA e JONES, 2004). Depois, a região dos olhos é estimada baseada na proporção da face. Em seguida, os olhos são localizados utilizando a transformada de Gabor, como em (LI *et al.*, 2006). Depois, a região da pupila (centro do olho) é utilizada como primeiro template.

Para efetuar o *template matching*, o *template* inicial percorre a região de olhos do novo frame utilizando uma medida de similaridade, pixel a pixel. O ponto onde houver a maior similaridade é considerado o olho e a região da pupila desse ponto é considerada o novo template. Na Figura 15, é apresentada a arquitetura do método.

Figura 15 - Arquitetura do método.



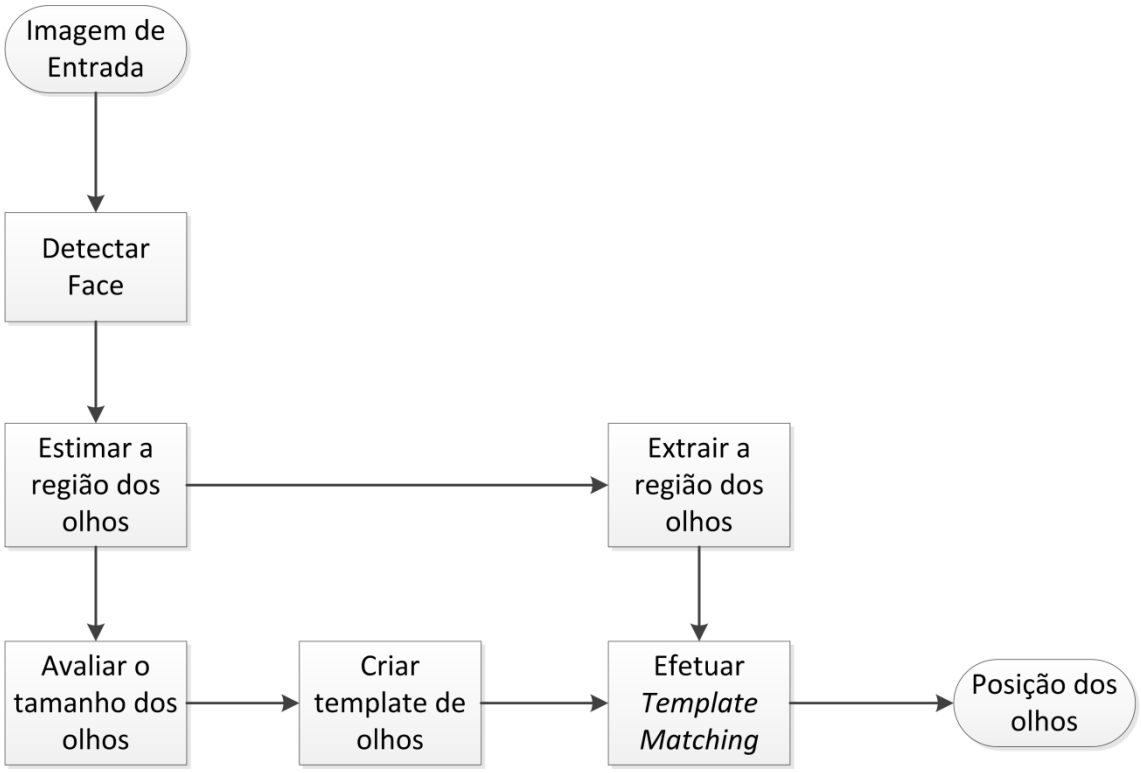
Fonte: Adaptado de (JING *et al.*, 2010).

O sistema foi testado com uma webcam Darkhorse com vídeo de resolução 640 x 480 pixels e 30 quadros por segundo, com variação de iluminação e background, mostrando-se eficaz mesmo quando há mudança de posição das faces.

3.3 O método *Rapid Human-Eye Detection Based on an Integrated Method*

O método RHEDBIM (do inglês *Rapid Human-Eye Detection Based on an Integrated Method*) é um método para detecção rápida de olhos em faces humanas utilizando correlação (XIA *et al.*, 2010). Este método é indicado para pessoas sem óculos e uma adaptação dele é usada no módulo de normalização proposto no presente trabalho. Na Figura 16, é apresentada a arquitetura do método.

Figura 16 - Arquitetura do método de (XIA *et al.*, 2010).



Fonte: Adaptado de (XIA *et al.*, 2010).

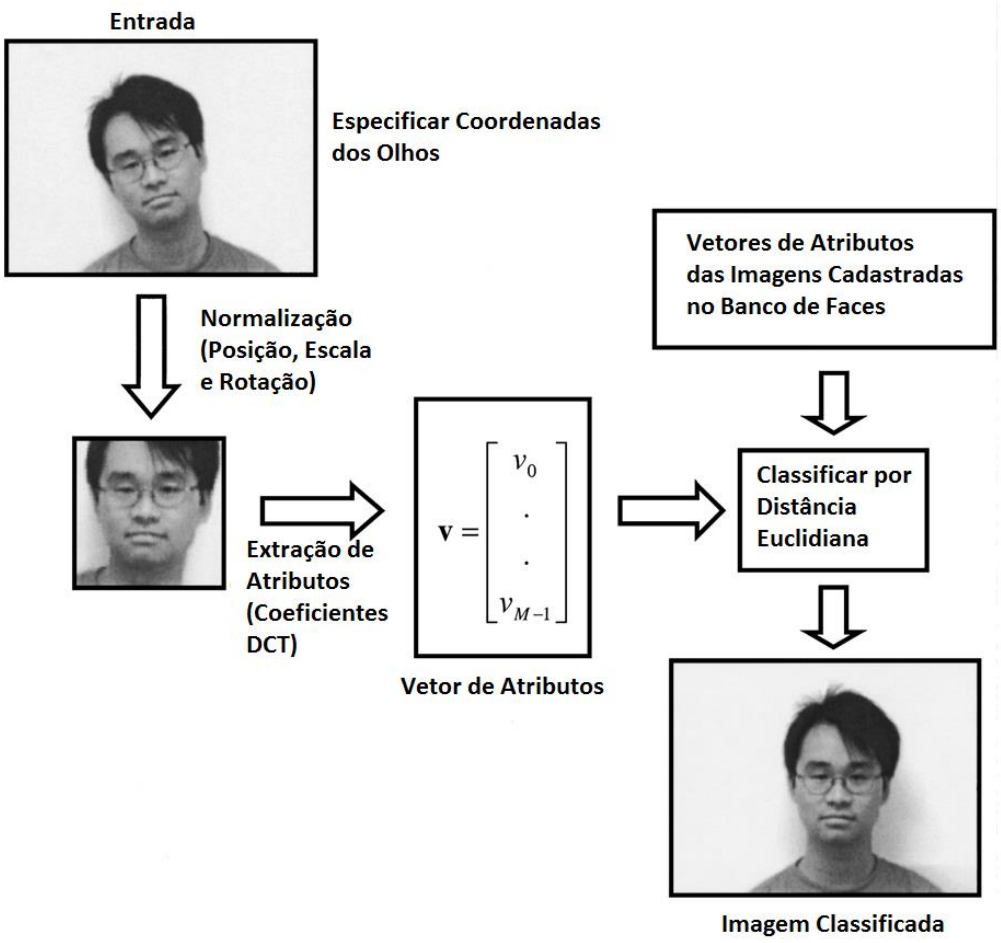
Neste método, a face é detectada usando uma variação do algoritmo Adaboost (XIA e SU, 2008). Então, na imagem de face, a região dos olhos é estimada utilizando um filtro de gradiente e posteriores projeções vertical e horizontal da imagem de gradiente. O tamanho dos olhos é obtido a partir da região de olhos previamente estimada. Então, imagens de templates de olhos são geradas a partir da imagem média de recortes manuais de olhos do banco de faces com que se está trabalhando para que se obtenha a posição dos centros dos olhos nas faces detectadas utilizando correlação entre as faces e os templates.

O método foi testado em 227 imagens de 29 indivíduos do banco de faces ORL, com todos os indivíduos sem óculos, alcançando uma taxa de 94,7% de acertos.

3.4 O Sistema de Hafed e Levine

Neste trabalho, os autores desenvolveram um sistema de reconhecimento de faces holístico que utiliza a Transformada Cosseno Discreta (HAFED e LEVINE, 2001). Para realizar a comparação das imagens, as DCT das imagens de faces são computadas e delas são selecionados os coeficientes de baixa e média frequência. Estes coeficientes formam um vetor de atributos e são comparados utilizando-se um classificador de vizinho mais próximo com distância euclidiana como métrica de comparação. A imagem de face de consulta é associada à imagem de face de treinamento cuja distância entre elas seja mínima. Na Figura 17, é apresentada a arquitetura do sistema.

Figura 17 - Arquitetura do Sistema de Hafed e Levine.



Fonte: Adaptado de (HAFED e LEVINE, 2001).

Para funcionar, o sistema necessita das coordenadas dos olhos das faces, porém, ao invés de ele calcular essas coordenadas automaticamente, ele usa algoritmos de terceiros. As coordenadas dos olhos são utilizadas para extrair a região da face da imagem.

A face é normalizada em relação à iluminação e posição, com o objetivo de diminuir as variações existentes entre as imagens. Da imagem normalizada são selecionados os 49 menores coeficientes (em uma região quadrática) DCT para compor o vetor de atributos.

Para testar o reconhecimento de faces em fotos foi utilizado o procedimento da validação cruzada *two fold*, o qual consiste em separar os dados em dois conjuntos, que podem ou não ser do mesmo tamanho, um conjunto de treinamento e um conjunto de testes. Dentre outros bancos de faces, o sistema foi testado no banco ORL dividindo as faces de cada pessoa do banco em cinco faces de testes e cinco de treinamento, atingindo uma taxa de acertos de 92,5%.

3.5 O Trabalho de Matos

A autora desenvolve e compara diversos métodos de reconhecimento de faces que utilizam a DCT (MATOS, 2008). Sete diferentes seletores de atributos são combinados com três diferentes classificadores para descobrir qual método é o mais eficiente. Dentre esses métodos, o que apresentou melhor taxa de acertos durante o reconhecimento facial foi o que utiliza um seletor de baixas frequências de DCT combinado com um classificador de Vizinho Mais Próximo (NN), sendo uma abordagem semelhante ao trabalho de (HAFED e LEVINE, 2001). No seletor de baixas frequências, uma região quadrática da DCT, contendo os coeficientes mais significativos é escolhida e a classificação é utilizada considerando-se apenas os coeficientes presentes nessa região. Etapas prévias de pré-processamento e normalização das imagens em relação à iluminação e posição são dispensadas, pois espera-se que variações de tais aspectos em coeficientes DCT bem selecionados sejam pouco representadas.

Para fazer a classificação das faces, o sistema implementa um classificador de Vizinho Mais Próximo que utiliza como medida de similaridade a distância de Minkowski de ordem um. A menor distância é utilizada para classificar uma face. Para testar o reconhecimento de faces em fotos foi utilizado o procedimento da validação cruzada *leave-one-out*, um caso específico da validação cruzada *two-fold* em que o conjunto de testes possui apenas um objeto. Então, no caso do trabalho de (MATOS, 2008), o sistema mantêm uma face fora do conjunto de treinamento para realizar a

classificação, ou seja, usa 9 faces para o treinamento. Por exemplo, o banco de faces ORL possui dez poses para cada pessoa, então dez rodadas são realizadas. Na primeira rodada, a primeira pose de todas as pessoas é excluída do conjunto de treinamento, essas faces excluídas são utilizadas como faces de teste a serem classificadas. Na segunda rodada, a segunda pose é excluída do conjunto de treinamento e é utilizada como face de teste, e assim por diante.

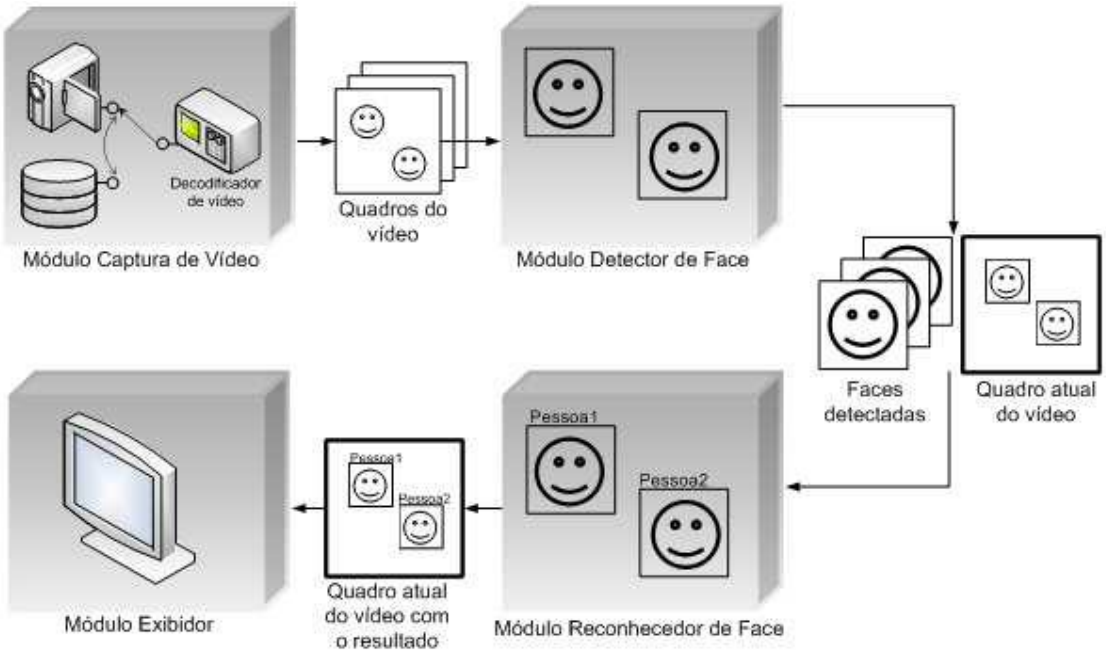
Este método se mostrou compatível com o estado da arte, atingindo uma taxa de acertos de 99,5% em 400 testes ao ser testado no banco de faces ORL, utilizando 36 coeficientes (*Olivetti Research Lab*) (AT&T LABORATORIES, 2002). Verifica-se uma taxa maior de acertos ao ser comparado com o trabalho de (HAFED e LEVINE, 2001), porém naquele trabalho são utilizadas 5 faces para o treinamento do classificador.

3.6 O Sistema para Detecção e Reconhecimento de Face em Foto e Vídeo Utilizando a Transformada Cosseno Discreta de Omaia

O anteriormente mencionado SDRF de (OMAIA, 2009) efetua reconhecimento facial holístico utilizando a DCT e um seletor de baixas frequências com uma região elíptica, sendo assim uma variação do melhor método apresentado em (MATOS, 2008).

O SDRF recebe como entrada um vídeo com faces para classificação, e tem como saída o vídeo com o resultado da detecção e do reconhecimento das faces presentes. Na Figura 1Figura 18, é apresentada a arquitetura geral do sistema.

Figura 18 - Arquitetura Geral do SDRF.



Fonte: (OMAIA, 2009).

A arquitetura é composta por quatro módulos: módulo de captura de vídeo, módulo detector de face, módulo reconhecedor de face, e módulo de exibição. O módulo de captura realiza a obtenção das imagens de entrada, que podem ser obtidas de um arquivo de vídeo previamente capturado, ou podem ser lidas diretamente de algum dispositivo de captura de vídeo. Então, o vídeo é decodificado e transformado em uma sequência de quadros, os quais são passados ao módulo de detecção de face.

O módulo de detecção de face detecta as faces presentes em todos os quadros de vídeo. O módulo reconhecedor de face analisa apenas as regiões que foram consideradas como face pelo detector de face. O reconhecimento de face é realizado sobre essas regiões. O módulo de exibição apenas exibe o resultado da detecção e reconhecimento sobre o vídeo que está sendo analisado. O SDRF não possui nenhum módulo de normalização, e assim, não realiza nenhum tipo de pré-processamento nas imagens antes de realizar o reconhecimento de face.

Conforme a Figura 6, o módulo reconhecedor de faces possui um submódulo seletor de atributos, que utiliza como atributos os coeficientes das imagens DCT extraídos de uma região elíptica como a região mostrada na Figura 19. O módulo reconhecedor também possui um submódulo para efetuar a classificação (reconhecimento) das imagens. Para isso, este submódulo também implementa um

classificador de Vizinho Mais Próximo que utiliza como métrica de comparação a distância de Minkowski de ordem um e no qual a menor distância é utilizada para classificar uma face.

Figura 19 - Exemplo de região elíptica.

Fonte: (Omaia, 2009).

Um banco de faces e vídeo (UFPB-Fotos e UFPB-Video) com pessoas da Universidade Federal da Paraíba - UFPB foi criado para testar o sistema. Para poder comparar o desempenho com o de outros métodos propostos na literatura, o SDRF também foi testado com o banco de faces ORL.

Para testar o reconhecimento de faces em fotos também foi utilizado o procedimento da validação cruzada *leave-one-out*. No banco UFPB-Fotos o reconhecedor atingiu uma taxa de 97,75% de acertos e no banco ORL 99,75% de acertos. Nos testes em vídeos, os resultados não foram promissores quando aplicados a um vídeo contendo múltiplas pessoas do banco de faces, resultando numa taxa inferior a 5% de acertos. A baixa taxa de acerto foi atribuída ao movimento dos indivíduos e da câmera, bem como à iluminação irregular do ambiente.

O autor propõe como trabalhos futuros o teste do SDRF nos mesmos bancos, mas normalizados em relação à posição e escala. Também propõe uma modificação na etapa de seleção de coeficientes, utilizando para isso um filtro passa-baixas não ideal. Estas propostas são objeto de pesquisa da presente dissertação.

3.7 Considerações Finais do Capítulo

Neste Capítulo, foram apresentados alguns trabalhos relacionados com o trabalho desenvolvido nessa dissertação. Os trabalhos apresentados dividem-se em dois grupos, sendo um composto por trabalhos que envolvem localização de olhos e

outro composto por trabalhos que envolvem reconhecimento facial. Os trabalhos de (XIA *et al.*, 2010) e de (OMAIA, 2009) receberam um maior destaque, pois foram utilizados na presente dissertação. Na Tabela 1 é apresentado um quadro dos métodos de localização de olhos apresentados neste Capítulo.

Tabela 1: Quadro Comparativo dos Localizadores de Olhos.

Localização de Olhos		
Trabalho	**Banco de Testes**	**Eficiência**
(KE e KANG, 2010)	Subconjunto do FERET	83,47%
	Subconjunto do FERET (grupo 2)	96,6%
(JING *et al.*, 2010)	Vídeo próprio	Não informado
(XIA *et al.*, 2010)	Subconjunto do ORL	94,7%

Note-se que não é possível fazer uma comparação destes trabalhos em termos da taxa de acertos da localização de olhos porque eles não foram testados com os mesmos bancos e não usam o mesmo tipo de validação.

Na Tabela 2 é apresentado um quadro comparativo dos sistemas reconhecedores de face apresentados neste Capítulo.

Tabela 2: Quadro Comparativo dos Reconhecedores de Face.

Reconhecimento Facial		
Trabalho	**Banco de Testes**	**Eficiência**
(HAFED e LEVINE, 2001)	ORL	92,5%
(MATOS, 2008)	ORL	99,25%
(OMAIA, 2009)	ORL	99,75%
	UFPB	97,75%

Nota-se que os trabalhos seguintes ao de (HAFED e LEVINE, 2001) buscaram e conseguiram um melhoramento deste.

No Capítulo 4 será detalhado o desenvolvimento do trabalho desta dissertação.

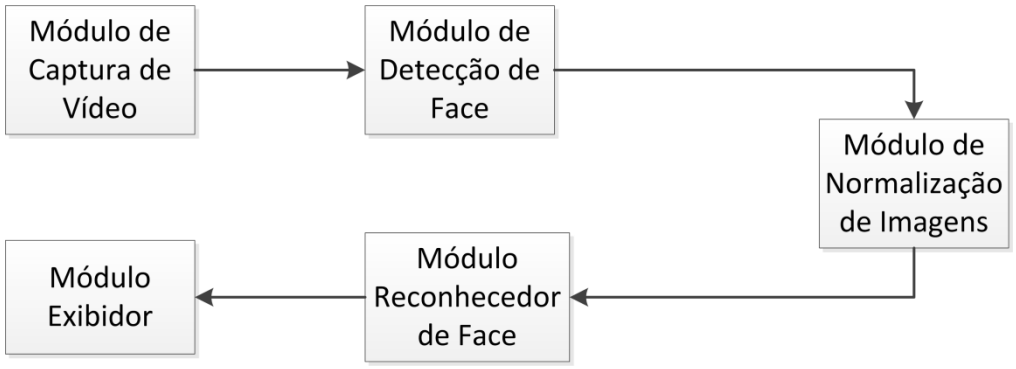
4. DESENVOLVIMENTO

Neste capítulo, é descrito todo o desenvolvimento das duas melhorias propostas para o SDRF, a saber: um módulo de normalização de imagens em relação à rotação e escala, e a uma modificação na etapa de seleção de atributos do SDRF, por meio da inserção de um filtro passa-baixas não ideal. A Seção 4.1 foca na etapa de normalização, enquanto que a Seção 4.2 na modificação da seleção de atributos.

4.1 Módulo de Normalização da Imagem

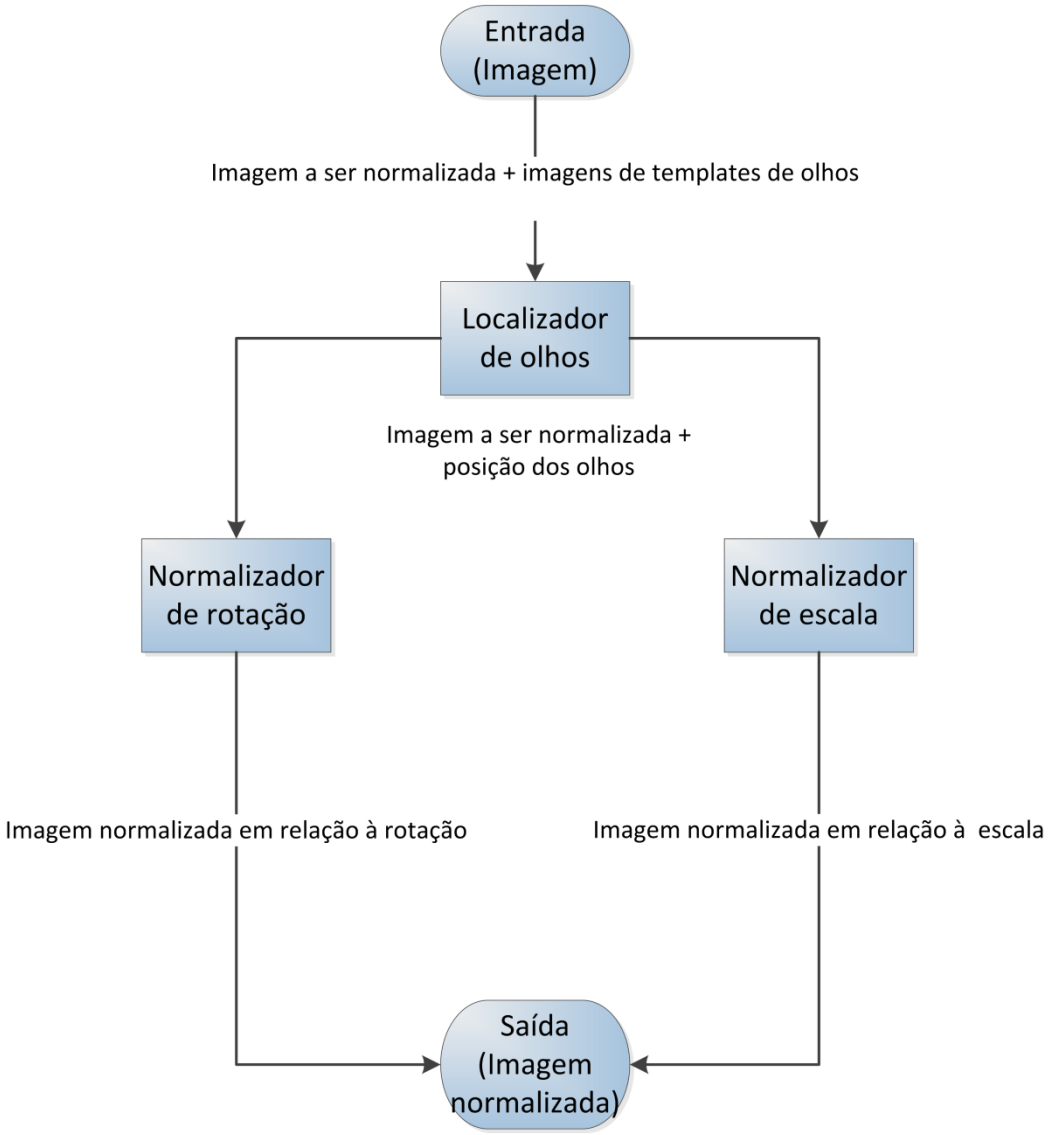
Na Figura 18 é apresentada a arquitetura do SDRF. Nela é possível notar que o sistema não possui um módulo de normalização de imagens conforme a arquitetura geral de um sistema de reconhecimento facial mostrada na Figura 6. Sem esse módulo não é possível adequar as imagens capturadas às mesmas condições das imagens utilizadas na etapa de treinamento do módulo reconhecedor de face. Assim, na Figura 20 é apresentada a posição do módulo normalizador na arquitetura do SDRF enquanto que na Figura 21 é apresentada a visão geral do módulo normalizador de imagens. Ele é dividido em dois submódulos: o localizador de olhos e o normalizador em si.

Figura 20 - Arquitetura do SDRF com o módulo de normalização de imagens.



O módulo normalizador recebe como entrada a imagem de face a ser normalizada e imagens de *template* de olhos (um *template* de olho direito e outro de olho esquerdo). Então, o sistema localiza os olhos na imagem de entrada por meio de uma operação de *template matching* que utiliza correlação como métrica de comparação. De posse da posição dos olhos, as faces ou são normalizadas em relação à rotação ou são normalizadas em relação à escala. Maiores detalhes sobre os tipos de normalização escolhida são apresentados nas subseções 4.1.2 e 4.1.3.

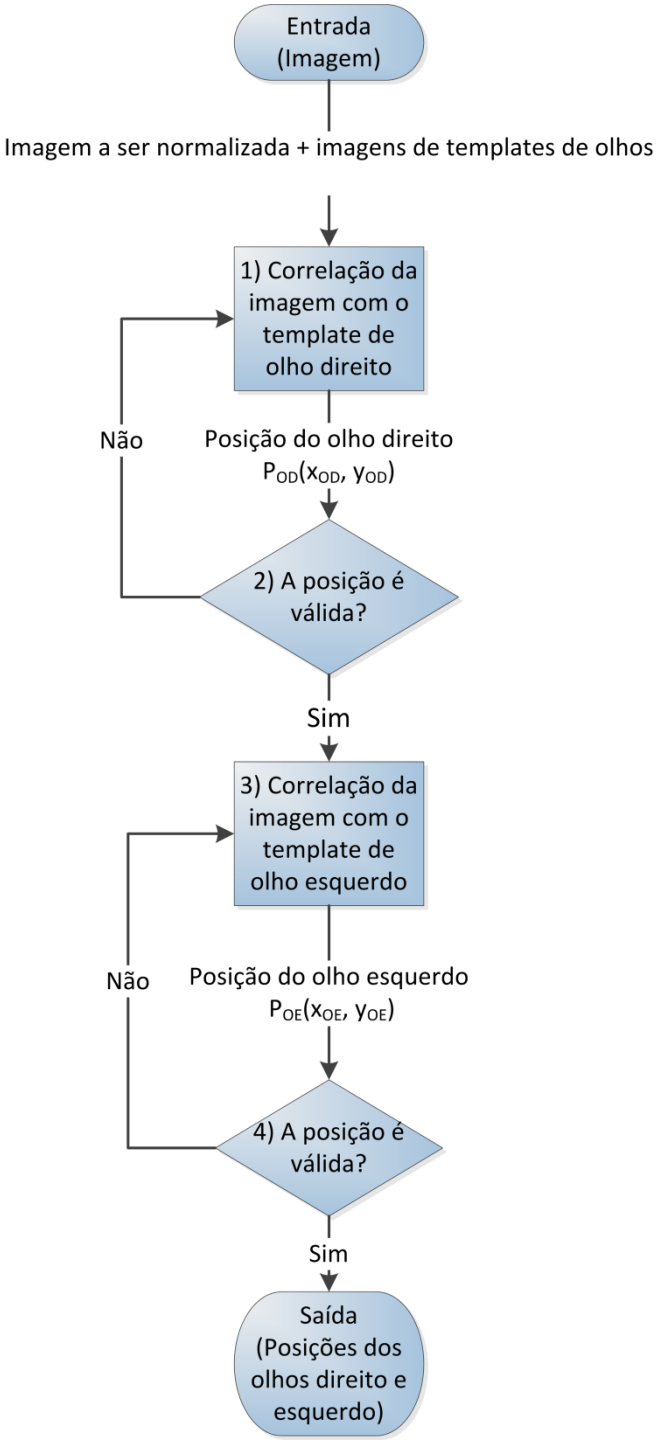
Figura 21 - Visão geral do normalizador de imagens.



4.1.1 Submódulo localizador de olhos

Na Figura 22 é apresentado o fluxo do submódulo localizador de olhos, uma implementação parcial do método proposto em (XIA *et al.*, 2010). No presente trabalho, a etapa de estimação de olhos foi removida após testes preliminares. Ao invés disso, os olhos são localizados por meio de *template matching* em toda a imagem, sendo a metade direita da face para o olho direito e a metade esquerda para o olho esquerdo. Esta abordagem foi escolhida pelo seu bom desempenho.

Figura 22 - Fluxo do Localizador de Olhos.



O localizador recebe como entrada a imagem de face e os *templates* de olhos. Os *templates* são gerados especificamente para cada banco, a partir da média de 30 recortes de olhos direito e olhos esquerdo de pessoas presentes no banco. Eles são produzidos de modo que o centro do olho esteja no centro do *template*, pois os *templates* funcionam como uma janela deslizante no *template matching* O **passo 1** consiste em efetuar *template matching* entre a metade direita da face com o *template*

de olho direito. Como dito antes, a correlação é utilizada como métrica de comparação e é feita de acordo com as Equações 1.5 e 1.6.

Como a correlação é uma operação pontual, em relação ao centro da janela deslizante, o ponto que produzir o maior valor de correlação durante o *template matching* é considerado como sendo a localização do centro do olho direito, $P_{OD}(x_{od}, y_{od})$.

O **passo 2** usa a Relação 3.1 para verificar se a posição do centro do olho direito P_{OD} é uma posição válida:

$$1/3 * altura_{Face} < y_{od} < 1/2 * altura_{Face} \quad \text{Relação (3.1)}$$

Sendo:

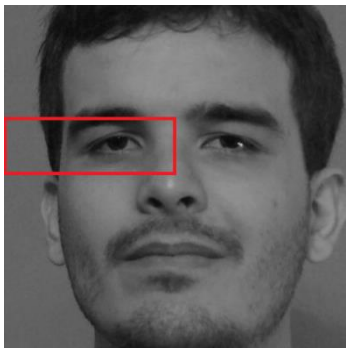
$altura_{Face}$ a altura da imagem da face;

y_{od} a ordenada do centro do olho direito e,

A origem do eixo de coordenadas da face $(0, 0)$ é o canto superior esquerdo.

Esta Relação é oriunda das proporções da face humana (DRAW23, 2012). Na Figura 23 é mostrado um exemplo de região onde o olho direito pode ser encontrado numa imagem de face, de acordo com a Relação (3.1).

Figura 23 - Região válida para localizar o olho direito.



Sendo P_{OD} uma posição válida, o localizador efetua *template matching* entre a metade esquerda da imagem da face e o *template* de olho esquerdo (**passo 3**). O ponto de maior valor da correlação, $P_{OE}(x_{oe}, y_{oe})$, é considerado como a localização do centro do olho esquerdo.

O **passo 4** utiliza as seguintes Relações para verificar se P_{OE} é uma posição de centro de olho esquerdo válida:

$$largura_{template} < distancia\_horizontal\_entre\ olhos < 4 \times largura_{template} \quad \text{Relação (3.2)}$$

Sendo:

$largura_{template}$ a largura da imagem de *template* de olho (esquerdo ou direito, já que ambos possuem o mesmo tamanho);

$distancia\_horizontal\_entre\_olhos$ a distância horizontal existente entre o olho esquerdo localizado e o olho direito previamente localizado, isto é: (x_{oe} - x_{od});

$$distancia\_vertical\_entre\ olhos < 3 \times altura_{template} \quad \text{Relação (3.3)}$$

Sendo:

$distancia\_vertical\_entre\_olhos$ a distância vertical existente entre o olho esquerdo localizado e o olho direito previamente localizado, isto é: (y_{oe} - y_{od}) ;

$$1/3 \times altura_{\ face} < y_{oe} \quad \text{Relação (3.4)}$$

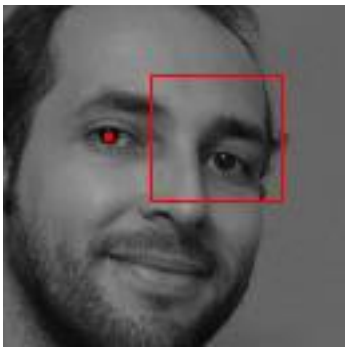
Sendo:

$altura_{\ Face}$ a altura da imagem de face; e

y_{oe} é a ordenada do centro do olho esquerdo.

Estas Relações também são devido às proporções da face humana. Na Figura 24 é apresentado um exemplo de região válida para a localização do olho esquerdo, uma vez que um olho direito já tenha sido detectado.

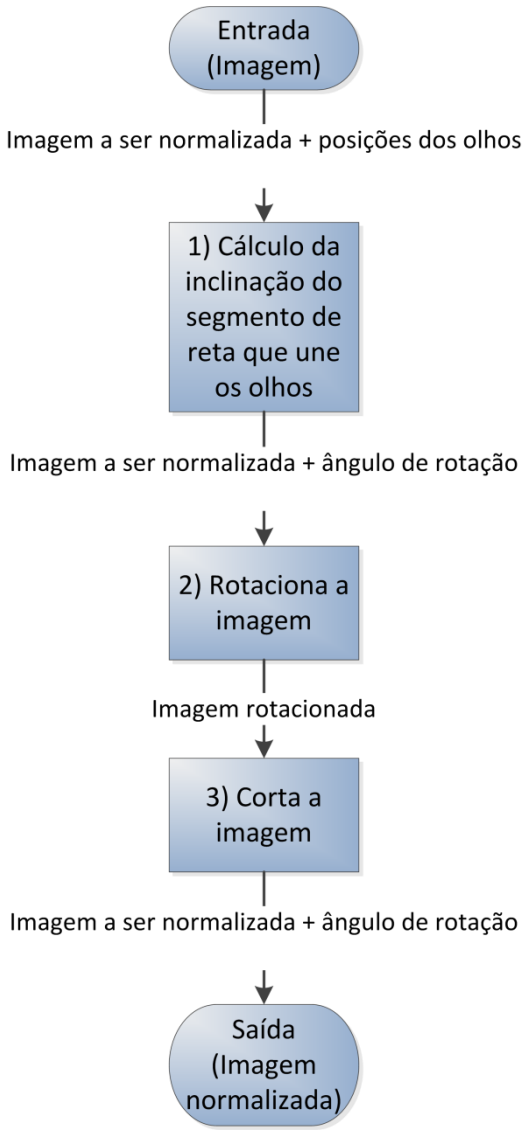
Figura 24 - Região da imagem onde o olho esquerdo encontrado será válido.



4.1.2 Submódulo normalizador de rotação

O submódulo normalizador de rotação tem como objetivo eliminar, se existir, a inclinação do segmento de reta que une os olhos de uma imagem de face. Uma vez que as coordenadas dos olhos são obtidas, calcula-se tal inclinação para então aplicar uma operação de rotação e assim normalizar as imagens. Este tipo de normalização foi escolhido porque geralmente as pessoas não são filmadas com a cabeça inclinada para terem suas faces reconhecidas. Na Figura 25 é apresentado o fluxo deste submódulo.

Figura 25 - Fluxo do submódulo normalizador de rotação.

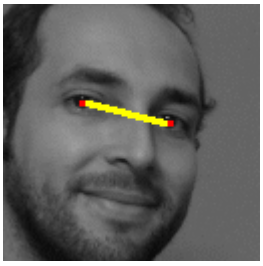


O **passo 1** consiste em calcular a inclinação do segmento de reta que une os dois olhos. Isto é feito calculando-se o arco tangente das posições dos olhos (Figura 26 (a)).

No **passo 2**, a imagem é rotacionada, de acordo com a Equação 1.1, de modo a eliminar a inclinação calculada no passo 1 (Figura 26 (b)).

A operação de rotação pode deixar a imagem com espaços em branco no plano de fundo. Assim, no **passo 3** uma janela de tamanho 100 x 100 pixels é utilizada para cortar a imagem, de modo que não existam espaços em branco nela (Figura 26 (c)).

Figura 26 - Passos do submódulo normalizador de posição.



(a) Localização dos olhos.

(b) Aplicação de rotação na imagem.

(c) Aplicação de recorte na imagem.

4.1.3 Submódulo normalizador de escala

O submódulo normalizador de escala tem como objetivo deixar todas as imagens de um determinado banco de faces com a mesma distância entre os olhos, fazendo assim com que todas elas fiquem com a mesma distância para a câmera. Esta normalização foi escolhida porque os erros do SDRF no teste em fotos eram causados por diferenças de escala de acordo com (OMAIA, 2009). Na Figura 27 é apresentado o fluxo deste submódulo.

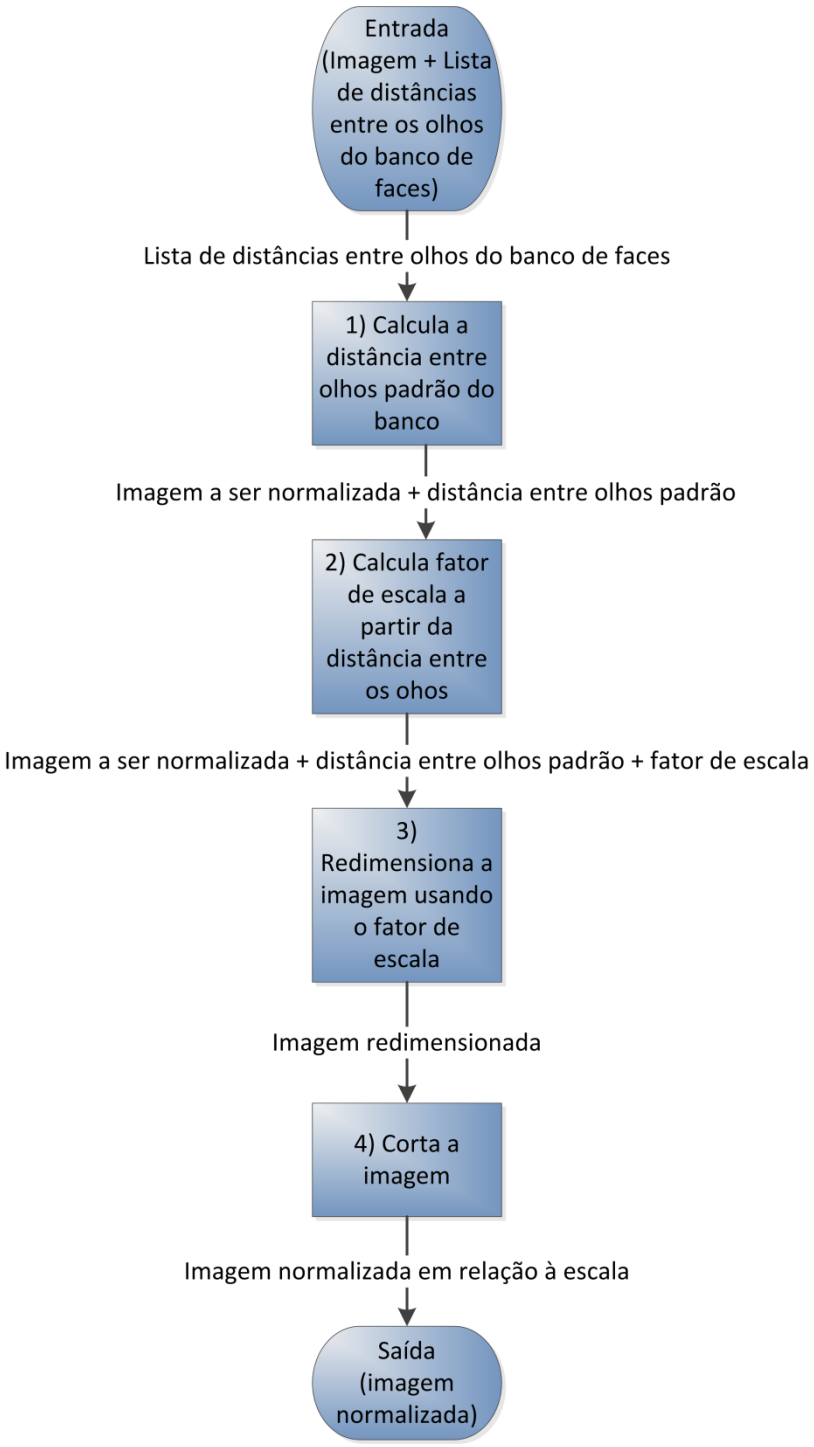
O **passo 1** pode ser feito à parte. Ele consiste em calcular a distância entre olhos padrão $D_{padrão}$ de um dado banco de faces, a partir da moda de todas as D_k distâncias entre olhos desse banco.

No **passo 2**, calcula-se para a imagem a ser normalizada um fator de escala, de acordo com a Equação 3.1 (Figura 28 (a)).

$$fatorDeEscala = D_{padrão} / D_k \tag{3.1}$$

No **passo 3**, a imagem é redimensionada utilizando-se o fator de escala definido na Equação (3.1) e a operação de redimensionamento definida na Equação 1.2 (Figura 28 (b)).

Figura 27 - Fluxo do submódulo normalizador de escala.



A operação de redimensionamento deixa as imagens do banco de faces com tamanhos diferentes, mas o SDRF necessita que elas sejam do mesmo tamanho para o teste do reconhecimento em fotos. Assim, é necessário o **passo 4**. Ele recorta as imagens para um mesmo tamanho (Figura 28(c)).

Figura 28 - Passos do submódulo normalizador de escala.

(a) Imagem original ($D_k = 50$) **(b) Imagem redimensionada ($D_k = 43$)** **(c) Imagem recortada ($D_k = 43$)**

4.2 Modificação na Etapa de Seleção de Atributos do SDRF

No SDRF, a seleção de atributos é feita utilizando um seletor de baixas frequências com uma região elíptica. Isto funciona como um filtro passa-baixas ideal (PEYNOT, 2011), pois os coeficientes da DCT que não são utilizados no processo de reconhecimento têm os seus valores transformados em zero.

Um filtro passa-baixas ideal elimina completamente as altas frequências fora da área de raio D_0, a partir da origem da imagem, enquanto mantém (ou seja, deixa passar) as frequências mais baixas dentro da área de raio de D_0. A sua função de transferência $H(u,v)$ é definida na Equação 3.2.

$$H(u,v) = \begin{cases} 1, & se\ D(u,v) \leq D_0 \\ 0, & se\ D(u,v)\ \geq D_0 \end{cases}. \tag{3.2}$$

Sendo:

$H(u,v)$ a função de transferência do filtro;

$D_0 >= 0$; e

$D(u,v)$ a distância de um ponto (u,v) para a origem da DCT, conforme a Equação 3.3.

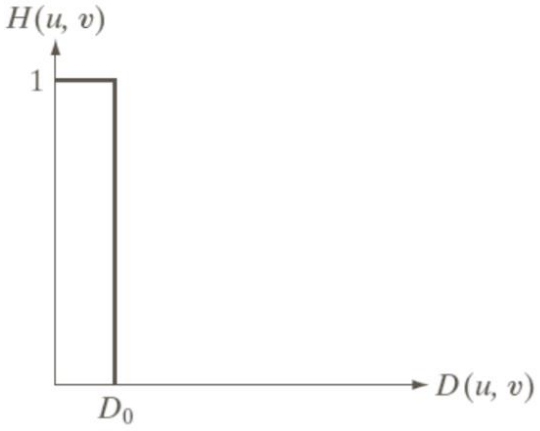
$$D(u,v) = \sqrt{(u^2 + v^2)}. \tag{3.3}$$

Ou seja, para gerar uma nova imagem $G(u,v)$ no domínio da frequência a partir de uma imagem de entrada $F(u,v)$ também no domínio da frequência e de um filtro passa-baixas ideal definido pelo sua função de transferência $H(u,v)$ basta aplicar a Equação (3.4).

$$G(u,v) = F(u,v)H(u,v). \tag{3.4}$$

A resposta em frequência é uma função retangular, como mostrado na Figura 29. O ponto de transição entre $H(u,v) = 1$ e $H(u,v) = 0$ é chamado de frequência de corte.

Figura 29 - Resposta em frequência de um filtro passa-baixas ideal.



Fonte: (PEYNOT, 2011).

O trabalho de (OMAIA, 2009) deixa um questionamento: alterar a etapa de seleção de atributos inserindo um filtro passa baixas não ideal de modo que um número maior de coeficientes seja utilizado no processo de reconhecimento facial, mesmo que isso aumente o custo computacional, faria a taxa de acertos do SDRF aumentar?. Um exemplo desse tipo de filtro é o filtro de Butterwoth (PEYNOT, 2011), cuja função de transferência é dada na Equação 3.5 e a resposta em frequência na Figura 30. É possível ver na Figura 30 como a variação de n afeta o decaimento dos valores na função de transferência, ou seja, quanto maior o valor de n, mais abrupto será o decaimento.

$$H(u,v) = \frac{1}{1 + [D(u,v)/D_0]^{2n}}. \tag{3.5}$$

Sendo:

$H(u,v)$ a função de transferência do filtro.

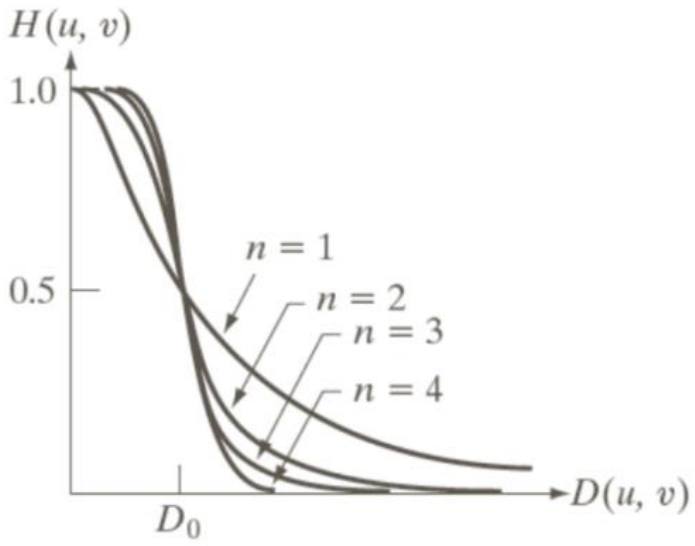
n a ordem.

$D_0 >= 0$.

$D(u,v)$ a distância de um ponto (u,v) para a origem da DCT, conforme a Equação 3.3.

Assim, a modificação na etapa de seleção de atributos do SDRF consiste em selecionar todos os coeficientes da imagem DCT e aplicar sobre eles o filtro de Butterwoth. A comparação é feita da mesma forma que no SDRF original. Dentre os filtros passa-baixas não ideais, o filtro de Butterwoth foi escolhido por ter uma resposta em frequência plana (Figura 30) e ser um dos mais utilizados na literatura.

Figura 30 - Resposta em frequência do filtro de Butterwoth.



Fonte: (PEYNOT, 2011).

4.3 Considerações Finais do Capítulo

Neste capítulo foi apresentado o desenvolvimento das duas melhorias para o SDRF que foram propostas nos objetivos específicos deste trabalho. A primeira consiste em um módulo normalizador de imagens em relação à rotação e escala, e a segunda consiste em uma alteração na etapa de seleção de atributos do SDRF.

Desenvolver um módulo normalizador para o SDRF foi uma escolha natural porque um módulo desse tipo é citado como parte integrante dos sistemas de reconhecimento facial e porque é sabido que normalizar imagens em sistemas de classificação, de modo que elas fiquem nas mesmas condições que as imagens de treinamento, leva à melhores resultados.

Alterar a etapa de seleção de atributos foi uma sugestão retirada da seção de trabalhos futuros de (OMAIA, 2009) que visa investigar se haverá um ganho na porcentagem de acertos do sistema ao utilizar um número maior de coeficientes DCT no processo de reconhecimento.

No Capítulo 5 serão apresentados os resultados dessas modificações.

5. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste Capítulo, são descritos os resultados alcançados pelas duas melhorias. Na Seção 5.1, são apresentados os resultados do módulo normalizador de rotação e escala. Na Seção 5.2, são apresentados os resultados da modificação na etapa de seleção de atributos do SDRF no reconhecimento sobre fotos (Seção 5.2.1) e sobre vídeos (Seção 5.2.2).

5.1 Resultados do Módulo Normalizador de Imagens

Nesta Seção são descritos os resultados obtidos com o módulo normalizador de imagens. Ele foi testado em fotos nos bancos UFPB (Seção 5.1.1) e ORL (Seção 5.1.2).

5.1.1 Testes no Banco UFPB

Dois pares principais de *templates* foram gerados para o banco de faces UFPB-Fotos, o primeiro é o Par #1, mostrado nas Figuras 31 (a) e (b).

Figura 31 - Par #1.

(a) Olho Direito (b) Olho Esquerdo

O segundo é o Par #2, mostrado nas Figuras 32 (a) e (b). Ambos os pares possuem 16 x 8 pixels.

Figura 32 - Par #2.

(a) Olho Direito (b) Olho Esquerdo

Para escolher o par de *templates* que apresente a melhor taxa de acertos na localização de olhos para ser usado na normalização das imagens, os pares #1 e #2 foram combinados em oito testes, cujos resultados são mostrados na Tabela 3.

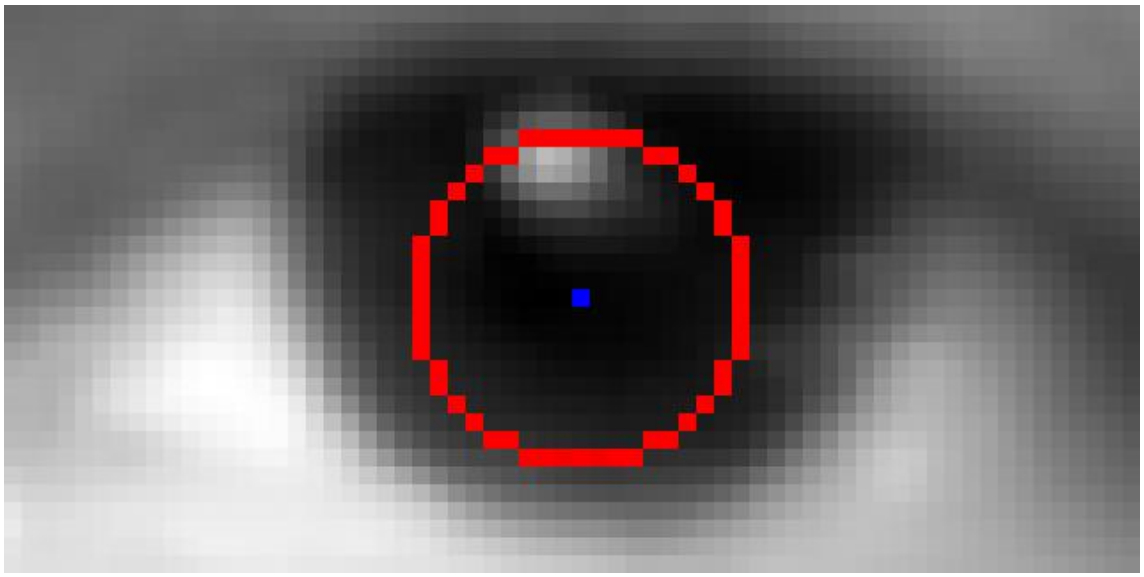
Tabela 3: Testes de localização de olhos no banco UFPB.

Teste	Template de Olho Direito	Template de Olho Esquerdo	% de Acertos
1	Direito_1	Esquerdo_1	84,00%
2	Direito_1	Esquerdo_2	83,50%
3	Direito_2	Esquerdo_1	83,12%
4	Direito_2	Esquerdo_2	83,38%
5	Direito_1	Direito_1 (invertido)	84,50%
6	Esquerdo_1 (invertido)	Esquerdo_1	79,75%
7	Direito_2	Direito_2 (invertido)	83,12%
8	Esquerdo_2 (invertido)	Esquerdo_2	83,12%

Em cada um destes testes, uma lista contendo as posições de olhos de cada imagem do banco é gerada automaticamente utilizando o submódulo localizador de olhos (Seção 4.1.1). Esta lista é comparada com uma lista de posições corretas do banco (gabarito), marcadas manualmente. A comparação consiste em calcular a diferença das posições encontradas automaticamente e as posições gabaritadas dos olhos de cada imagem. Se essa diferença, para ambos os olhos, estiver dentro de uma margem de erro de quatro pixels para mais ou para menos, então um acerto é computado. Essa margem foi escolhida empiricamente após análise visual, pois qualquer pixel que esteja na região de raio 4 pixels definida a partir do centro da íris ainda está na região da íris. Este fato está ilustrado na Figura 33, que mostra um *template* de olho ampliado em 4 vezes e na Figura 34, que mostra esse mesmo template ampliado em 10 vezes de tal forma que podemos contar os pixels.

Figura 33 - Template de Olho (Ampliado 4x).

Figura 34 - Template de Olho (Ampliado 10x).



Uma vez que a localização de olhos desenvolvida exige que o sujeito esteja com os olhos abertos e olhando para a câmera, resolveu-se montar uma nova versão do banco UFPB-Fotos em que as imagens de pessoas com os olhos fechados ou com a cabeça muito inclinada, doravante chamadas de imagens inadequadas, foram retiradas do banco com o objetivo de melhorar o desempenho do localizador de olhos. Na Figura 35, são apresentados exemplos das imagens retiradas. Foram retiradas 260 imagens, o que corresponde a 32,5% do banco UFPB-Fotos, restando 540 imagens.

Figura 35 - Exemplo de imagens inadequadas.

Os oito testes da Tabela 3 foram repetidos nessa versão do banco e a Tabela 4 mostra os resultados.

Tabela 4: Testes de localização de olhos no banco UFPB sem imagens inadequadas.

Teste	Template de Olho Direito	Template de Olho Esquerdo	% de Acertos
9	Direito_1	Esquerdo_1	87,59%
10	Direito_1	Esquerdo_2	87,59%
11	Direito_2	Esquerdo_1	86,48%
12	Direito_2	Esquerdo_2	87,04%
13	Direito_1	Direito_1 (invertido)	87,04%
14	Esquerdo_1 (invertido)	Esquerdo_1	84,26%
15	Direito_2	Direito_2 (invertido)	87,04%
16	Esquerdo_2 (invertido)	Esquerdo_2	87,04%

Comparando os dados da Tabela 4 e da Tabela 3, nota-se um ganho de 3,09 pontos percentuais no melhor caso (Teste 5 com Testes 9 e 10). Diante deste fato, foi decidido analisar, em números absolutos, a quantidade de erros e acertos nas 540 imagens consideradas adequadas e nas 260 imagens consideradas inadequadas. Estes números são mostrados na Tabela 5.

A partir da Tabela 5, notou-se que havia muitos acertos na localização de olhos dento do subconjunto de imagens consideradas inadequadas. Foi concluído então que as imagens não foram separadas em imagens adequadas e inadequadas da melhor maneira e esse teste foi abandonado. Apenas o banco UFPB-Fotos em sua totalidade foi escolhido para ser normalizado.

Tabela 5: Acertos e Erros nas imagens inadequadas.

	Quantidade de Acertos	Quantidade de Erros	Total
Imagens adequadas	470	70	540
Imagens inadequadas	162	98	360

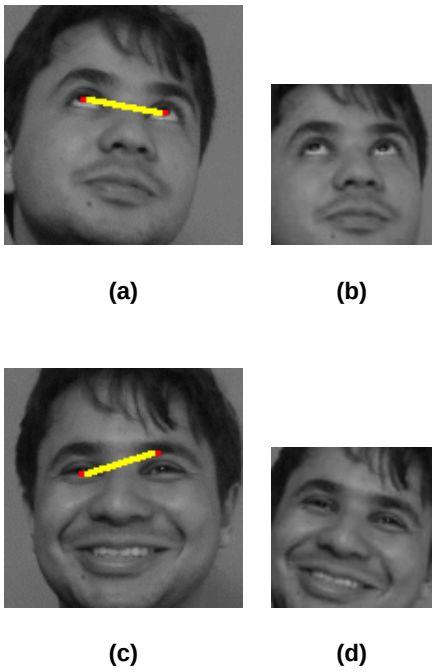
Para localizar os olhos no UFPB-Fotos e depois proceder à normalização, foi escolhido o par de *templates* (Direito_1, Direito_1 (invertido)), por ter obtido o melhor resultado na localização automática de olhos, de acordo com o Teste 5 da Tabela 3.

Com a lista de posições de olhos gerada automaticamente após o *template matching* de todas as imagens do banco com o par (Direito_1, Direito_1 (invertido)), o próximo passo foi normalizar, separadamente, as imagens em relação à rotação (Seção 4.1.2) e escala (Seção 4.1.3).

Primeiramente, o banco UFPB foi normalizado em relação à rotação, seguindo os passos da Seção 4.1.2 Na Figura 36 são apresentados exemplos de imagens do banco usadas nesta normalização. As Figuras 36 (a) e 36 (b) mostram exemplos de

normalização a partir da localização correta dos olhos e as Figuras 36 (c) e 36 (d) mostram exemplos de normalização a partir de uma localização incorreta.

Figura 36 - Imagens utilizadas nesta normalização.



(a) **(b)**

(c) **(d)**

Em seguida o banco UFPB foi normalizado em relação à escala, seguindo os passos da Seção 4.1.3. Nas Figuras 37 (a) e 37 (b), são apresentados exemplos de imagens do banco usadas nesta normalização.

O SDRF foi executado, então, sobre os bancos UFPB original, UFPB normalizado em relação à posição e UFPB normalizado em relação à escala. O SDRF atingiu taxa de 97,25% de acertos no banco UFPB original, 96% (10 erros a mais) no normalizado em relação à posição e 97,12% (1 erro a mais) em relação à escala.

A redução na taxa de acertos ao testar o banco normalizado em relação à posição é atribuída aos erros inseridos pela localização dos olhos (por exemplo, Figura 366 (c)). Entenda-se por erros aquelas imagens nas quais os olhos foram localizados incorretamente e assim as imagens também foram rotacionadas incorretamente.

A leve redução no banco normalizado em relação à escala é relacionada à perda de informação ao redimensionar as imagens. Também é importante observar que as imagens reconhecidas erroneamente não são as mesmas nos três testes.

Figura 37 - Exemplos de imagens utilizadas na normalização.

(a) **(b)**

Como o localizador de olhos inseriu erros onde antes não havia. As taxas atingidas pelos bancos normalizados não são úteis para fins de comparação com o banco UFPB-Fotos não normalizado testado em (OMAIA, 2009). Além disso, o banco UFPB-Fotos testado em (OMAIA, 2009), além de possuir mais informação que sua versão normalizada no presente trabalho, como mostra a Figura 38, mantém a região acima da testaFigura 38 - Diferença de informação entre imagem original e normalizada.. Essa região contém, muitas vezes, parte do cabelo, que pode variar muito de uma captura para outra da mesma pessoa, ou seja, contem informação que não faz parte da face em si. Para contornar estes fatos e testar a utilidade de um módulo normalizador no SDRF, foi feito o seguinte:

- Foi gerada uma versão do banco UFPB-Fotos de tamanho 100x100, a partir do mesmo corte 100x100 utilizado no normalizador;
- O banco UFPB-Fotos (original) foi normalizado em relação à rotação e em relação à escala a partir da lista de posições de olhos corretas, eliminando assim os erros da detecção automática de olhos.

Figura 38 - Diferença de informação entre imagem original e normalizada.

(a) Imagem original **(b) Imagem normalizada.**

O SDRF foi executado então sobre estes três bancos, tendo como resultado 96,62% de acertos no banco UFPB-Fotos 100 x 100, 97% no banco UFPB-Fotos normalizado em relação à rotação e 97,12% no banco UFPB-Fotos normalizado em

relação à escala. Em ambos os casos a taxa de acertos para os bancos normalizados subiu, comprovando a necessidade de um módulo normalizador para o SDRF.

5.1.2 Testes no Banco ORL

Para o banco ORL, foi gerado um par de *templates* de olho, também com 16 x 8 pixels: o Par #3, ilustrado nas Figuras 39Figura 39 (a) e (b).

Figura 39 – Par #3.

(a) Olho Direito (b) Olho Esquerdo

Este par e suas variantes invertidas horizontalmente foram combinados em 4 testes de localização de olhos (Tabela 6).

Tabela 6: Testes de localização de olhos no banco ORL.

Teste	Template de Olho Direito	Template de Olho Esquerdo	% de acertos
17	Direito_3	Esquerdo_3	59,00%
18	Direito_3	Direito_3 (invertido)	58,75%
19	Esquerdo_3 (invertido)	Esquerdo_3	53,75%
20	Esquerdo_3 (invertido)	Direito_3 (invertido)	53,75%

Observando os dados da Tabela 6, decidiu-se normalizar o banco ORL com o par de *templates* (Direito_3, Esquerdo_3), que é o Par #3 original, por ter apresentando o melhor resultado na localização automática para este banco.

Com a lista de posições de olhos gerada automaticamente após a correlação de todas as imagens do banco com o par de *template* escolhido, o próximo passo foi normalizar, separadamente, as imagens em relação à posição e escala.

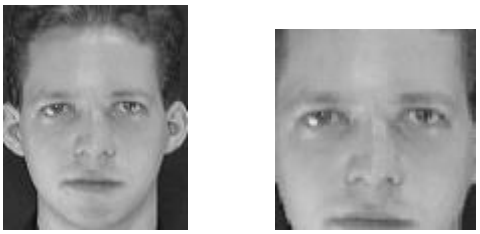
Primeiramente, o banco ORL foi normalizado em relação à posição, seguindo os passos da Seção 4.1.2. Em seguida o banco foi normalizado em relação à escala, seguindo os passos da Seção 4.1.3.

O SDRF foi executado, então, sobre os bancos ORL original, ORL normalizado em relação à posição e ORL normalizado em relação à escala. O SDRF atingiu taxa de 99% de acertos (4 erros) no banco original, 86,5% de acertos (50 erros a mais) no normalizado em relação à posição e 94% de acertos (20 erros a mais) em relação à escala. Novamente, a redução na taxa de acertos ao testar o banco normalizado em relação à posição é atribuída aos erros inseridos pela localização dos olhos, enquanto

que a redução na taxa de acertos ao testar o banco normalizado em relação à escala é atribuída à perda da informação ao redimensionar as imagens.

Além dos erros inseridos pela localização automáticas de olhos nas versões normalizadas do banco ORL, o banco ORL original testado em (OMAIA, 2009) também contém mais informação que as versões normalizadas no presente trabalho. Este fato é ilustrado Figura 40.

Figura 40 - Diferenças entre ORL original e normalizado.



(a) Imagem original (b) Imagem normalizada.

Assim, para checar a necessidade do módulo normalizador, as imagens do banco ORL original também foram modificadas de modo a conter uma área de face semelhante a das áreas das faces dos bancos ORL normalizado; e, o banco ORL original foi normalizado em relação à rotação e em relação à escala utilizando a lista de posições corretas de olhos. Essas versões do banco ORL foram então testadas no SDRF, obtendo como resultados 92,5% de acertos para o banco ORL cortado, 92,5% de acertos para o banco ORL normalizado em relação à rotação e 94% de acertos para o banco ORL normalizado em relação à escala. Nota-se que a taxa de acertos para o banco normalizado em relação à rotação não foi alterada, enquanto que a taxa de acertos para o banco ORL normalizado em relação à escala subiu 1,5 ponto percentual. Novamente, isto confirma a necessidade de um módulo normalizador para o SDRF.

5.2 Resultados da Modificação na Etapa de Seleção de Atributos do SDRF

Nesta Seção, são descritos os resultados da modificação na etapa de seleção de atributos do SDRF. Na Seção 5.2.1, são descritos os resultados sobre fotos, testando todos os bancos, enquanto que na Seção 5.2.2 são descritos os resultados sobre vídeo, testando o banco UFPB-Vídeo e Vsoft-Vídeos.

5.2.1 Resultados no Reconhecimento sobre Fotos

Para testar a modificação proposta na Seção 4.1 em fotos foram utilizados todos os bancos descritos no Capítulo 2.

Os bancos UFPB e ORL foram testados com as seguintes variantes: original, normalizado em relação à rotação, normalizado em relação à escala, modificado pelo filtro de Log, modificado pelo filtro de Log About e modificado pelo filtro de Histograma. Os filtros de normalização de iluminação são utilizados para avaliar o comportamento da etapa de seleção de atributos alterada quando a iluminação do banco de faces é modificada artificialmente.

O banco Yale foi testado primeiramente com as seguintes variantes: original, recortada com 128 x 128 pixels de dimensões, recortada com 100 x 100 pixels de dimensões. Estes cortes foram feitos no banco para avaliar o comportamento do SDRF (original e modificado) na presença e ausência do *background* das imagens dessa banco. Posteriormente, a versão 128 x 128 foi escolhida para gerar mais três variantes para o teste: modificado pelo filtro de Log, modificado pelo filtro de Log About e modificado pelo filtro de Histograma. A razão de testar esses filtros é a mesma dos bancos ORL e UFPB, com a diferença de que a iluminação original do banco Yale é não uniforme.

O banco GTAV foi testado em sua versão original e na variante recortada de 128 x 128 pixels de dimensões. Este banco também foi recortado para avaliar a influência do *background* de suas imagens no resultado do SDRF original e modificado.

Por fim, o banco Vsoft-Fotos foi testado em sua variante original.

Na Tabela 7, são apresentados os resultados de reconhecimento em fotos antes e depois da modificação no sistema.

Para o Banco UFPB, pode-se observar uma leve queda na taxa de acertos, em todas as variações testadas

Para o banco ORL, a taxa de acertos mantém-se estável na versão original e na versão modificada pelo filtro de Log, cai 0,25 pontos percentuais no banco normalizado em relação à rotação, cai 1 ponto percentual na versão modificada pelo filtro de LogAbout e sobe 0,25 pontos percentuais no banco normalizado em relação à escala. Este comportamento não permite conclusões.

Tabela 7: Resultados do SDRF com e sem a modificação na etapa de seleção de atributos

Banco de Faces	Variação	% de acertos no SDRF original	% de acertos no SDRF modificado	Diferença no % de acertos
UFPB	128 x 128	97,25%	96,75%	-0,5
	*100 x 100 (Normalizado em relação à rotação)	96%	95,62%	-0,38
	*100 x 100 (Normalizado em relação à escala)	97,12%	96,37%	-0,75
	128 x 128 (Modificado com Log)	93,62%	93,25	-0,37
	128 x 128 (Modificado com Log About)	93,62%	93,12	-0,5
	128 x 128 (Modificado com Histograma)	96%	95,87	-0,13
ORL	92 x 112	99%	99%	0
	*100 x 100 (Normalizado em relação à rotação)	86,5%	86,25%	-0,25
	*100 x100 (Normalizado em relação à escala)	94%	94,25%	0,25
	100 x100 (Modificado com Log)	96,50%	96,50%	0
	100 x100 (Modificado com Log About)	97,50%	96,50%	-1
	100 x100 (Modificado com Histograma)	97,75%	96,25%	-1,5
Yale	320 x 243	90,90%	92,12%	2,12
	128 x 128 (cortado)	87,88%	89,70%	1,82
	100 x 100 (cortado)	83,64%	84,24%	0,6
	128 x 128 (Log)	93,93%	93,93%	0
	128 x 128 (Log About)	95,75%	94,54%	-1,21
	128 x 128 (Histograma)	90,30%	89,69%	-0,61
GTAV	320 x 240	92,51%	92,51%	0
	128 x 128 (cortado)	80,40%	80,68%	0,28
Vsoft-Fotos	250 x250 (cortado)	97,62%	98,81%	1,19

O banco Yale apresenta resultados significativos, uma vez que nas três primeiras variantes do banco (original, recortado com 128 x 128 pixels e recortado com 100 x 100 pixels) a taxa de acertos do reconhecimento aumenta em relação ao SRDF original, diferente do banco UFPB e do banco ORL.

A característica mais peculiar ao banco Yale em relação aos bancos ORL e UFPB é a iluminação não uniforme. Para averiguar se a iluminação não uniforme é uma das causas do melhor desempenho do SDRF modificado no banco Yale, foram criadas as variantes do banco modificadas pelos filtros de Log, Log About e Equalização de Histograma. Ao testar essas variantes no SDRF original e modificado nota-se um pior desempenho do SDRF modificado. Testar essas variantes do Yale foi a razão de criação das variantes de iluminação dos bancos UFPB e ORL, de modo a comparar o desempenho destes no SDRF modificado de uma forma melhor.

O banco GTAV apresenta uma pequena melhora (0,28 pontos percentuais) quando suas imagens são cortadas de modo a conterem apenas partes da face, sem *background*, porém a versão original do banco mantém a taxa de acertos.

Por fim, para o banco Vsoft-Fotos é observada uma melhora de 1,19 pontos percentuais na taxa de acertos do reconhecimento.

5.2.2 Resultados no Reconhecimento sobre Vídeos

Para validar o SDRF, em (OMAIA, 2009) foram realizados testes em vídeos individuais e também em um vídeo coletivo, ou seja, vídeo com mais de uma pessoa presente. O teste em vídeo coletivo apresentou uma taxa de acertos muito baixa no reconhecimento: cerca de 5% de acertos analisando o vídeo quadro a quadro. Assim, resolveu-se avaliar vídeos coletivos em busca de um melhor resultado. Para tanto, foram realizados os seguintes experimentos:

I. Execução do SDRF com a etapa de seleção de atributos modificada sobre o mesmo vídeo coletivo usado em (OMAIA, 2009).
II. Execução do SDRF (original e modificado) sobre os vídeos coletivos criados no banco Vsoft-Vídeos, tendo assim um meio de comparar resultados em vídeos de melhor qualidade.

Resultados sobre o banco UFPB-Vídeo

Neste teste, o vídeo testado é o vídeo coletivo do banco UFPB. Duas variantes do banco UFPB-Fotos são utilizadas no treinamento: a original (região inteira da face) e uma de 80 x 54 pixels, que representa uma região parcial da face que engloba olhos e nariz (Figura 41).

Figura 41 - Exemplo de imagem de região parcial da face.

Na Tabela 8, são comparados os resultados do reconhecimento em vídeo do SDRF original com o SDRF modificado com o filtro de Butterwoth.

Tabela 8: Comparação dos resultados no banco UFPB-Vídeo

Região usada no reconhecimento	% acertos no SDRF original	% acertos no SDRF modificado	Diferença no % de acertos
Parcial	5,17%	4,76%	-0,41
Inteira	7,12%	6,09%	-1,03

O vídeo é analisado quadro a quadro nos testes. O tamanho mínimo da face a ser detectada é de 120 pixels e o número total de faces detectadas foi 6.517 faces. É importante salientar que os parâmetros n e D_0 do filtro de Butterwoth são diferentes para cada tipo de região de face. Ao testar o reconhecimento utilizando a região inteira, tem-se $n = 6$ e $D_0 = 10$, enquanto que utilizando a região parcial, temos $n = 5$ e $D_0 = 10$. Estes valores foram escolhidos empiricamente após testar as combinações de n e D_0 que produziam as melhores taxas de reconhecimento. Uma vez que o banco UFPB-Fotos é utilizado na etapa de treinamento, a queda na taxa de acertos do reconhecimento desse vídeo é condizente com a queda na taxa de acertos ao testar o banco UFPB-Fotos no SDRF modificado.

Resultados sobre o banco Vsoft-Vídeo

Na Tabela 9, são comparados os resultados do reconhecimento em vídeo do SDRF original com o SDRF modificado com o filtro de Butterwoth. Foram testados os dois vídeos coletivos do banco Vsoft-Vídeos. Uma variante do banco Vsoft-Fotos-2 com dimensões de 120 x 120 pixels foi utilizada no treinamento.

Tabela 9: Comparação dos resultados no banco Vsoft-Vídeos

Vídeo	% acertos no SDRF original	% acertos no SDRF modificado	Diferença no % de acertos
Vídeo 1	23,32%	62,20%	38,88
Vídeo 2	6,54%	8,88%	2,34

Novamente, os vídeos foram analisados quadro a quadro. O tamanho mínimo da face a ser detectada novamente foi de 120 pixels e o número total de faces detectadas foi 373 faces. Os valores dos parâmetros do filtro de Butterwoth foram $n = 5$ e $D_0 = 55$, escolhidos novamente após testes para a verificação dos melhores valores.

Em todos os testes foram observadas taxas de acerto superiores às apresentadas no vídeo do banco UFPB. Isto era esperado, uma vez que os vídeos do banco Vsoft-Vídeo foram filmados com uma melhor iluminação e com os indivíduos olhando diretamente para a câmera, sem muita variação de posição.

5.3 Considerações Finais do Capítulo

Neste capítulo foram apresentados os resultados das modificações desenvolvidas de acordo com o Capítulo 4.

A primeira modificação feita no SDRF foi o desenvolvimento de um módulo normalizador de imagens. Este módulo pode normalizar as imagens em relação à rotação ou à escala. A normalização em relação à rotação é feita de modo a retirar a inclinação existente entre os olhos, pois assim as faces ficam nas mesmas condições de posição que as imagens de treinamento. A normalização em relação à escala é feita de modo a padronizar a distância das faces para a câmera, eliminando assim disparidades de zoom. Isto é feito, pois o único erro do SDRF em relação ao banco de faces ORL está relacionado a diferenças de zoom, conforme visto em (OMAIA, 2009).

Os bancos UFPB-Fotos e ORL foram normalizados em relação à rotação e escala utilizando detecção automática de olhos. Como a localização de olhos não se mostrou perfeita e inseriu erros onde antes não havia, esses bancos foram normalizados então utilizando suas respectivas listas de posições corretas de olhos. Além disso, os bancos UFPB-Fotos e ORL foram cortados para que se adequassem ao nível de informação dos bancos normalizados. Isto foi feito para que a necessidade de um módulo normalizador de imagens para o SDRF fosse devidamente averiguada. Na Tabela 10 é apresentado um resumo do desempenho, em termos de porcentagem de acertos, dos bancos UFPB e ORL cortados e normalizados no SDRF original.

Tabela 10: Resumo do desempenho dos bancos UFPB e ORL no SDRF.

Banco	Variação		
	Cortado (100 x 100)	Normalizado em relação à rotação	Normalizado em relação à escala
UFPB	96,62%	97%	97,12%
ORL	92,5%	92,5%	94%

Nota-se que a taxa de acertos nos bancos normalizados foi igual ou superior que os bancos sem normalização, o que confirma a utilidade do módulo normalizador para o SDRF. Além disso, confirma o averiguado em (OMAIA, 2009) no que diz respeito à diferença de distância para câmera no banco ORL.

A segunda modificação feita no SDRF consiste em uma alteração na etapa de seleção de atributos, visando investigar o desempenho do SDRF ao utilizar todos os coeficientes DCT. Esta modificação foi testada nos bancos UFPB-Fotos, GTAV, Yale, ORL e Vsoft. Na Tabela 7 são apresentados os resultados dos testes em fotos.

O que se pode concluir dos resultados da Tabela 7 é que a modificação via inserção do filtro de Butterwoth da etapa de seleção de atributos do SDRF pode melhorar o desempenho do reconhecimento em bancos de faces de iluminação não uniforme, porém o ganho na taxa é pouco e isto aumenta o custo computacional.

Quanto aos testes em vídeo, aplicou-se o SDRF modificado no vídeo coletivo realizado por (OMAIA, 2009) e o resultado foi um decremento de cerca de 1% na taxa de acertos, o que condiz com o decremento nas taxas de acerto do banco de faces UFPB-Fotos, que é utilizado na etapa de treinamento do teste de vídeo. O SDRF (original e modificado) também foi testado nos vídeos coletivos do banco Vsoft-Vídeos, com o intuito de realizar testes de vídeo coletivo em melhores condições. Em ambos os testes o SDRF obteve melhores resultados que quando testado com o vídeo coletivo do banco UFPB-Vídeo, comprovando a má qualidade deste. O SDRF modificado também foi testado no banco Vsoft-Vídeos, melhorando o desempenho em relação ao SDRF original. Isto é condizente com o melhor desempenho do SDRF modificado no banco Vsoft-Fotos, que é utilizado para o treinamento do reconhecimento facial do banco Vsoft-Vídeos.

No Capítulo 6 serão apresentadas as considerações finais sobre o presente trabalho bem como sobre trabalhos futuros.

6. CONSIDERAÇÕES FINAIS

O presente teve como objetivo geral adequar o SDRF a ambientes não controlados. Para tanto, foram definidos os seguintes objetivos específicos:

1. Desenvolver um módulo de normalização de imagens em relação à rotação e escala. Este módulo deverá retirar as inclinações das imagens de face, de modo que o ângulo entre os olhos seja de 0º. Ele também corrigirá as diferenças de escala entre as imagens, de modo que as imagens de face fiquem com a mesma distância em relação à câmera.
2. Modificar a etapa de seleção de atributos do SDRF por meio da inserção de um filtro passa baixas não ideal.

Em relação ao objetivo específico 1 (página 20), o módulo normalizador de imagens foi desenvolvido, conforme descrito no Capítulo 4, seção 4.1. Ele é considerado a maior contribuição do presente trabalho, pois de acordo com os resultados apresentados no Capítulo 5, seções 5.1 e 5.3, sua adição ao SDRF influencia positivamente no desempenho do reconhecimento facial.

Em relação ao objetivo específico 2 (página 20), alterar a etapa de seleção de atributos do SDRF não se mostrou uma contribuição proveitosa, pois apesar de gerar um melhor desempenho em alguns dos bancos testados, a alteração gera também um maior custo computacional. O SDRF modificado por essa alteração pode ser utilizado desde que o usuário o calibre para melhor se adequar ao banco de faces que irá ser reconhecido. Por ora, os resultados apresentados no Capítulo 5, seção 5.2 confirmam que é melhor fazer reconhecimento facial utilizando poucos coeficientes DCT.

Assim, o SDRF com o módulo de normalização de imagens não é um sistema voltado para ambientes não controlados, mas é um sistema com mais opções de ambientes para uso.

As modificações desenvolvidas neste trabalho não são imunes à erros que podem ser corrigidos em trabalhos futuros. Exemplos destes trabalhos que podem ser feitos são:

- a inclusão de um modelo de face que amplie a heurística do submódulo localizador de olhos;
- a geração de melhores *templates* de olhos para que seja feita uma localização automática de olhos mais adequada; e
  a comparação de outros filtros passa baixas não ideais para verificar qual apresentaria o melhor desempenho na modificação da etapa de seleção de atributos do SDRF.

REFERÊNCIAS

AHMAD, T.; JAMEEL, A.; AHMAD, B. **Pattern recognition using statistical and neural techniques**. 2011 International Conference on Computer Networks and Information Technology (ICCNIT). [S.l.]: [s.n.]. 2011. p. 87-91.

AT&T LABORATORIES. The ORL Database of Face, 2002. Disponivel em: <http://www.cl.cam.ac.uk/research/dtg/attarchive/pub/data/att_faces.zip>. Acesso em: Julho 2011.

BATISTA, L. V. **Notas de Aula da Disciplina Introdução ao Processamento Digital de Imagens**. João Pessoa. 2005.

BELHUMEUR, P. N.; HESPANHA, P.; KRIEGMAN,. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. **IEEE Trans. Pattern Analysis and Machine Intelligence**, p. 711-720, 1997.

BRUNELLI, R.; POGGIO, T. A. Face recognition: Features versus templates. **IEEE Trans. Pattern Anal. Mach. Intell.**, Washington, v. 15, n. 10, p. 1042-1052, Outubro 1993. ISSN 0162-8828.

CHAUDHARI, S. T.; KALE, A. **Face Normalization:** Enhancing Face Recognition. Third International Conference on Emerging Trends in Engineering & Technology (ICETET 2010). [S.l.]: IEEE Computer Society. 2010. p. 520-525.

DRAW23. How to Draw Face. **Draw23**, 2012. Disponivel em: <http://www.draw23.com/drawing-face>. Acesso em: 18 Março 2012.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2ª. ed. [S.l.]: Wiley-Interscience, 2001. ISBN 0471056693.

GEORGHIADES, A. S. **Yale Face Database**, 1997. Disponivel em: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. **Digital Image Processing Using MATLAB**. Upper Saddle River: Prentice-Hall, 2003. ISBN 0130085197.

HAFED, Z. M.; LEVINE, M. D. Face Recognition Using the Discrete Cosine Transform. **International Journal of Computer Vision**, n. 3, p. 167-188, 2001.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: a review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4-37, Janeiro 2000. ISSN 0162-8828.

JAIN, A. K.; KLARE, B.; PARK, U. **Face recognition:** Some challenges in forensics. 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011). [S.l.]: [s.n.]. 2011. p. 726-733.

JAIN, A. K.; LI, S. Z. **Handbook of Face Recognition**. 2ª. ed. Secaucus: Springer-Verlag New York, 2011. 1-3 p. ISBN 038740595X.

JING, Z.; ZHIXING, L.; LI, Z. **An adaptive template eye location based on Gabor transform method using web camera**. 2010 International Conference on Computer Design and Applications (ICCDA). [S.l.]: [s.n.]. 2010. p. 178-181.

KAKADIARIS, I. A.; PASSALIS, G.; TODERICI, G.; EFRATY, E.; PERAKIS, P.; CHU, D. Face Recognition Using 3D Images. In: LI, S. Z.; JAIN, A. K. **Handbook of Face Recognition**. 2ª. ed. [S.l.]: Springer, 2011. Cap. 17, p. 428-459.

KANADE, T. **Picture Processing System by Computer Complex and Recognition of Human Faces**. Tese de Doutorado, Kyoto University. Kyoto, Japão. 1973.

KE, L.; HUANG, Y. **Eyes Location Based on Dual-Orientation Gabor Filters and Templates**. CISP '09. 2nd International Congress on Image and Signal Processing, 2009. [S.l.]: [s.n.]. 2009. p. 1-4.

KE, L.; KANG, J. **Eye location method based on Haar features**. 3rd International Congress on Image and Signal Processing (CISP), 2010. [S.l.]: [s.n.]. 2010. p. 925-929.

KHAYAM, S. A. **The Discrete Cosine Transform (DCT): Theory and Application**. Michigan State University. [S.l.]. 2003.

LI, S. Z.; JAIN, A. K. Introduction. In: LI, S. Z.; JAIN, A. K. **Handbook of Face Recognition**. 2ª. ed. [S.l.]: Springer, 2011. Cap. 1, p. 1-18. ISBN 978-0-85729-931-4.

LI, S. Z.; YI, D. Face Recognition Using Near Infrared Images. In: LI, S. Z.; JAIN, A. K. **Handbook of Face Recognition**. 2ª. ed. [S.l.]: Springe, 2011. Cap. 15, p. 382-400.

LI, S.; LIU, D. H.; SHEN, L. S. Eye location using Gabor Transform. **Measurement & Control Technology**, v. 32, p. 27-29, Maio 2006.

LIU, H.; GAO, W.; MIAO, J.; LI, J. **A Novel Method to Compensate Variety of Illumination in Face Detection**. 6th Joint Conference on Information Sciences (ICCVPRIP). [S.l.]: [s.n.]. 2002. p. 692-695.

LORDÃO, F. A. F. **Reconhecimento de Formas Utilizando Modelos de Compressão de Dados e Espaço de Escalas de Curvatura**. Disserntação de Mestrado (Mestrado em Informática), Universidade Federal da Paraíba. João Pessoa. 2009.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. [S.l.]: Brasport, 1999. ISBN 8574520098.

MATOS, F. M. D. S. **Reconhecimento de Faces Utilizando a Transformada Cosseno Discreta**. Dissertação de Mestrado (Mestrado em Informática), Universidade Federal da Paraíba. João Pessoa. 2008.

OMAIA, D. **Um Sistema para Detecção e Reconhecimento de Face em Vídeo Utilizando a Transformada Cosseno Discreta**. Dissertação de Mestrado (Mestrado em Informática), Universidade Federal da Paraíba. João Pessoa. 2009.

PARKER, J. R. **Algorithms for Image Processing and Computer Vision**. [S.l.]: John Wiley & Sons, 1996.

PEYNOT, T. Chapter 4 - Filtering in the Frequency Domain. **Digital Image Processing**, Julho 2011. Disponivel em: <http://www.acfr.usyd.edu.au/courses/amme4710/Lectures/AMME4710-Chap4-FrequencyFiltering.pdf>. Acesso em: 28 Abril 2012.

PHILLIPS, J.; MOON, H.; RAUSS, P.; RIZVI, S. A. **The FERET Evaluation Methodology for Face-Recognition Algorithms**. Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR). Washington, DC, USA: IEEE Computer Society. 1997.

ROSS, A. A.; NANDAKUMAR, K.; JAIN, A. K. **Handbook of Multibiometrics**. 1ª. ed. Secaucus: Springer-Verlag New York, 2006. 1-3 p. ISBN 0387222960.

SELLAHEWA, H.; JASSIM, S. Image-Quality-Based Adaptive Face Recognition. **IEEE Transactions on Instrumentation and Measurement**, v. 59, n. 4, p. 805-813, 2010.

SINGH, K. R.; ZAVERI, M. A.; RAGHUWANSHI, M. M. Illumination and Pose Invariant Face Recognition: A Technical Review. **International Journal of Computer Information Systems and Industrial Management Applications**, v. 02, p. 29-38, 2010. ISSN 2150-7988.

SOLOMON, C.; BRECKON, T. **Fundamentals of Digital Image Processing**. [S.l.]: Wiley-Blackwell, 2011.

TARRÉS, F.; RAMA, A. GTAV Face Database, 2011. Disponivel em: <http://gpstsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm>. Acesso em: Setembro 2011.

THIAN, N. **Biometric Authentication System**. Dissertação de Mestrado, USM. Penang, Malásia. 2001.

TURK, M.; PENTLAND, A. Eigenfaces for face recognition. **J. Cognitive Neuroscience**, Cambridge, MA, USA, v. 3, n. 1, p. 71-86, Janeiro 1991. ISSN 0898-929X.

VIOLA, P.; JONES, M. Robust Real-Time Face Detection. **International Journal on Computer Vision**, Hingham, MA, USA, v. 57, n. 2, p. 137-154, 2004. ISSN 09205691.

VIOLA, P.; JONES, M. **Rapid object detection using a boosted cascade of simple features**. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.. [S.l.]: [s.n.]. 2001. p. 511-518.

WEBB, A. R. **Statistical Pattern Recognition**. 2ª. ed. [S.l.]: John Wiley and Sons, 2002.

WISKOTT, L.; FELLOUS, J.-M.; KRÜGER, ; VON DER MALSBURG,. Face Recognition By Elastic Bunch Graph Matching. **IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE**, v. 19, p. 775-779, 1997.

XIA, L.; SU, L. **Research on Face Detection Classifier Using an Improved Adboost Algorithm**. ISCSCT '08. International Symposium on Computer Science and Computational Technology, 2008. [S.l.]: [s.n.]. 2008. p. 78-81.

XIA, L.; YONGQING, D.; SU, L.; CHAO, H. **Rapid human-eye detection based on an integrated method**. Communications and Mobile Computing (CMC), 2010 International Conference on. [S.l.]: [s.n.]. 2010. p. 3-7.

ZHAO, W.; CHELLAPPA, R. Image-based Face Recognition: Issues and Methods. In: JAVIDI, B. **Image Recognition and Classification:** Algorithms, Systems and Applications. 1ª. ed. [S.l.]: CRC Press, 2002.

ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P. J.; ROSENFELD, A. P. Face recognition: A literature survey. **ACM Comput. Surv.**, New York, v. 35, n. 4, p. 399-458, Dezembro 2003. ISSN 0360-0300.