UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MATHEUS ANTÔNIO CORRÊA RIBEIRO

GERENCIAMENTO E AUTENTICAÇÃO DE IDENTIDADES DIGITAIS USANDO FEIÇÕES FACIAIS

Porto Alegre
(2008)

MATHEUS ANTÔNIO CORRÊA RIBEIRO

GERENCIAMENTO E AUTENTICAÇÃO DE IDENTIDADES DIGITAIS USANDO FEIÇÕES FACIAIS

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.

Área de concentração: Automação e Instrumentação Eletro-Eletrônica

Orientador: Jacob Scharcanski

Co-orientador: Adalberto Schuck Júnior

Porto Alegre

(2008)

MATHEUS ANTÔNIO CORRÊA RIBEIRO

GERENCIAMENTO E AUTENTICAÇÃO DE IDENTIDADES DIGITAIS USANDO FEIÇÕES FACIAIS

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: _____

Prof. Dr. Jacob Scharcanski, UFRGS

Doutor pela Universidade de Waterloo, Canadá

Banca Examinadora:

Profa. Dra. Letícia Vieira Guimarães, UFRGS

Doutora pelo Muroran Institute of Technology – Muroran, Japão

Prof. Dr. Walter Fetter Lages, UFRGS

Doutor pelo Instituto Tecnológico de Aeronáutica – São José dos Campos, Brasil

Prof. Dr. Cláudio Rosito Jung, UNISINOS

Doutor pela Universidade Federal do Rio Grande do Sul – Porto Alegre, Brasil

Coordenador do PPGEE: _____

Prof. Dr. Arturo Suman Bretas

Porto Alegre, maio de 2008.

DEDICATÓRIA

Este trabalho é dedicado à minha família e aos meus amigos, em especial pela dedicação, compreensão e apoio incondicionais.

AGRADECIMENTOS

Aos professores que, de alguma forma, me orientaram ao longo do mestrado.

Aos colegas do Departamento de Engenharia Elétrica (DELET), pela colaboração na construção do banco de faces.

Aos colegas do Laboratório de Instrumentação Eletro-Eletrônica (IEE), pela ambiente agradável e pelo auxílio durante o desenvolvimento do trabalho.

Ao Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), pela oportunidade de realização de trabalhos em minha área de pesquisa.

À CAPES pela provisão da bolsa de mestrado.

RESUMO

Em nossa vida diária, são utilizadas identidades digitais (IDDs) para acessar contas de e-mail, bancos e lojas virtuais, locais restritos, computadores compartilhados, e outros. Garantir que apenas usuários autorizados tenham o acesso permitido é um aspecto fundamental no desenvolvimento destas aplicações. Atualmente, os métodos de controle de acesso simples como senhas ou números de identificação pessoal não devem ser considerados suficientemente seguros, já que um impostor pode conseguir estas informações sem o conhecimento do usuário. Ainda, no caso de utilização de dispositivos físicos como cartões de identificação, estes podem ser roubados ou forjados. Para tornar estes sistemas mais confiáveis, técnicas de autenticação de identidades utilizando múltiplas verificações são propostas. A utilização de características biométricas surge como a alternativa mais confiável para tratar este problema, pois são, teoricamente, únicas para cada pessoa. Contudo, algumas características biométricas como a aparência facial podem variar com o tempo, implicando em um grande desafio para os sistemas de reconhecimento facial. Neste trabalho é combinado o acesso tradicional por senha com a análise da face para realizar a autenticação. Um método de aprendizagem supervisionada é apresentado e sua adaptação é baseada na melhora contínua dos modelos faciais, que são representados por misturas de gaussianas. Os resultados experimentais, obtidos sobre um conjunto de teste reduzido, são encorajadores, com 98% de identificação correta dos usuários e custo computacional relativamente baixo. Ainda, a comparação com um método apresentado na literatura indicou vantagens do método proposto quando usado como um pré-selecionador de faces.

Palavras-chave: Autenticação biométrica; Biometria; Controle de acesso; Feições faciais; Gerenciamento de identidades digitais; Modelagem de usuários.

ABSTRACT

In our daily life, we use digital identities (DIDs) to access e-mails, e-banks, e-shops, physical environments, shared computers, and so on. Guarantee that only authorized users are granted access is an important aspect in the development of such applications. Nowadays, the simple access control methods like passwords or personal identification numbers can not be considered secure enough, because an impostor can obtain and use these information without user knowledge. Also, physical devices like ID cards can be stolen. To make these systems more reliable, multimodal DID authentication techniques combining different verification steps are proposed. Biometric features appears as one of the most reliable alternatives to deal with this problem because, theoretically, they are unique for each person. Nevertheless, some biometric features like face appearances may change in time, posing a serious challenge for a face recognition system. In this thesis work, we use the traditional password access combined with human face analysis to perform the authentication task. An intuitive supervised appearance learning method is presented, and its adaptation is based on continuously improving face models represented using the Gaussian mixture modeling approach. The experimental results over a reduced test set show encouraging results, with 98% of the users correctly identified, with a relatively small computational effort. Still, the comparison with a method presented in the literature indicated advantages of the proposed method when used as a pre-selector of faces.

Keywords: Access control; Biometric authentication; Biometrics; Identity management systems; Facial features; User modeling.

LISTA DE ILUSTRAÇÕES

Figura 1 - Mistura de duas Gaussianas 1-D: (a) FDPs de duas distribuições normais; (b) Mistura das FDPs. 32

Figura 2 - Amostras dos indivíduos que compõem o BFIEE. 35

Figura 3 - Diferentes aparências para um indivíduo do BFIEE. 35

Figura 4 - Diagrama de blocos do algoritmo proposto. 37

Figura 5 - Resultado típico do detector de tons de pele: (a) Imagem reduzida rI ; (b) *Pixels* em rI classificados como pele. 38

Figura 6 - Ilustração do procedimento de busca por *blobs*: (a) Mapa de tons de pele; (b) Máscara da face; (c) Parte superior da máscara da face; (d) *Blobs* detectados; (e) *Blobs* restantes após a filtragem. 41

Figura 7 - Exemplos de localização aproximada dos olhos: Linha superior: resultados corretos; Linha inferior: resultados incorretos. (a) e (d) $EyeMap$; (b) e (e) $EyeMap'''$ obtido; (c) e (f) Localização dos olhos correta e incorreta, respectivamente. 43

Figura 8 - Ilustração do refinamento da localização do centro do olho: (a) Região do olho em tons de cinza R_{eye}^{I} ; (b) Mapa de bordas eR_{eye}^{I} ; (c) Magnitudes do gradiente; (d) Acumulador de Hough $accR_{eye}^{I}$; (e) Acumulador de Hough limitado $CaccR_{eye}^{I}$; (f) Localização final do centro do olho. 45

Figura 9 - Exemplos de triângulos para faces detectadas corretamente. 46

Figura 10 - Processo de normalização: (a) Faces detectadas; (b) Molde para normalização; (c) Faces normalizadas. 48

Figura 11 - Feições faciais selecionadas: (a) Olho esquerdo; (b) Olho direito; (c) Nariz. 49

Figura 12 - Evolução do processo de registro de usuários e formação de classes no espaço de feições direita: $\overrightarrow{FV}$ projetados em 2-D; esquerda: visão 3-D dos modelos das classes): (a) 1 classe; (b) 2 classes; (c) 10 classes. 55

Figura 13 - Representação das classes: (a) nuvem de pontos formada por vetores de atributos 2-D de duas aparências de um mesmo usuário (vermelho: aparência 1; azul: aparência 2); (b) representação usando modelo adaptativo único; (c) representação usando múltiplos modelos. 56

Figura 14 - Classes confundidas mais freqüentemente - esquerda: classe de entrada; direita: resultado da autenticação 61

Figura 15 - Curvas *recall-precision* obtidas. 70

LISTA DE TABELAS

Tabela 1 - Seqüência para acesso à BFIEE. 59
Tabela 2 - Matriz de confusão de classes para a CT1. 62
Tabela 3 - Matriz de confusão de classes para a CT2. 63
Tabela 4 - Matriz de confusão de classes para a CT3. 64
Tabela 5 - Matriz de confusão de classes para a CT4. 65
Tabela 6 - Ranking de recuperação de classes limitado nas 10 primeiras posições (utilizando 2500 amostras de autenticação). 66
Tabela 7 - Ranking de recuperação de classes limitado nas 5 primeiras posições (utilizando as 50 amostras medianas). 67
Tabela 8 - Ranking de recuperação de classes comparativo. 68
Tabela 9 - Ranking de recuperação de classes comparativo limitado nas 5 primeiras posições. 69
Tabela 10 - Taxas de autenticação (TA) para o segundo conjunto de testes. 71

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componentes Principais
AMMI	Autenticação Multi-Modal de Identidades
BFIEE	Banco de Faces do IEE-UFRGS
CIE	Comissão Internacional de Iluminação (Commission Internationale de l'Eclairage)
CMM	Comprimento Mínimo de Mensagem
CT	Configuração de Teste
FDP	Função Densidade de Probabilidade
IDD	Identidade Digital
IEE	Laboratório de Instrumentação Eletro-Eletrônica
ME	Maximização da Expectativa
MCC	Matriz de Confusão de Classes
MMG	Modelo de Mistura de Gaussianas
MSV	Máquina de Suporte Vetorial
PPGEE	Programa de Pós-Graduação em Engenharia Elétrica
SGID	Sistema de Gerenciamento de Identidades Digitais
SRF	Sistema de Reconhecimento Facial
TCC	Taxa de Classificações Corretas
TCA	Transformada de Confiança Adaptativa
TDC	Transformada Discreta dos Co-senos
TH	Transformada de Hough

TLL	Transformada Linear Local
TSA	Taxa de Sucesso de Aprendizagem
UFRGS	Universidade Federal do Rio Grande do Sul

SUMÁRIO

1 INTRODUÇÃO 13

2 REVISÃO DE LITERATURA 16
2.1 AUTENTICAÇÃO MULTI-MODAL DE IDENTIDADES 16
2.2 DETECÇÃO DE FACES E LOCALIZAÇÃO DOS OLHOS 19
2.3 SEGMENTAÇÃO DE IMAGENS BASEADA EM TONS DE PELE 22
2.4 ANÁLISE DE TEXTURAS 23
2.5 MODELAGEM DE CONJUNTOS DE DADOS 23

3 FUNDAMENTAÇÃO TEÓRICA 25
3.1 DETECÇÃO DE TONS DE PELE 25
3.2 MAPA CROMÁTICO DOS OLHOS 26
3.3 TRANSFORMAÇÕES GEOMÉTRICAS 27
3.4 TRANSFORMADA DE HOUGH 28
3.5 OBTENÇÃO DE DESCRITORES FACIAIS 29
3.6 MODELAGEM DE MISTURA DE GAUSSIANAS 31
3.7 CLASSIFICAÇÃO POR DECISÃO BAYESIANA 32

4 MÉTODOS E MATERIAIS 34
4.1 VISÃO GERAL DO ALGORITMO 36
4.2 EXTRAÇÃO DE DESCRITORES FACIAIS E REPRESENTAÇÃO DA FACE 37
4.2.1 DETECÇÃO DA FACE E LOCALIZAÇÃO DOS OLHOS 37
4.2.1.1 Segmentação baseada em tons de pele 38
4.2.1.2 Construção da máscara da face 39
4.2.1.3 Localização dos olhos na face detectada 40
4.2.1.4 Verificação da Face 45
4.2.2 NORMALIZAÇÃO DA FACE 47
4.2.3 OBTENÇÃO DE DESCRITORES FACIAIS 48
4.2.3.1 Seleção das feições 48
4.2.3.2 Extração de atributos 49
4.3 AUTENTICAÇÃO MULTI-MODAL DE USUÁRIOS 51
4.3.1 REGISTRO DE NOVOS USUÁRIOS 51
4.3.2 AUTENTICAÇÃO DE USUÁRIOS 52
4.3.2.1 Subclasses com probabilidades a priori iguais 53
4.3.2.2 Classes com probabilidades a priori iguais 54
4.3.3 ATUALIZAÇÃO DO MODELO DE APARÊNCIA FACIAL 54
4.3.3.1 Classes com modelo adaptativo único 56
4.3.3.2 Classes contendo múltiplos modelos 56
4.3.4 UM EXEMPLO DE AUTENTICAÇÃO DE USUÁRIO 57

5 RESULTADOS EXPERIMENTAIS 58
5.1 CONFIGURAÇÕES DE TESTE 59
5.2 AVALIAÇÃO DO ALGORITMO DE APRENDIZAGEM 60
5.3 AVALIAÇÃO DA CONFUSÃO INTER-CLASSES 60
5.4 AVALIAÇÃO DO CLASSIFICADOR COMO PRÉ-FILTRO 66
5.5 COMPARAÇÃO DO MÉTODO PROPOSTO 67
5.5.1 MÉTODOS COMO AUTENTICADORES DE IDENTIDADES 67
5.5.2 MÉTODOS COMO PRÉ-FILTROS 69
5.6 RESULTADOS EXPERIMENTAIS NA BASE DE DADOS EXTENDIDA 70
6 CONCLUSÕES 72
REFERÊNCIAS 74

1 INTRODUÇÃO

A utilização de identidades digitais (IDDs) em nosso dia-a-dia tem aumentado significativamente, principalmente, devido à popularização de serviços disponibilizados via internet como contas de e-mail, bancos e lojas virtuais. Identidades digitais também são utilizadas diariamente para acessar ambientes restritos em empresas, clubes, escolas, computadores compartilhados e terminais de auto-atendimento.

Com o avanço das pesquisas é provável que em um futuro próximo se possa utilizar uma única IDD universal (GOTH, 2005; SHIM; BHALLA; PENDYALA, 2005) que tornará o mundo virtual muito mais seguro. No entanto, atualmente, é possível ter uma IDD diferente para cada um dos ambientes citados acima e, para garantir os requisitos básicos de privacidade e segurança de dados e usuários, são necessários sistemas de gerenciamento de identidades digitais~(SGID) confiáveis (PHIRI; AGBINYA, 2006).

Um SGID deve ser capaz de executar três tarefas básicas: (a) gerenciar as IDDs; (b) verificar a relação legítima entre a identidade do usuário no mundo real e a sua correspondente no mundo digital (i.e., autenticar a identidade) e; (c) permitir acesso do usuário apenas às informações e/ou ambientes a que ele possui privilégio. A tarefa de autenticação das identidades cumpre papel fundamental no sistema e é responsável principal por sua confiabilidade.

A autenticação de identidades pode ser realizada utilizando diversas informações (e.g., senhas, números de identificação pessoal e características biométricas) e dispositivos (e.g., cartões de identificação e chaves), sendo que todas estas formas de autenticação estão sujeitas à falhas ou fraudes, com maior ou menor grau de segurança (RIHA; MATYÁS, 2000). No entanto, a utilização de características biométricas constitui um dos meios mais confiáveis para se realizar esta tarefa por serem, teoricamente, únicas para cada indivíduo e o acompanharem por toda a vida.

Diversas metodologias de autenticação foram propostas recentemente, utilizando as mais diversas características biométricas fisiológicas (e.g., impressões digitais, íris, retina, face, formato e padrão das veias das mãos) e comportamentais (e.g., voz, assinatura e ritmo de digitação) (RIHA; MATYÁS, 2000). As disponíveis comercialmente, baseadas na análise impressões digitais e íris, por exemplo, são confiáveis, mas dependem da cooperação do usuário e de leitores biométricos especiais.

Neste contexto, a utilização de faces como característica biométrica é uma abordagem atrativa, pois: (a) está presente em todos os indivíduos; (b) a aquisição das imagens pode ser realizada com equipamento popular (e.g., *webcam*) e; (c) pode ser obtida independentemente da cooperação do usuário. As desvantagens, quando comparados à íris e impressões digitais, ficam por conta da não estabilidade das feições com o passar do tempo e da aquisição em ambientes não controlados. Por este motivo, sistemas de reconhecimento facial (SRF) geralmente falham quando uma mudança considerável na aparência é apresentada ou as condições de aquisição variam significativamente.

O desafio é, portanto, desenvolver sistemas capazes de aprender com as variações de aparência do usuário com a manutenção das taxas de autenticação obtidas pelos sistemas disponíveis comercialmente. Para atingir estas metas uma nova linha de pesquisa trata o problema sob a ótica de autenticação multi-modal de identidades (AMMI). Nessa abordagem são realizadas autenticações combinadas de diferentes tipos de informações como senhas, dados pessoais, dispositivos físicos e características biométricas.

Neste trabalho é proposta uma metodologia para gerenciamento de identidades digitais utilizando AMMI, que combina informações de senha e aparência facial do usuário (i.e., modelo biométrico). Os modelos biométricos, baseados em misturas de Gaussianas e construídos por meio de um método de aprendizagem supervisionado bastante intuitivo, são constantemente atualizados para manter o controle das mudanças de aparência facial do usuário quando elas ocorrem. Ainda, além de serem utilizados como identificadores biométricos para a autenticação, os modelos podem proporcionar um estágio de pré-seleção de usuários[1].

Para testar o desempenho do algoritmo foi simulada uma situação onde uma estação de trabalho (e.g., um computador) é compartilhada entre vários usuários. Ao utilizar a estação

[1] Os usuários pré-selecionados podem ser enviados a um sistema especialista (e.g., um SRF) para análise mais criteriosa das faces.

de trabalho o usuário passa pela autenticação de senha e aparência facial. Se um possível impostor é identificado, o supervisor é acionado para: (a) garantir o acesso e atualizar o modelo biométrico caso o usuário seja legítimo ou; (b) rejeitar o acesso em caso de confirmação de fraude.

O restante do volume está organizado conforme a estrutura apresentada a seguir.

No capítulo 2 é apresentada a revisão de literatura. A autenticação multi-modal de identidades constitui o objetivo principal do trabalho e, portanto, da revisão. Ainda, as outras técnicas de processamento utilizadas são contextualizadas.

No capítulo 3 é realizada a fundamentação teórica acerca das técnicas utilizadas ao longo do trabalho e, no capítulo 4, o método proposto para o gerenciamento e autenticação de identidades é descrito de forma detalhada.

Os resultados experimentais obtidos são apresentados e avaliados no capítulo 5, juntamente com uma comparação entre o método proposto e um algoritmo para representação e reconhecimento de faces buscado na literatura.

Finalizando o volume, no capítulo 6 é realizada a discussão dos resultados obtidos e são apresentadas algumas sugestões para a continuação do trabalho.

2 REVISÃO DE LITERATURA

Neste capítulo é apresentada a revisão de literatura das diversas áreas de processamento de imagens utilizadas no desenvolvimento deste trabalho. As técnicas mencionadas são parte de etapas distintas do algoritmo proposto e, portanto, a revisão de literatura é dividida em subseções, cada qual apresentando os trabalhos em ordem cronológica.

A saber, a divisão das subseções respeita a seguinte estrutura de apresentação: (a) estado da arte para autenticação multi-modal de identidades; (b) revisão dos métodos para detecção de faces e localização dos olhos em imagens; (c) segmentação baseada em tons de pele; (d) análise de texturas e; (e) modelagem de conjuntos de pontos.

2.1 AUTENTICAÇÃO MULTI-MODAL DE IDENTIDADES

A autenticação de identidades consiste no módulo central de SGIDs, sendo, portanto, o principal objetivo deste trabalho. As técnicas de autenticação biométrica simples não são abordadas nesta revisão, podendo ser consultadas em vários trabalhos disponíveis na literatura, como o de (RÍHA; MATYÁS, 2000). Nesta seção são apresentados os trabalhos que realizam a autenticação multi-modal de identidades (AMMI) utilizando pelo menos uma característica biométrica do usuário.

Em (BIGUN *et al.*, 1997) é proposta a autenticação de usuários por meio da combinação de duas máquinas especialistas. A primeira é responsável pela autenticação da face utilizando filtros de Gabor e, a segunda, pela autenticação da voz do usuário por meio de coeficientes de predição linear. Testes mostram que a utilização de um supervisor treinado a partir de estatísticas Bayesianas produz melhores resultados que a combinação através da simples média dos resultados das máquinas.

Em (BEM-YACOUB; ABDELJAOUED; MAYORAZ, 1999) são avaliados alguns esquemas de classificação binária para combinar resultados de autenticação facial e da voz. Os esquemas testados são baseados em máquinas de suporte vetorial (MSVs), redes neurais, discriminantes lineares de Fisher e classificador Bayesiano. O melhor resultado foi obtido pelo classificador Bayesiano e o desempenho do sistema multi-modal é consideravelmente superior quando comparado à autenticação biométrica simples.

No trabalho de (O'GORMAN, 2003) é realizado um estudo comparativo entre autenticações utilizando senhas, dispositivos físicos e atributos biométricos de forma isolada e combinada. São apresentados protocolos para cada autenticador proposto e situações de aplicação, discutindo onde cada método possui vantagens e desvantagens.

Em (WANG; TAN; JAIN, 2003), a autenticação combina atributos biométricos da íris e da face, sendo usadas duas estratégias para realizar a fusão dos classificadores. Na primeira são calculadas somas ponderadas das distâncias entre a amostra de entrada e o usuário legítimo e, na segunda, são utilizados discriminantes de Fisher e redes neurais. Os resultados são comparados aos obtidos pelos classificadores individuais e os autores concluem que, além de aumentar a acurácia da autenticação, a combinação das características biométricas promove um aumento na população de usuários que podem ser diferenciados.

Em (BIGUN *et al.*, 2003) é realizada a autenticação de identidades por meio da combinação de informações de voz e impressões digitais. Um novo supervisor adaptativo baseado na qualidade do sinal biométrico de entrada é avaliado. Resultados experimentais sobre dados coletados de telefones celulares são apresentados, demonstrando os benefícios do esquema proposto.

Em (KRAWCZYK, 2005) é apresentado um sistema de autenticação de identidades para utilização em *tablet* PCs. A assinatura *on-line* e a voz são utilizadas para a autenticação biométrica por serem mais convenientes, já que este tipo de equipamento vem aparelhado com os sensores e *hardware* adequados. O sistema foi testado em um banco de 100 usuários, obtendo uma taxa de erro médio de 0,72%.

Em (SNELICK, 2005) é realizado um estudo para avaliação em larga escala dos métodos considerados estado da arte em AMMI utilizando faces e impressões digitais, sendo que os métodos avaliados são utilizados em equipamentos comercializados. Os estudos, realizados sobre um banco de 1000 usuários, comprovam que a utilização combinada destas

feições apresenta um ganho significativo na exatidão dos sistemas. Ainda, são introduzidos novos métodos para normalização e fusão dos atributos biométricos.

Em (PHIRI; AGBINYA, 2006) é proposta uma metodologia para SGID concebida com o intuito de manter a segurança e privacidade de usuários e provedores de serviço em ambientes virtuais de forma transparente, confiável e eficiente. O sistema realiza a fusão de atributos identificadores quaisquer por meio de uma rede neural e, posteriormente, realiza a autenticação do usuário.

Em (SHU; DING, 2006) os autores introduzem a transformada de confiança adaptativa (TCA) para solucionar o problema de incomparabilidade entre saídas de diferentes classificadores. Testes foram realizados sobre um banco de 100 usuários, utilizando informações de face, íris e assinaturas *on-line* e *off-line*. Os resultados indicam que para o mesmo tipo de fusão, a TCA produz desempenho melhor e mais robusto que outros métodos de normalização.

Em (KOREMAN *et al.*, 2006) é apresentado um sistema de autenticação multi-modal utilizado para permitir a comunicação móvel segura a partir de *handhelds* (e.g., PDAs), chamado *SecurePhone*. São utilizados três atributos biométricos: (a) voz; (b) face e; (c) assinatura. Para a aquisição dos atributos o usuário lê uma mensagem para que a voz seja captada pelo microfone, assina sobre a tela e captura uma imagem facial utilizando a câmera. São realizados diversos experimentos para testar a fusão dos atributos e, segundo os autores, um nível de autenticação “aceitável” para diversas aplicações foi atingido.

Em (MELEN, 2006) é proposto um sistema baseado em AMMI altamente seguro que combina cartões de identificação sem contato e impressões digitais. Após a verificação biométrica, o cartão assegura que o usuário está presente e desbloqueia a estação de trabalho, dando os privilégios de usuário permitidos. A força do mecanismo é garantida pela limitação em tempo e espaço da associação lógica entre a identificação biométrica no mundo digital e o portador do cartão no mundo real.

Em (ZHOU *et al.*, 2007) é proposta a combinação de informações da face e impressões digitais para realizar a autenticação de identidades. Para tentar eliminar a influência da variação de pose da face, são obtidas três fotografias do usuário a partir de diferentes pontos de vista. Estas imagens são processadas e dão origem a outra imagem chamada de "face 2-D ótima". A fusão dos classificadores é realizada por meio de SVMs e

experimentos demonstram que os resultados são melhores dos que os obtidos quando os atributos biométricos são utilizados isoladamente.

2.2 Detecção de Faces e Localização dos Olhos

O primeiro trabalho voltado para detecção de faces que se tem conhecimento é o proposto por (SAKAI; NAGAO; KANADE, 1972). Este trabalho utilizava linhas traçadas em fotografias com o intuito de localizar feições faciais na imagem. As técnicas eram rígidas, baseadas em heurísticas simples e conhecimentos antropométricos. Ainda, não era robusto, sendo aplicável somente em condições bastante limitadas com respeito à pose, escala, iluminação e rotação.

Em (LEUNG; BURL; PERONA, 1995) é apresentado um algoritmo para detectar faces quase frontais em cenários confusos. O algoritmo utiliza a combinação de detectores de atributos locais com modelos estatísticos das distâncias mútuas entre as feições faciais. Os detectores de atributos locais geram constelações de candidatos, sendo que a procura pela melhor constelação pode ser tratada como um problema de "casamento" aleatório de grafos. São reportados resultados de 95% de detecções corretas para imagens quase frontais.

Em (ROWLEY; BALUJA; KANADE, 1998) é proposta a utilização de redes neurais para a detecção de faces frontais. Uma rede neural multi-camada é utilizada para aprender padrões de face verdadeiros e falsos. Com a rede treinada, uma pequena janela percorre a imagem e, com base em heurísticas, o classificador neural decide se na posição atual da janela existe ou não uma face. Os melhores resultados mencionados são de 90,3% para o conjunto de 130 faces da base de imagens *CMU Face Database*.

Em (SUNG; POGGIO, 1998) é proposto um método para detectar faces frontais baseado na aprendizagem por exemplos. A técnica modela os padrões de distribuição de faces humanas por meio de *clusters*, formados por misturas de Gaussianas, representando faces e não-faces. Sobre cada localização escolhida da imagem, um vetor de diferenças de feições e extraído e comparado com o modelo baseado em distribuições. A detecção é determinada por um classificador neural treinado através de medidas dos vetores de diferenças. Os autores propõem várias métricas, sendo que, para o melhor caso, a taxa de detecção atingiu 96,7% utilizando uma base de dados própria.

Em (KAWAGUCHI; HIDAKA; RIZON, 2000) é aplicado um detector de vales de intensidade sobre a imagem da face em tons de cinza e, a partir de modelos que tratam as íris como círculos, são selecionados os melhores *blobs*[2]. Sobre a região que contém estes *blobs* são aplicados o detector de bordas de Canny e a transformada de Hough (TH) para detectar círculos, sendo que os círculos mais votados são associados às íris. Como mais de dois *blobs* podem ser aceitos, é calculada uma função custo para definir as íris corretas. O método obteve uma taxa de acerto de 96,5%, mas uma parte do processo é realizada de forma interativa.

Devido ao aumento das pesquisas envolvidas com o processamento de faces, uma grande quantidade de técnicas foram propostas para detecção de faces. Sendo assim, em (HJELMAS; LOW, 2001) os autores propõem uma divisão das técnicas utilizadas para detecção de faces até o momento. A classificação obedece duas abordagens: (a) baseadas em feições e; (b) baseadas na imagem. Estas abordagens são subdivididas dependendo das técnicas e feições empregadas.

As abordagens baseadas em feições utilizam informações de baixo nível (e.g., cor, movimento e formas) e tendem a ser algoritmos rápidos e invariantes à escala e rotação. Contudo o desempenho é limitado quando utilizado em imagens com baixa resolução e cenário confuso. Nas abordagens baseadas na imagem são obtidos modelos estatísticos da face utilizando um conjunto de treinamento. Após o treinamento a imagem é percorrida pelos modelos e as faces são detectadas por comparação. As técnicas deste grupo são tidas como estado da arte para detecção de faces, porém seus algoritmos são complexos e de elevado custo computacional, já que são necessários modelos de face para diversas orientações, poses e escalas.

Em (VIOLA; JONES, 2001) é utilizado o conceito de "imagem integral" para representar uma imagem. Um algoritmo de aprendizagem, baseado na técnica *AdaBoost*, seleciona uma pequena quantidade de feições visuais críticas, levando à obtenção de classificadores bastante rápidos. Vários classificadores são combinados em cascata fazendo com que o algoritmo rejeite rapidamente regiões de fundo e se dedique as regiões que podem conter uma face. São reportados resultados de 93,7% de faces detectadas.

Em (YANG; KRIEGMAN; AHUJA, 2002), os autores realizam uma revisão dos métodos para detecção de faces propostos até o momento. São realizadas a classificação e

[2] Um *blob* é uma região sem tom de pele totalmente cercada pela máscara da face, ou seja, é um "buraco" na máscara.

avaliação das técnicas, apontando vantagens e limitações. Ainda, são discutidos aspectos relevantes como: (a) obtenção dos dados; (b) métricas de avaliação e; (c) *benchmarking*.

Em (HSU; ABDEL-MOTTALEB; JAIN, 2002) é proposto um sistema de detecção de faces bastante complexo no qual a tarefa é dividida em duas etapas: (a) obtenção de regiões candidatas e; (b) verificação das faces encontradas. A obtenção das regiões candidatas é realizada no espaço de cores YC_bC_r, após compensação da iluminação e detecção de *pixels* com tom de pele. Sobre as regiões de pele é realizada uma segmentação baseada na variância e no agrupamento das regiões conectadas. Na seqüência, são obtidos mapas cromáticos específicos para localizar os olhos e a boca. Estas feições formam um triângulo que deve respeitar o conjunto de regras para a face ser validada. São apresentados resultados para três conjuntos de imagens, sendo que o melhor resultado indica taxa de detecção de 91,63% com 27 falsos positivos para ao banco de imagens *Champion Database*.

Em (WANG *et al.*, 2005) os autores mostram que a taxa de reconhecimento de faces cresce significativamente com o aumento na precisão da localização dos centros dos olhos. Segundo os autores, isto ocorre porque a normalização é mais eficiente, proporcionando um melhor alinhamento das imagens.

Em (LEE; PARK; PARK, 2005) é proposto um método para detecção de faces em imagens coloridas e em tons de cinza. A busca por feições é baseada somente em intensidades, sendo que, nas imagens coloridas, ela é acelerada pela eliminação de regiões sem tons de pele. Um mapa de vales de intensidade é construído e as localizações dos olhos e da boca são obtidas, formando um triângulo de feições. As regiões que formam este triângulo de feições são submetidas à classificação baseada na similaridade com um modelo pré-definido. O algoritmo foi testado para três conjuntos de imagens e os melhores resultados obtidos foram de 96,67% de taxa de detecção, para uma base contendo 270 imagens construída pelo autor.

Em (PENG *et al.*, 2005) são combinadas duas técnicas para determinar os centros dos olhos em imagens em tons de cinza. Em uma primeira etapa, é aplicado o operador gradiente sobre a face e, a partir de projeções verticais e horizontais combinadas com regras heurísticas, a região dos olhos é determinada. Posteriormente, são determinados os modelos dos olhos e procede-se uma busca pela região que melhor "casa" com estes padrões para determinar os centros dos olhos. O algoritmo demonstrou uma taxa de detecção de 95,2% em faces sem a presença de óculos.

Em (WARING; LIU, 2005) é apresentado um método de detecção de faces utilizando histogramas espectrais e MSVs. Cada pequena porção da imagem é representada por seu histograma espectral, que é um vetor de atributos, obtido de histogramas de imagens filtradas. As MSVs são treinadas com 4500 amostras de faces e 8000 amostras que não são faces e um algoritmo de compensação de iluminação é utilizado. A taxa de detecção chega a 96,7% para o melhor caso relatado.

Em (SAAD *et al.*, 2006) é proposto um algoritmo para detectar faces frontais em imagens coloridas cujo diferencial é o processo de agrupamento, divisão e recorte de regiões com tons de pele. Este procedimento faz com que restem apenas candidatos com bom grau de simetria, que é uma característica das faces. Em cada região candidata é realizada a busca por olhos, baseada em relações geométricas e informações de intensidade. Passando por esta etapa, a face é alinhada com um modelo e a correlação é calculada para então ocorrer a classificação final. Os resultados, obtidos sobre um conjunto de 200 imagens contendo 240 faces, apresentam taxa de detecção igual a 95,5% com apenas 3 falsos positivos.

2.3 SEGMENTAÇÃO DE IMAGENS BASEADA EM TONS DE PELE

A segmentação de imagens utilizando tons de pele é uma técnica bastante empregada no processamento de faces para reduzir o espaço de busca por estas estruturas. Diversos são os métodos apresentados na literatura, sendo que, neste trabalho, é utilizado um classificador de *pixels* baseado em informações de cor. Um grande número de abordagens para modelar tons de pele utilizando classificação de *pixels* é apresentado em (VEZHNEVETS; SAZONOV; ANDREEVA, 2003).

Em (KOVAC; PEER; SOLINA, 2003) os autores estudam a utilização de diferentes espaços de cores para detectar tons de pele. São propostas regras simples e eficazes para detecção destas tonalidades nos espaço de cores RGB (suscetível as condições de iluminação) e no espaço YC_bC_r, resultando em classificadores bastante rápidos.

Em (FALIPOU, 2006) é apresentado um detector de faces baseado em feições de baixo nível no qual a segmentação por tons de pele é uma extensão da proposta em (KOVAC; PEER; SOLINA, 2003) para o espaço RGB. Duas novas regras, baseadas nas componentes R e G normalizadas, são adicionadas ao classificador para torná-lo mais robusto.

2.4 ANÁLISE DE TEXTURAS

Há diversas técnicas para análise imagens texturizadas disponíveis na literatura. Como o objetivo deste trabalho não é o desenvolvimento de uma nova técnica para a extração atributos de texturas, foram utilizados como referência os artigos de revisão apresentados a seguir.

Em (TUCERYAN; JAIN, 1998) é realizada uma revisão dos conceitos básicos, métodos e técnicas para o processamento de imagens contendo texturas. Modelos para extração de atributos de textura baseados em geometria, fractais e processamento de sinais são apresentados. Problemas de classificação, segmentação e extração de formas a partir de texturas são discutidos, bem como algumas aplicações.

Em (RANDEN; HUSOY, 1999) os autores se dedicam ao estudo de técnicas de extração de atributos de texturas baseados na filtragem das imagens. Estas abordagens incluem *wavelets*, *wavelet packet*, filtros de Gabor, preditores lineares, transformada discreta de co-senos (TDC), filtros FIR otimizados e outras, sendo que os atributos são calculados como a energia local da resposta dos filtros. Comparações com abordagens clássicas, não baseadas em filtragem, como matrizes de co-ocorrência (método estatístico) e atributos auto-regressivos (método baseado em modelagem) são realizadas. Finalizando o trabalho é apresentado um *ranking* das técnicas avaliadas.

Em (DRIMBAREAN; WHELAN, 2001) os autores visam avaliar a contribuição do atributo cor para a classificação de texturas. Três abordagens relevantes para extração de atributos de textura em tons de cinza, sendo elas as matrizes de co-ocorrência, filtros de Gabor e transformadas lineares locais (TLLs), são expandidas para imagens coloridas. Resultados experimentais indicam que a incorporação da cor na análise das texturas aumenta o desempenho das técnicas apresentadas.

2.5 MODELAGEM DE CONJUNTOS DE DADOS

Em (FUKUNAGA, 1990) são apresentados diversos métodos de modelagem paramétricos e não-paramétricos para reconhecimento estatístico de padrões. No entanto, optou-se por uma técnica bastante conhecida e utilizada em processamento de sinais, que é a modelagem por mistura de Gaussianas (MMG). É apresentado nesta revisão, e utilizado no

durante o desenvolvimento do trabalho, o algoritmo proposto em (FIGUEIREDO; JAIN, 2002), que representa uma poderosa ferramenta para modelagem de conjuntos de dados complexos utilizando modelos paramétricos.

Este algoritmo proposto por Figueiredo e Jain (2002) é um método não supervisionado para o aprendizado de modelos de misturas a partir de dados multidimensionais. O algoritmo é capaz de selecionar o número de componentes da mistura e, ao contrário do algoritmo baseado na maximização da expectativa (ME), segundo os autores, não necessita de uma inicialização cuidadosa. A técnica pode ser aplicada a qualquer tipo de mistura paramétrica, sendo apresentado, no artigo consultado, um exemplo utilizando misturas de Gaussianas.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentadas as definições teóricas acerca das técnicas de processamento envolvidas no algoritmo proposto, sendo que elas contribuem para: (a) a detecção de *pixels* com tons de pele nas imagens; (b) a localização das faces; (c) a localização dos olhos; (d) a extração de atributos de textura das regiões faciais; (e) a obtenção de modelos de aparência facial e; (f) a classificação das faces. É abordado, principalmente, o equacionamento referente às técnicas utilizadas, conforme segue.

3.1 DETECÇÃO DE TONS DE PELE

Para reduzir o espaço de busca por faces em imagens coloridas, uma técnica bastante utilizada é a segmentação da imagem baseada em tons de pele. Diversas abordagens baseadas na análise de cor dos pixels são apresentadas em (VEZHNEVETS; SAZONOV; ANDREEVA, 2003).

O método utilizado neste trabalho é o proposto em (FALIPOU, 2006), que é uma extensão da proposta de (KOVAC; PEER; SOLINA, 2003) para detecção de tons de pele sob iluminação natural (iluminante CIE-D65)[3]. A detecção é realizada por meio de um conjunto de regras bastante simples que classifica cada *pixel* da imagem, no espaço de cores RGB, como sendo "pele" ou "não-pele". Portanto, *pixel* é classificado como pertencente a uma região de pele na imagem se todas as regras da equação 3.1 são satisfeitas.

[3] Apesar de serem obtidas para iluminação natural, as regras se mostraram eficientes no caso *indoor* apresentado neste trabalho.

$$\begin{aligned} &R > 95 \quad e \quad G > 45 \quad e \quad B > 20, \\ &0.25 < nR < 0.65 \quad e \quad 0.22 < nG < 0.38, \\ &\left(\max(R,G,B) - \min(R,G,B)\right) > 15, \\ &\left|R - G\right| > 15 \quad e \quad R > G \quad e \quad R > B \end{aligned} \tag{3.1}$$

onde, nR e nG são as componentes cromáticas normalizadas dadas por:

$$nR = \frac{R}{R+G+B}; \quad nG = \frac{G}{R+G+B} \tag{3.2}$$

3.2 MAPA CROMÁTICO DOS OLHOS

No detector de faces proposto em (HSU; ABDEL-MOTTALEB; JAIN, 2002) é construído um mapa cromático para ressaltar a região dos olhos, utilizando o espaço de cores YC_bC_r. O mapa é obtido por meio da combinação de aspectos cromáticos e de luminância. Segundo os autores, a região dos olhos possui valores relativamente altos para a componente C_b e baixos para a componente C_r, quando comparados às regiões de pele. Utilizando-se desta característica, o mapa cromático MO_C proposto para localizar a região dos olhos é dado por:

$$MO_C(x,y) = \frac{1}{3} \cdot C_b^2(x,y) + \left(\tilde{C}_r(x,y)\right)^2 + \left(\frac{C_b(x,y)}{C_r(x,y)}\right) \tag{3.3}$$

onde, $C_b^2(x,y)$, $\left(\tilde{C}_r(x,y)\right)^2$ e $\frac{C_b(x,y)}{C_r(x,y)}$ são normalizados para a faixa $[0,255]$ e $\tilde{C}_r(x,y)$ é o negativo de $C_r(x,y)$ (i.e., $\tilde{C}_r(x,y) = 255 - C_r(x,y)$).

Os autores observam, ainda, que a região dos olhos possui *pixels* claros e escuros, simultaneamente. Portanto, também é proposto um mapa baseado na dilatação e erosão de tons de cinza sobre a imagem de luminâncias, definido como mapa de luminância dos olhos MO_L, dado por:

$$MO_L(x,y) = \frac{Y(x,y) \oplus g_\sigma(x,y)}{Y(x,y) \ominus g_\sigma(x,y) + 1} \tag{3.4}$$

onde, $Y(x,y)$ é o valor da luminância e $g_\sigma(x,y)$ é o valor da função estruturante no ponto (x,y).

O mapa de crominâncias é realçado por uma equalização de histograma e combinado com o mapa de luminância, resultando no mapa de olhos MO:

$$MO(x,y) = MO_C(x,y) \cdot MO_L(x,y) \tag{3.5}$$

3.3 TRANSFORMAÇÕES GEOMÉTRICAS

Durante o desenvolvimento do trabalho é necessário realizar transformações geométricas nas imagens para obter a normalização das faces. Assumindo que I é a imagem original, a versão transformada tI de cada uma das transformações (translação, rotação e escala) é dada por:

$$tI(x,y) = I(x+T, y+T) \tag{3.6}$$

$$tI(x,y) = I(x \cdot S, y \cdot S) \tag{3.7}$$

$$tI(x,y) = I(x \cdot R, y \cdot R) \tag{3.8}$$

T é a matriz de translação, S de escalonamento e R de rotação.

As equações 3.9 e 3.10 apresentam as matrizes de rotação e escalonamento para transformações diretas no plano (GONZALEZ; WOODS, 1992):

$$S = \begin{bmatrix} sx & 0 \\ 0 & sy \end{bmatrix} \tag{3.9}$$

$$R = \begin{bmatrix} \cos(\alpha) & sen(\alpha) \\ -sen(\alpha) & \cos(\alpha) \end{bmatrix} \tag{3.10}$$

cujas matrizes de transformação inversas são dadas por:

$$S^{-1} = \begin{bmatrix} 1/sx & 0 \\ 0 & 1/sy \end{bmatrix} \tag{3.11}$$

$$R^{-1} = \begin{bmatrix} \cos(\alpha) & -sen(\alpha) \\ sen(\alpha) & \cos(\alpha) \end{bmatrix} \tag{3.12}$$

onde, α é o ângulo de rotação e sx e sy são as constantes de escala para os eixos x e y, respectivamente.

Para evitar o surgimento de distorções nas imagens, as transformações de escala e rotação são realizadas por meio de mapeamento inverso. Ao invés de transportar os valores da imagem de entrada para a imagem de saída (mapeamento direto), é realizado o processo inverso. É construída a imagem de saída tI, com tamanho suficiente para acomodar a transformação e, para cada *pixel* de tI, o valor é buscado na imagem de entrada I utilizando uma técnica de interpolação (bi-linear neste trabalho). Após a transformação, os pontos de interesse (e.g., centros dos olhos) são transportados de I para tI utilizando o mapeamento direto.

3.4 TRANSFORMADA DE HOUGH

A transformada de Hough (TH) é uma técnica de processamento desenvolvida para detectar formas geométricas parametrizáveis (e.g., retas, círculos e elipses) em imagens digitais. Neste trabalho utilizamos a abordagem *declividade-interceptação* (GONZALEZ; WOODS, 1992), que é a forma mais simples para representar retas, dada por:

$$y = m \cdot x + b \tag{3.13}$$

onde, m é a declividade da reta e b é o valor de y onde a reta cruza o eixo das ordenadas.

Portando, uma reta na imagem I pode ser completamente caracterizada por m e b no espaço de parâmetros de Hough. De forma análoga, um ponto (x, y) na imagem I representa um reta no espaço de parâmetros.

A aplicação da transformada de Hough na imagem é precedida por duas operações de pré-processamento, sendo elas: (a) a detecção das bordas, que resulta em um mapa de bordas lógico MB e; (b) a obtenção dos gradientes nas direções x e y, resultando nos mapas G_x e G_y, respectivamente. Com base nestas informações, procede-se a aplicação da transformada de Hough.

Para cada *pixel* de borda detectado, ou seja, onde $MB(x, y) = 1$, a declividade local da borda $m(x, y)$ é calculada e uma linha é gerada no espaço de parâmetros de Hough. O espaço de parâmetros de Hough é contínuo e, portanto, um acumulador é implementado para discretizar este espaço e tornar o armazenamento dos resultados possível. Assim, as células do acumulador que representam a linha gerada no espaço de parâmetros de Hough são

incrementadas. O procedimento descrito acima é expresso matematicamente nas equações 3.14-3.16:

$$1)\quad m(x,y)=\frac{G_x^I(x,y)}{G_y^I(x,y)} \tag{3.14}$$

$$2)\quad yy = round\left(m(x,y)\cdot(x-xx)+y\right),\ xx\in[x-d:x+d] \tag{3.15}$$

$$3)\quad acc(xx,yy)=acc(xx,yy)+1 \tag{3.16}$$

onde, xx e yy são as coordenadas das células do acumulador de Hough que são incrementadas e d limita o comprimento da reta gerada em $2d+1$ células do acumulador.

3.5 OBTENÇÃO DE DESCRITORES FACIAIS

A obtenção de descritores faciais se mostrou uma etapa crítica no desenvolvimento de algoritmos para classificação, identificação ou reconhecimento de faces. Para realizar a diferenciação entre faces, é imprescindível que se tenham descritores capazes de captar diferenças sutis entre feições faciais de indivíduos diferentes. Os descritores podem ser globais, quando obtidos da face como um todo, ou locais, quando são extraídos de pontos, regiões ou estruturas específicas da face.

Depois de selecionadas as feições faciais de interesse, é realiza-se o processo de extração de atributos que descrevem estas regiões. Texturas e cores são aceitos como dois atributos-chave na análise de imagens (DRIMBAREAN; WHELAN, 2001) e, por isso, são utilizados neste trabalho para a diferenciação de faces. Dentre as diversas técnicas de processamento de imagens freqüentemente utilizados para a análise de texturas, destacam-se as matrizes de co-ocorrência, filtros de Gabor, *wavelets* e as transformadas lineares locais (TLLs), sendo esta última adotada neste trabalho.

As TLLs (UNSER, 1986) caracterizam as texturas por meio de um conjunto de atributos estatísticos, extraídos das imagens de saída de um banco de filtros que executam estas transformações. Estes filtros são sintonizados para capturar propriedades específicas da textura local. Conforme (DRIMBAREAN; WHELAN, 2001), a utilização da transformada discreta dos co-senos (TDC) apresenta melhores resultados para discriminação de texturas quando comparada aos filtros de Gabor e matrizes de co-ocorrência. Portanto, a TDC será utilizada, neste trabalho, como meio de extração de atributos estatísticos das texturas. Ainda, a

TDC é ortogonal e separável e, portanto, pode ser computada por algoritmos rápidos. Assim, os autores sugerem que um vetor base TDC h_m, com dimensão $N \times 1$, pode ser calculado por,

$$h_m(k) = \begin{cases} \dfrac{1}{\sqrt{N}} & se\ m = 0 \\ \sqrt{\dfrac{2}{N}} \cos \dfrac{(2k-1)m\pi}{2N} & se\ m > 0 \end{cases} \tag{3.17}$$

Estes vetores são usados para obter filtros TDC-2D, com N^2 coeficientes, por meio do produto interno $h_{mn} = h_m h_n^T$. Considerando a imagem original $I(x, y)$, os atributos de textura f_{mn} são definidos como sendo a variância da imagem I_{mn}, filtrada com a máscara h_{mn}.

$$I_{mn} = I * h_{mn} \tag{3.18}$$

$$f_{mn} = \frac{1}{M^2} \sum_{x,y}^{M} \left(I_{mn}(x, y) - \mu_{mn} \right)^2 \tag{3.19}$$

onde,

$$\mu_{mn} = \sum_{x,y=0}^{M} I_{mn}(x, y) \tag{3.20}$$

Para uma máscara de tamanho $N = 3$, temos que os vetores base TDC são $h_0 = [1,1,1]$, $h_1 = [1,0,-1]$ e $h_2 = [1,-2,1]$. Utilizando o produto interno descrito anteriormente, podem ser construídos os 9 filtros TDC-2D apresentados na equação 3.21. Estes filtros passa-banda são utilizados para a extração dos atributos, segundo as equações 3.18-3.20, e capturam aspectos específicos da textura.

$$\begin{aligned} h_{00} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & h_{01} &= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} & h_{02} &= \begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix} \\ h_{10} &= \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} & h_{11} &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} & h_{12} &= \begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ 1 & 0 & -1 \end{bmatrix} \\ h_{20} &= \begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix} & h_{21} &= \begin{bmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix} & h_{22} &= \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \end{aligned} \tag{3.21}$$

No caso de imagens em tons de cinza, a máscara h_{00} pode ser excluída pois geralmente não captura informações de textura, somente intensidade média. Porém, quando trabalhamos com imagens coloridas, elas podem conter informações relevantes (DRIMBAREAN; WHELAN, 2001).

3.6 MODELAGEM DE MISTURA DE GAUSSIANAS

Durante o desenvolvimento do trabalho é necessário construir modelos para "explicar" as aparências dos diferentes usuários e proceder a autenticação. Os modelos devem representar as diferentes classes a partir de conjuntos de pontos n-D, sendo fundamental que possam ser parametrizáveis. Com base nestes requisitos, a abordagem que parece mais adequada para este tipo de modelagem é a utilização de misturas finitas. Este tipo de modelagem representa uma eficiente ferramenta para descrever funções de densidade de probabilidade (FDPs) complexas a partir de dados uni- ou multi-dimensionais (FIGUEIREDO; JAIN, 2002). Como os classificadores Bayesianos utilizados neste trabalho baseiam-se no cálculo de probabilidades, freqüentemente dadas pela FDP da distribuição, esta abordagem se mostra adequada para o tratamento do problema.

O método padrão usado para ajustar modelos de mistura a conjuntos de dados é o algoritmo de maximização da expectativa (ME), que converge a uma estimativa dos parâmetros da mistura que proporcionam a máxima verossimilhança. No entanto, segundo (FIGUEIREDO; JAIN, 2002), esta abordagem possui dois aspectos negativos importantes, sendo eles: (a) a necessidade de inicialização adequada e; (b) o ME pode convergir para os limites do espaço de parâmetros. Para tratar este problema os autores propõem um algoritmo para construção não supervisionada de modelos de misturas, baseado no critério de comprimento mínimo de mensagem (CMM), que é utilizado neste trabalho. As misturas finitas contendo k componentes são expressas por:

$$p(y \mid \theta) = \sum_{m=1}^{k} \alpha_m \cdot p(y \mid \theta_m) \tag{3.22}$$

onde, $\alpha_1, \mathrm{K}, \alpha_k$ são as probabilidades das componentes da mistura, cada θ_m é um conjunto de parâmetros que define a *m*-ésima componente da mistura, e $\theta \equiv \{\theta_1, \mathrm{K}, \theta_k, \alpha_1, \mathrm{K}, \alpha_k\}$ é o conjunto completo de parâmetros necessários para especificar a mistura.

Tratando-se de probabilidades, α_m deve satisfazer as condições:

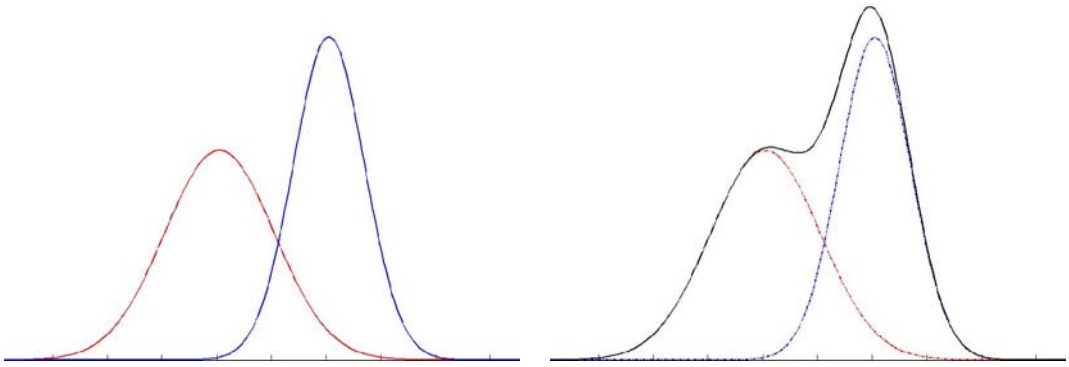
$$\alpha_m \geq 0, \quad m = 1,\text{K}\ ,k \quad e \quad \sum_{m=1}^{k} \alpha_m = 1 \tag{3.23}$$

Se as funções utilizadas são Gaussianas (caso deste trabalho), os modelos obtidos são os bem conhecidos modelos de misturas de Gaussianas (MMG), dados por:

$$MMG = \sum_{m=1}^{k} \alpha_m \cdot G(\mu_m, \Sigma_m) \tag{3.24}$$

onde, $G(\mu_m, \Sigma_m)$ representa a Gaussiana m da mistura.

A Figura 1 ilustra a FDP resultante para a mistura de duas Gaussianas 1-D.



(a) (b)

Figura 1 - Mistura de duas Gaussianas 1-D: (a) FDPs de duas distribuições normais; (b) Mistura das FDPs.

Além da abordagem proposta em (FIGUEIREDO; JAIN, 2002), destacam-se os trabalhos de (ZIVKOVIC; HEIJDEN, 2004) e (LAW; FIGUEIREDO; JAIN, 2004), que realizam a atualização *on-line* de modelos de misturas finitas.

3.7 CLASSIFICAÇÃO POR DECISÃO BAYESIANA

Neste trabalho é implementado um classificador baseado no teorema de Bayes. Na verdade, conhecendo a função densidade de probabilidade (FDP) condicional de todas as classes, o problema torna-se um teste de hipóteses estatísticas, conforme (FUKUNAGA, 1990) na definição utilizada para introduzir a classificação por decisão Bayesiana apresentada a seguir.

Suponha uma situação onde tem-se duas classes, ω_1 e ω_2, e uma amostra que entra no sistema deve pertencer a uma destas classes. As probabilidades *a priori* e as FDPs condicionais são conhecidas. Seja X um vetor de observação e o objetivo é determinar se X pertence a classe ω_1 ou ω_2, conforme a regra de decisão baseada simplesmente em probabilidades:

$$\begin{aligned} &\omega_1 \quad se \quad q_1(X) > q_2(X) \\ &\omega_2 \quad se \quad q_1(X) < q_2(X) \end{aligned} \tag{3.25}$$

onde, $q_i(X)$ é a probabilidade *a posteriori* de ω_i, dado X.

A equação 3.25 indica que, se a probabilidade de ω_1 dado X é maior que a probabilidade de ω_2 dado X, X é classificado como pertencente a classe ω_1 e vice-versa. A probabilidade *a posteriori* $q_i(X)$ pode ser calculada através da probabilidade *a priori* P_i e a da FDP condicional $p_i(X)$, usando o teorema de Bayes, dado por:

$$q_i(X) = \frac{P_i \cdot p_i(X)}{p(X)} \tag{3.26}$$

onde, $p(X)$ é a FDP resultante da mistura das FDPs das classes.

Sendo que $p(X)$ é positivo e comum aos dois lados da desigualdade, a regra de decisão entre as classes ω_1 e ω_2, na equação 3.25, pode ser expressa por:

$$\begin{aligned} &\omega_1 \quad se \quad P_1 \cdot p_1(X) > P_2 \cdot p_2(X) \\ &\omega_2 \quad se \quad P_1 \cdot p_1(X) < P_2 \cdot p_2(X) \end{aligned} \tag{3.27}$$

Estas relações entre classes consistem na teoria básica referente aos classificadores utilizados neste trabalho. Modificações nestas relações são propostas para validar as metodologias de classificação propostas, conforme será apresentado no próximo capítulo.

4 MÉTODOS E MATERIAIS

Neste capítulo é apresentada, de forma detalhada, a metodologia empregada na implementação do algoritmo proposto. Diversas técnicas de processamento de imagens são utilizadas no desenvolvimento do método, sendo que as escolhas são justificadas ao longo do texto, considerando aspectos práticos como complexidade de implementação, desempenho e adequação às necessidades do trabalho.

Para avaliar o algoritmo proposto foi construído um banco de imagens para simular o acesso de usuários a uma estação de trabalho de uso coletivo (e.g., um computador). As imagens foram adquiridas no laboratório de Instrumentação Eletro-Eletrônica da Universidade Federal do Rio Grande do Sul (IEE-UFRGS) e denominado *Banco de Faces do IEE* (BFIEE). O BFIEE é composto por 2500 imagens de 22 pessoas diferentes, divididas em 50 conjuntos compostos de 50 imagens. Cada conjunto representa uma seqüência de vídeo obtida quando um usuário inicia uma sessão na estação de trabalho. As imagens foram adquiridas, com intervalo de 0.5 segundos, por uma *webcam* Genius VídeoCam Messenger 350K, com resolução de 640x480 *pixels*, formato de saída RGB-24 bits e iluminação controlada.

Na Figura 2 é apresentada uma amostra de cada um dos 22 indivíduos do BFIEE. Na Figura 3 são apresentadas as diferentes aparências para um indivíduo do banco, sendo que o número de aparências varia para cada indivíduo. Todos os algoritmos foram desenvolvidos e testados no ambiente Matlab 5.3 e o computador utilizado foi um AtlhonXP +2400, *clock* de 2.0GHz e 1GB de memória RAM.

Figura 2 - Amostras dos indivíduos que compõem o BFIEE.

Figura 3 - Diferentes aparências para um indivíduo do BFIEE.

4.1 VISÃO GERAL DO ALGORITMO

O algoritmo proposto neste trabalho utiliza atributos faciais como informação biométrica para realizar a autenticação dos usuários que utilizam uma estação de trabalho com acesso restrito, bem como gerenciar suas mudanças de aparência facial. Em um primeiro momento, o usuário é registrado no sistema de gerenciamento de identidades digitais (SGID), recebendo um nome de usuário ao qual é associada uma senha e uma representação da sua aparência facial atual, por meio de um modelo de mistura de Gaussianas (MMG)

Quando um usuário registrado acessa a estação de trabalho utilizando sua senha pessoal, um processo de autenticação biométrica é realizado. A autenticação é realizada por meio da comparação entre a aparência atual do usuário e os modelos biométricos presentes na base de identidades digitais (IDDs), sendo que um usuário pode ser associado a mais de um modelo, como será detalhado mais tarde. Para garantir o acesso, o modelo de aparência facial do usuário deve combinar com o armazenado na base de IDDs, confirmando que ele não se trata de um impostor.

Para verificar a IDD do usuário, um classificador Bayesiano, baseado em MMGs das representações faciais é executado. Se pelo menos um dos modelos de aparência do usuário é o que melhor combina (i.e., é o mais parecido entre todos os modelos da base de IDDs) sua identidade é autenticada. Caso contrário, um supervisor é acionado para verificar a identidade do usuário, sendo que:

1. *Se* a identidade do usuário é confirmada, sinalizando uma mudança de aparência, o modelo de aparência facial do usuário armazenado é atualizado;
2. *Senão,* o acesso é rejeitado e um provável impostor é detectado.

A seqüência de passos do método proposto é ilustrada pelo diagrama de blocos na Figura 4.

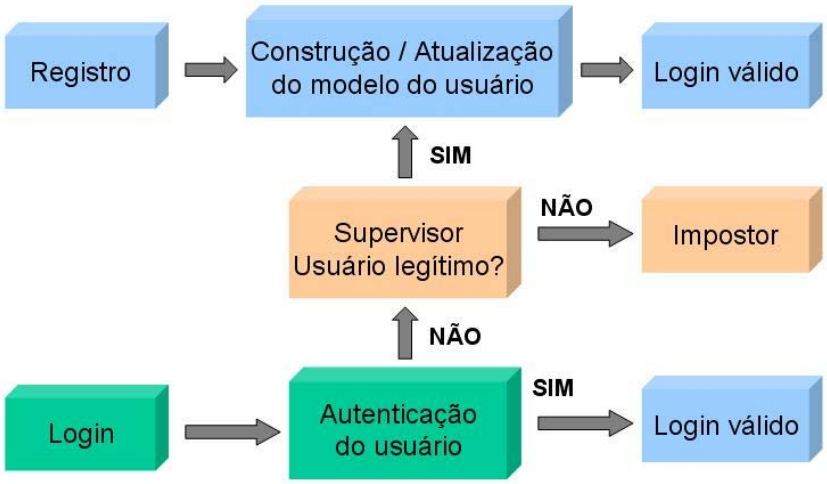


Figura 4 - Diagrama de blocos do algoritmo proposto.

No restante do capítulo são apresentados, de forma detalhada, todos os processos envolvidos na representação da aparência das faces e autenticação dos usuários.

4.2 Extração de Descritores Faciais e Representação da Face

A extração de descritores faciais consiste em um estágio de pré-processamento, realizado durante o processo de autenticação da identidade do usuário. Para cada imagem adquirida são realizados os seguintes procedimentos: (a) a região da face e a localização dos olhos são detectadas; (b) a região da face detectada é verificada e, se a face é confirmada, ocorre a normalização em termos de translação, escala e rotação; (c) feições faciais são selecionadas e atributos estatísticos são extraídos delas e; (d) a aparência facial é representada por um vetor de feições.

Uma imagem é considerada válida (i.e, contém a face de um usuário), se todas as etapas acima descritas são executadas com sucesso. Na seqüência desta seção, cada um dos procedimentos citados é apresentado em detalhes.

4.2.1 Detecção da face e localização dos olhos

A tarefa inicial na análise da imagem é a localização de regiões com grande probabilidade de ser faces humanas. Conforme apresentado na revisão de literatura, várias abordagens têm sido propostas para tratar este problema, sendo que foi desenvolvido um detector de faces que combina a análise de baixo nível (cores, bordas) com feições baseadas

no conhecimento prévio sobre a geometria da face. Estes atributos são escolhidos porque tendem a ser robustos para localização de faces com variabilidade de escala, rotação e expressão.

Considerando que, neste trabalho, o ambiente no qual as imagens foram obtidas é razoavelmente controlado com respeito à iluminação, escala e cenário, tentou-se implementar o detector de faces mais simples possível. As etapas do algoritmo de detecção são: (a) segmentação da imagem baseada em tons de pele; (b) criação da máscara da face; (c) busca por *blobs*; (d) localização dos olhos; e (e) verificação da face.

Algumas destas etapas são executadas sobre a imagem em tamanho original (640x480 *pixels*), enquanto outras utilizam uma versão reduzida da imagem para diminuir o tempo de processamento. No restante do trabalho, a imagem original é representada por I, sendo rI a representação de sua versão reduzida.

4.2.1.1 Segmentação baseada em tons de pele

Para reduzir o espaço de busca por faces em imagens coloridas, uma técnica amplamente utilizada é a segmentação baseada em tons de pele. Além de proporcionar a redução do tempo de processamento nas etapas seguintes, este procedimento ajuda a eliminar estruturas que poderiam ser confundidas com faces em imagens em tons de cinza. Para tanto, foi implementado o detector de tons de pele proposto por Falipou (2006), cujas regras de classificação dos *pixels*, apresentadas na equação 3.1, torna a utilização deste detector bastante atrativa devido à simplicidade e rapidez do processo. A Figura 5 ilustra o resultado típico da sua aplicação sobre a imagem rI, sendo que o resultado obtido é a imagem binária denominada $SkinMap$, onde *pixels* com valor "1" estão localizados em regiões com tom de pele.

(a) (b)

Figura 5 - Resultado típico do detector de tons de pele: (a) Imagem reduzida rI; (b) *Pixels* em rI classificados como pele.

Devido ao grande número de tons de pele possíveis, é difícil avaliar o desempenho do detector. Se limites mais rígidos são impostos na equação 3.1, a faixa de tons de pele detectados torna-se mais restrita. Por outro lado, se os limites são menos rígidos, os resultados obtidos são mais suscetíveis a ruído (i.e., *pixels* não pertencentes à classe "pele").

4.2.1.2 Construção da máscara da face

A abordagem utilizada neste trabalho tem por objetivo preservar somente regiões com tons de pele, descartando regiões espúrias. Portanto, é construída uma máscara da face, denominada *FaceMask* , que determina a região da imagem que contém a face completa, já que, freqüentemente, outras áreas da imagem são cobertas por tons de pele (e.g. pescoço e cabelo). Deve ser observado, ainda, que em algumas situações (e.g., uma pessoa usando barba) a área da face coberta por tons de pele se torna bastante reduzida.

Sendo esta tarefa essencial para a correta localização da face, várias abordagens foram testadas para selecionar a provável região de face. Algumas das tentativas foram: (a) busca de elipses via transformada de Hough, (b) operações recursivas de agrupamento e filtragem de regiões de pele e; (c) análise de componentes conectados. Nenhuma dessas abordagens mostrou-se suficientemente genérica para acomodar todas as situações testadas e, em alguns casos, o tempo consumido para a seleção da região da face foi proibitivo.

Considerando o foco deste trabalho em imagens obtidas próximas ao dispositivo de captura (i.e., um computador com *webcam*), é esperado que somente uma face esteja presente em cada imagem e que esta face corresponda a maior região de pele encontrada. Sendo assim, uma estratégia bastante simples de agrupamento e filtragem foi adotada. A fragmentação das regiões de pele encontradas é minimizada por meio da aplicação de operadores morfológicos (Equação 4.1), gerando regiões de pele conectadas e removendo buracos e defeitos nas bordas da máscara (ver Figura 7(b)). Após estas operações, a maior região de pele é considerada a candidata mais provável para conter uma face.

$$\begin{aligned} &1)\quad FaceMask = SkinMap \oplus SE \\ &2)\quad FaceMask' = fill(FaceMask) \\ &3)\quad FaceMask'' = FaceMask' \Theta SE \end{aligned} \tag{4.1}$$

onde, $FaceMask''$ é a máscara da face obtida, $\oplus$ e Θ denotam as operações de dilatação e erosão morfológicas, respectivamente, SE é um elemento estruturante circular com raio de 5 *pixels*, e **fill** denota a operação de preenchimento de buracos na imagem[4].

4.2.1.3 Localização dos olhos na face detectada

Utilizando-se do conhecimento prévio da estrutura da face humana, algumas características são utilizadas para localizar os olhos na máscara da face. Como exemplo, considere uma face frontal. Sabe-se que os olhos estão localizados na parte superior da face e as pupilas não apresentam tons de pele. Portanto, somente a parte superior da máscara da face, denominada $FaceMask'''$, é mantida para as próximas etapas da busca pelos olhos (Equação 4.2) e conjuntos de *pixels* conectados sem tons de pele (i.e., *blobs*) são procurados em $FaceMask'''$ usando a equação 4.3. Resultados típicos para estas duas operações são apresentados nas Figuras 6(c) e 6(d), respectivamente.

$$FaceMask''' = \left\{ FaceMask''(x,y) \mid y \in \left[y_{\min} + \frac{y_{\max} - y_{\min}}{2}, y_{\max} \right] \right\} \qquad (4.2)$$

onde, $y_{\min}$ e $y_{\max}$ são as coordenadas y mínima e máxima dos *pixels* em $FaceMask'''$.

$$Blobs = FaceMask''' \wedge \overline{SkinMap} \qquad (4.3)$$

onde, '$\wedge$' denota a operação lógica *and*, e $\overline{SkinMap}$ é o complemento lógico de $SkinMap$.

Agora, as maiores estruturas (i.e., os maiores conjuntos de *pixels*) restantes em $Blobs$ devem corresponder aos olhos e sobrancelhas. Contudo, *blobs* espúrios associados com fragmentos de cabelo e cenário podem aparecer e mais processamento é necessário para removê-los. Neste trabalho, para uma face ser válida, seu ângulo máximo de rotação deve ser $\alpha \leq 10^{o}$ (i.e., os olhos estão quase alinhados horizontalmente). As faces detectadas que são severamente contaminadas por *blobs* espúrios acima da região das sobrancelhas tendem a ser descartadas no estágio de verificação porque a localização dos olhos não é confiável, ou seja, o ângulo de rotação estimado freqüentemente excede 10^{o}. Todavia, uma etapa de processamento específica é utilizada para remover *blobs* espúrios localizados próximo a linha imaginária que une os centros dos olhos. Restrições geométricas tendem a ser ineficazes na

[4] A linha poligonal que melhor se ajusta aos limites da máscara da face é usada para delimitar a região a ser preenchida.

eliminação destes *blobs* e um filtro de projeções PF é usado, conforme a equação 4.4. São usadas as projeções verticais para possibilitar a separação dos olhos, que são verificados como sendo 2 picos no histograma de projeções. O mesmo não ocorre quando se utiliza a projeção horizontal (i.e., nas linhas).

Assumindo que $Blobs(x, y) = 1$, se o *pixel* está em um *blob* detectado:

$$PF(x) = \sum_{y=y_{\min}}^{y=y_{\max}} Blobs(x, y) \tag{4.4}$$

onde, $[x_{\min} : x_{\max}, y_{\min} : y_{\max}]$ é o envelope que contém o conjunto de *blobs* detectados.

As colunas com menor contagem relativa de *pixels* são descartadas de $Blobs$, ou seja, as colunas onde $PF(x) < \frac{\max(PF(w))}{2}$, $\forall w \in [x_{\min}, x_{\max}]$. Este procedimento de filtragem das projeções normalmente elimina regiões espúrias que não são os olhos, conforme ilustrado na Figura 6(e).

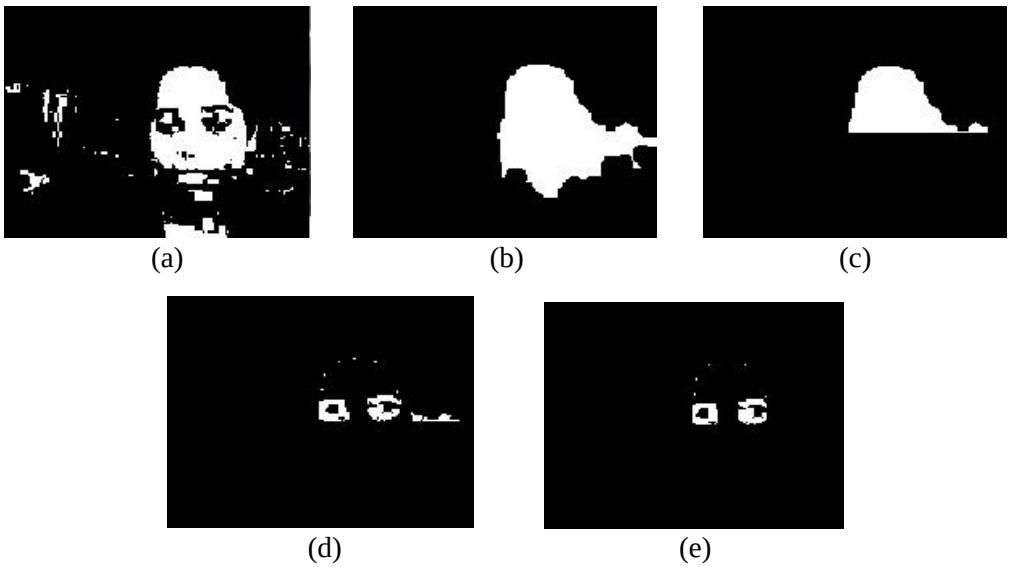


Figura 6 - Ilustração do procedimento de busca por *blobs*: (a) Mapa de tons de pele; (b) Máscara da face; (c) Parte superior da máscara da face; (d) *Blobs* detectados; (e) *Blobs* restantes após a filtragem.

O processo de localização dos olhos é realizado se pelo menos dois componentes conectados são encontrados em $Blobs$. Caso isso não ocorra (e.g., piscar de olhos), uma nova imagem é obtida e o processo é reiniciado. O algoritmo de localização é similar ao proposto por Hsu, Abdel-Mottaleb e Jain (2002), porém algumas mudanças foram introduzidas a fim de adequar o método às necessidades deste trabalho. A metodologia proposta utiliza dois estágios

para localizar os olhos nas faces. Em um primeiro momento é realizada uma localização aproximada, na versão reduzida da imagem rI e, posteriormente, esta localização é refinada utilizando a imagem no tamanho original I.

Para obter a localização aproximada dos olhos, um mapa de olhos é construído, utilizando as componentes cromáticas e acromática do espaço de cores YC_bC_r, conforme apresentado nas equações 3.3-3.5. No entanto, o mapa de luminância proposto por Hsu, Abdel-Mottaleb e Jain (2002) não se mostrou confiável nos experimentos realizados neste trabalho e, portanto, optou-se por uma abordagem diferente para tratar a luminância. Sabendo-se que as pupilas tendem a ser as estruturas mais escuras da face, *pixels* do mapa de olhos são ponderados de acordo com a falta de brilho (i.e., $255 - l$, onde l é a componente luminância da imagem) da região onde eles estão localizados. Sendo assim, as regiões das pupilas recebem pesos mais altos.

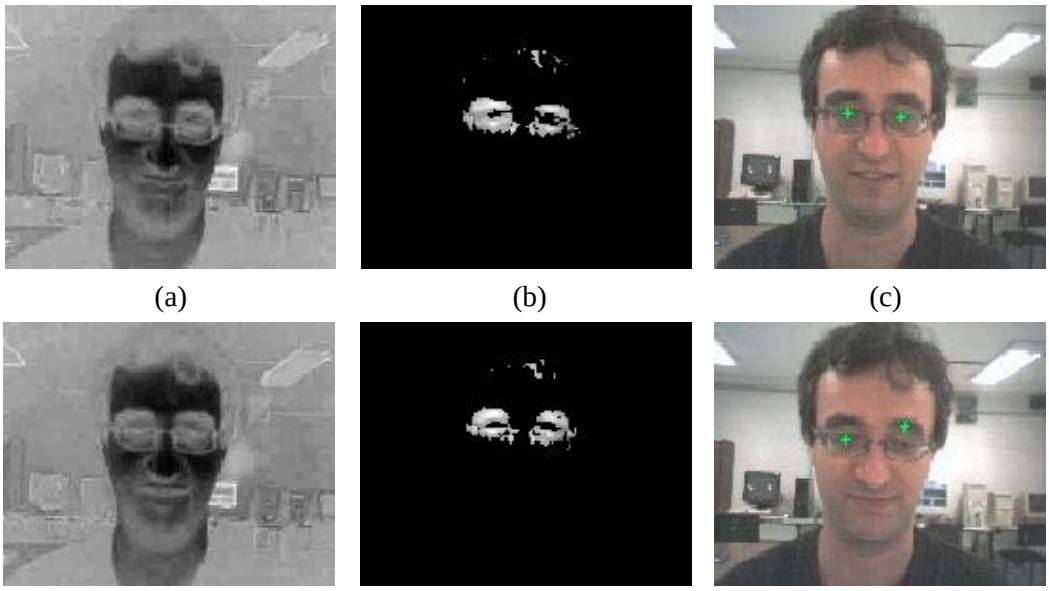
Lembrando que os olhos são localizados de forma aproximada na imagem reduzida rI por operações realizadas *pixel* a *pixel*, utilizando a equação 4.5, onde: (a) as componentes C_b e C_r são as componentes cromáticas azul e vermelha de rI, respectivamente; (b) $LowI$ é a versão em tons de cinza de rI e; (c) $\{Blobs\}$ é um conjunto de componentes conectadas representadas por uma matriz binária **Blb** sendo o valor "1" atribuído às componentes conectadas. O algoritmo para localização aproximada dos olhos é descrito $\forall(x,y) \in rI$ por:

$$\begin{aligned} 1)\quad & EyeMap(x,y) = \frac{1}{3}\cdot\left[C_b^2(x,y) + \left(\widetilde{C}_r(x,y)\right)^2 + \frac{C_b(x,y)}{C_r(x,y)}\right] \\ 2)\quad & EyeMap'(x,y) = EyeMap(x,y)\cdot[255 - LowI(x,y)] \\ 3)\quad & EyeMap''(x,y) = EyeMap'(x,y)\cdot Blb(x,y) \end{aligned} \tag{4.5}$$

onde, $C_b^2(x,y)$, $\left(\widetilde{C}_r(x,y)\right)^2$ e $\frac{C_b(x,y)}{C_r(x,y)}$ são normalizados para a faixa $[0,255]$ e $\widetilde{C}_r(x,y)$ é o negativo de $C_r(x,y)$ (i.e., $\widetilde{C}_r(x,y) = 255 - C_r(x,y)$.

Os dois máximos locais de $EyeMap''$, denominados $(x_{left\ eye}, y_{left\ eye})$ e $(x_{right\ eye}, y_{right\ eye})$, fornecem a localização aproximada dos olhos. A Figura 7 ilustra dois resultados típicos da localização usando este método, onde o exemplo relacionado às localizações, correta e incorreta, dos olhos são apresentados nas Figuras 7(a)-(c) e Figuras 7(d)-(f), respectivamente. Nos casos ilustrados, a localização incorreta ocorre porque

sobrancelhas muito escuras apresentam componentes cromáticas e acromáticas similares às da pupila.



(a) (b) (c)

(c) (e) (f)

Figura 7 - Exemplos de localização aproximada dos olhos: Linha superior: resultados corretos; Linha inferior: resultados incorretos. (a) e (d) *EyeMap*; (b) e (e) *EyeMap*''' obtido; (c) e (f) Localização dos olhos correta e incorreta, respectivamente.

Quando os olhos são incorretamente localizados (ver Figura 7(f)), um processo de refinamento é executado. Uma região quadrada R_{eye}^{rI} em torno da localização aproximada dos olhos, definida na equação 4.6, é selecionada na imagem reduzida rI. As regiões correspondentes na imagem original I, R_{eye}^{I}, são localizadas por meio da transformação de coordenadas apresentada na equação 4.7:

$$1)\quad R_{eye}^{rI} = rI\left(x_{\min} : x_{\max}, y_{\min}, y_{\max}\right) \tag{4.6}$$

$$2)\quad R_{eye}^{I} = I\left(\frac{x_{\min}}{r_f} : \frac{x_{\max}}{r_f}, \frac{y_{\min}}{r_f}, \frac{y_{\max}}{r_f}\right) \tag{4.7}$$

onde, $x_{\min} = x_p - D_{offset}$, $x_{\max} = x_p + D_{offset}$, $y_{\min} = y_p - D_{offset}$, $y_{\max} = y_p + D_{offset}$, $\left(x_p, y_p\right) \in \left\{\left(x_{left\ eye}, y_{left\ eye}\right), \left(x_{right\ eye}, y_{right\ eye}\right)\right\}$, D_{offset} é a dimensão do lado da região quadrada em torno da localização aproximada dos olhos e $r_f < 1$ é o fator de escala usado para obter rI.

Após selecionar a região R_{eye}^{I} em torno de cada olho na versão em tons de cinza da imagem I, três mapas são calculados: um mapa de localização de bordas eR_{eye}^{I}, que é obtido com o detector de Canny (1986); e dois mapas direcionais de gradientes, denominados $G_x R_{eye}^{I}$ e $G_y R_{eye}^{I}$, ao longo das direções x e y, respectivamente, obtidos usando os operadores de Prewitt (GONZALEZ; WOODS, 1992). As Figuras 8(b) e 8(c) ilustram o mapa de bordas de Canny e o mapa de magnitudes do gradiente, respectivamente. Estes mapas e a transformada de Hough (TH) são utilizados para detectar os centros das circunferências das pupilas ou íris seguindo a metodologia apresentada na fundamentação teórica.

No entanto, quando as regiões são ruidosas ou pouco definidas (e.g., quando os olhos estão semi-abertos), os máximos dos acumuladores de Hough ($accR_{eye}^{I}$) podem levar a estimativas incorretas do centro dos olhos. Portanto, mais restrições baseadas no conhecimento destas estruturas são impostas para aumentar a robustez da localização. As pupilas não possuem tons de pele e, de fato, são regiões escuras cercadas por regiões mais claras. Sendo assim, são descartados os *pixels* em R_{eye}^{I} com tons de pele ($RSkinMap$) e os que possuem intensidade mais alta que o limiar $\kappa \cdot \min\left(R_{eye}^{I}\right)$, onde $\kappa = 1{,}1$, foi determinado experimentalmente (veja as equações 4.8 e 4.9). Desta forma, um espaço de parâmetros de Hough limitado, $CaccR_{eye}^{I}$, é calculado como mostra a equação 4.10, preservando somente as posições dos acumuladores de Hough que tem uma localização válida em $RLowMap'$ (veja a equação 4.9).

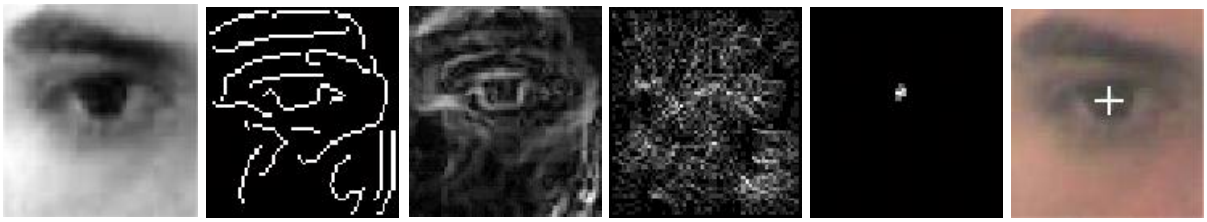
$$1)\quad RLowMap = R_{eye}^{I} < 1{,}1 \cdot \min\left(R_{eye}^{I}\right) \tag{4.8}$$

$$2)\quad RLowMap' = RLowMap \wedge \overline{RSkinMap} \tag{4.9}$$

$$3)\quad CaccR_{eye}^{I}(x,y) = RLowMap'(x,y) \cdot accR_{eye}^{I}(x,y) \tag{4.10}$$

onde, $\overline{RSkinMap}$ é o complemento lógico de $RSkinMap$.

Os *votos* atribuídos ao centro de círculo em (x,y), $accR_{eye}^{I}(x,y)$, são considerados somente se $RLowMap'(x,y) = 1$. Uma ilustração do efeito destas restrições na transformada de Hough para encontrar círculos é apresentada nas Figuras 8(d) e 8(e). O centro do olho é detectado como o centro do círculo associado ao máximo do espaço de parâmetros de Hough limitado, $CaccR_{eye}^{I}$, conforme apresentado nas Figuras 8(e) e 8(f).



(a) (b) (c) (d) (e) (f)

Figura 8 - Ilustração do refinamento da localização do centro do olho: (a) Região do olho em tons de cinza R^I_{eye}; (b) Mapa de bordas eR^I_{eye}; (c) Magnitudes do gradiente; (d) Acumulador de Hough $accR^I_{eye}$; (e) Acumulador de Hough limitado $CaccR^I_{eye}$; (f) Localização final do centro do olho.

4.2.1.4 Verificação da Face

Neste estágio do método proposto, um candidato à face contendo os dois prováveis olhos foi selecionado. Para confirmar que a face é válida, dois critérios básicos devem ser satisfeitos: (a) o ângulo de rotação da face α deve ser menor que 10^o e; (b) a distância entre o centro dos olhos $\left\|\overrightarrow{dEE}\right\|$ deve ser de, pelo menos, 64 *pixels*.

Os parâmetros α e $\left\|\overrightarrow{dEE}\right\|$ foram determinados experimentalmente, após análise do banco de faces. Como foi adotada uma condição de aquisição específica (i.e., usuário em frente a uma estação de trabalho), a escala e a rotação da cabeça tendem a ter variabilidade limitada. A imposição de restrições para a rotação da face e distância entre olhos garante que as feições faciais sejam *visíveis* e tenham resolução adequada. Portanto, as restrições são usadas para aumentar a robustez do detector de faces.

Contudo, mesmo que os olhos detectados satisfaçam as condições impostas acima, erros ainda podem ocorrer principalmente porque as cores de pele, cabelo e cenário podem ser confundidas, aumentando indevidamente a área da máscara da face e fazendo com que *blobs* espúrios apareçam. Portanto, uma terceira condição de verificação é imposta. Com base na estimativa das coordenadas dos centros dos olhos, um triângulo eqüilátero, chamado aqui de triângulo de feições, é construído. Dois dos vértices do triângulo são os centros dos olhos, sendo o terceiro vértice dado pela boca hipotética, calculada conforme a equação 4.11[5]. Se a localização dos olhos está correta, a proporção de *pixels* com tom de pele dentro do triângulo de feições (i.e., a razão entre o número de *pixels* com tom de pele e o número total de *pixels*

[5] Foi verificado experimentalmente, por meio da análise de 50 imagens de faces frontais, que a razão entre a distância entre olhos e a distância entre o ponto médio dos olhos e a boca é, aproximadamente, 1.

na área do triângulo) será alta, mesmo nos casos em que o usuário usa óculos, barda ou bigode.

$$\begin{aligned} Mouth(y) &= EC_y + \left\|\overrightarrow{dEE}\right\| \cdot sen(\alpha) \\ Mouth(x) &= EC_x + \left\|\overrightarrow{dEE}\right\| \cdot \cos(\alpha) \end{aligned} \tag{4.11}$$

onde, (EC_x, EC_y) é o ponto médio entre os olhos, $\left\|\overrightarrow{dEE}\right\|$ é o comprimento da linha $\overrightarrow{dEE}$ que une os centros dos olhos, $(x_{left\ eye\ center}, y_{left\ eye\ center})$ e $(x_{right\ eye\ center}, y_{right\ eye\ center})$, e o ângulo de rotação da face α é o ângulo que $\overrightarrow{dEE}$ faz com o eixo horizontal :

$$\alpha = \tan^{-1} \frac{(y_{left\ eye\ center} - y_{right\ eye\ center})}{(x_{left\ eye\ center} - x_{right\ eye\ center})} \tag{4.12}$$

Caso contrário, se os olhos são localizados incorretamente, a proporção de *pixels* com tom de pele dentro do triângulo tende a cair. Isto ocorre porque, quando os olhos são erroneamente localizados, pelo menos um deles está situado fora da face, sendo associado ao cenário. Foi definido experimentalmente que a proporção mínima de *pixels* com tom de pele para validar a face é de 70%. Se uma face candidata não passa em todas as verificações ela é descartada e uma nova imagem é obtida. Alguns exemplos de triângulos de feições para faces válidas são apresentados na Figura 9.

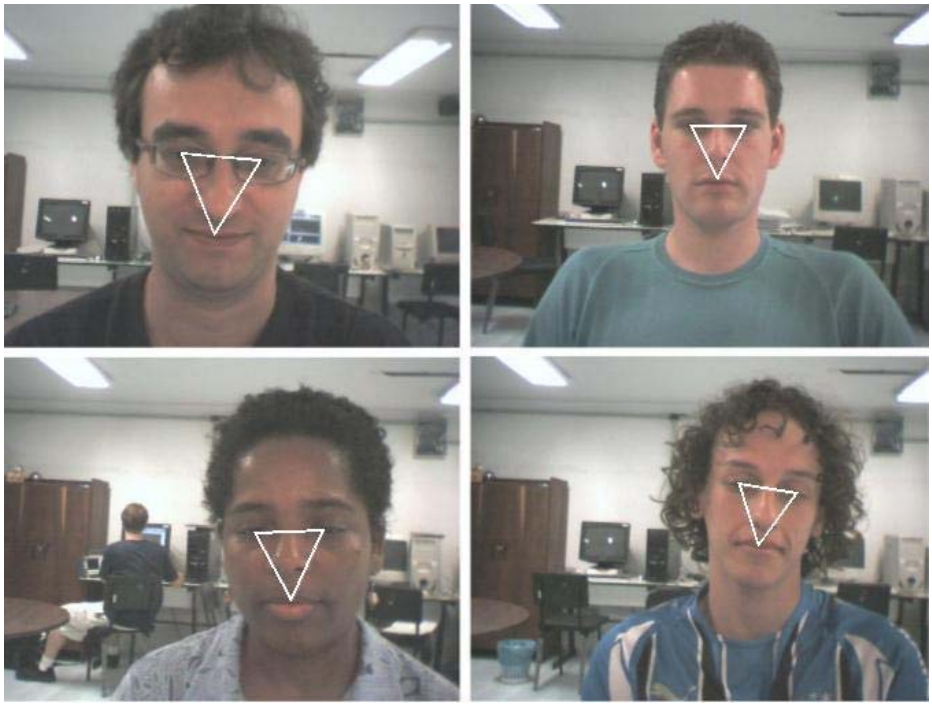


Figura 9 - Exemplos de triângulos para faces detectadas corretamente.

4.2.2 Normalização da face

Se uma face candidata passa em todos os testes especificados na seção anterior, ela é considerada válida. Neste caso, uma região de interesse F, que contém a face, é extraída da imagem original I como mostra a Figura 10(a). O tamanho de F é calculado com respeito à $\left\|\overrightarrow{dEE}\right\|$ e ao ponto médio entre os olhos $\left(EC_x, EC_y\right)$, conforme a equação 4.13:

$$F = \left\{ I(x,y) \mid x \in \left[EC_x - \vartheta_x \cdot \left\|\overrightarrow{dEE}\right\|,\ EC_x + \vartheta_x \cdot \left\|\overrightarrow{dEE}\right\| \right], \right. \\ \left. y \in \left[EC_x - \vartheta_y \cdot \left\|\overrightarrow{dEE}\right\|,\ EC_y + \vartheta_y \cdot \left\|\overrightarrow{dEE}\right\| \right] \right\} \tag{4.13}$$

onde, $\vartheta_x = 2$ e $\vartheta_y = 1{,}5$ em todos os experimentos.

A normalização é importante para assegurar que se possam extrair descritores comparáveis para todas as faces, os quais são usados no processo de autenticação do usuário. O procedimento de normalização consiste em levar todas as faces para as mesmas condições de translação, escala e rotação. Inicialmente, uma transformação de escala é realizada utilizando a matriz de escalonamento S:

$$S = \begin{bmatrix} \frac{ndEE}{\left\|\overrightarrow{dEE}\right\|} & 0 \\ 0 & \frac{ndEE}{\left\|\overrightarrow{dEE}\right\|} \end{bmatrix} \tag{4.14}$$

onde, $\left\|\overrightarrow{dEE}\right\|$ é a distância entre olhos atual e $ndEE$ é a distância entre olhos após a normalização. Esta distância normalizada é arbitrada como sendo a mínima distância entre olhos necessária para ocorrer a validação da face, ou seja, $ndEE = 64$ *pixels*.

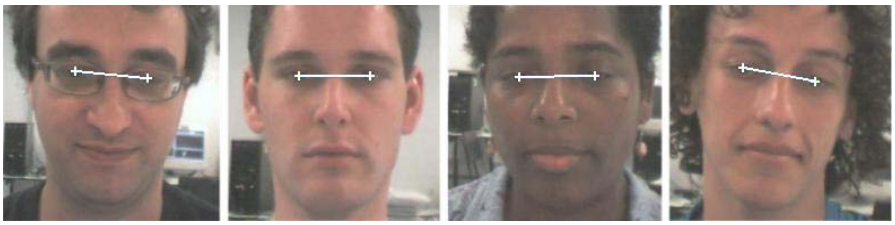
O segundo passo no processo de normalização é a aplicação da transformação de rotação R dada por:

$$R = \begin{bmatrix} \cos(-\alpha) & sen(-\alpha) \\ -sen(-\alpha) & \cos(-\alpha) \end{bmatrix} \tag{4.15}$$

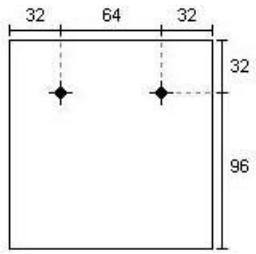
onde a correção da rotação é realizada com negativo do ângulo de rotação da face α.

Portanto, a face normalizada é $F = T(R(S(F)))$ e estas transformações geométricas são implementadas utilizando mapeamento inverso e interpolação bi-linear. Após cada etapa da normalização, as coordenadas dos olhos são recuperadas utilizando mapeamento direto, e as matrizes de transformação inversas S^{-1} e R^{-1}.

Finalmente, uma região normalizada da face F' é extraída obedecendo o molde de 128×128 *pixels*, apresentado na Figura 10(b). Este último passo completa o processo de normalização da face. Alguns exemplos de faces normalizadas são apresentados na Figura 10(c). A linha que conecta os olhos na Figura 10(a) somente indica visualmente o grau de rotação da face.



(a)



(b)

(c)

Figura 10 - Processo de normalização: (a) Faces detectadas; (b) Molde para normalização; (c) Faces normalizadas.

4.2.3 Obtenção de Descritores Faciais

Nesta seção são apresentados os critérios adotados para a escolha das feições faciais. Ainda, é descrita a metodologia utilizada para a obtenção de atributos de textura para discriminar as faces.

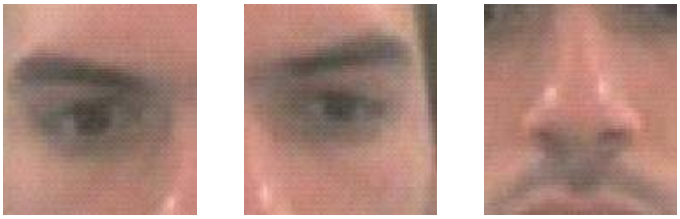
4.2.3.1 Seleção das feições

A escolha das feições sobre as quais serão extraídos os descritores deve ser cuidadosa. Para se ter um bom conjunto de discriminantes para as faces, as feições faciais devem obedecer as seguintes condições básicas: (a) estar presente em todas as faces; (a) permanecer

estáveis com o passar do tempo; (c) possuir boa capacidade de discriminação; (d) apresentar pequena variação elástica (quando da variação de expressão facial, por exemplo).

Dessa forma, neste trabalho, os olhos e o nariz foram as estruturas que se apresentaram mais adequadas para diferenciar faces, já que tenderam a ser menos variantes com a mudança de expressão, estiveram presentes, teoricamente, em todas as faces e sua localização foi relativamente fácil. Como contra-exemplo, a boca, que é uma estrutura marcante da face e possui alto poder de diferenciação, se mostrou extremamente elástica, podendo se deformar de tal maneira que se torna difícil comparar expressões diferentes até do mesmo indivíduo. Além disso, a boca é uma feição que pode sofrer variações temporais bruscas na cor e textura devido a maquiagem. Ainda, cicatrizes e manchas na pele (e.g., sardas) podem ser poderosos discriminantes entre faces, porém não são encontradas em todas elas. Por fim, outro fator que recomenda a utilização dos olhos e do nariz é que, em pessoas utilizando barba, a aparência global é bastante diferente da sua versão sem barba, contudo estas regiões não são afetadas.

Neste trabalho, portanto, a representação e discriminação das aparências dos usuários são baseadas em descritores estatísticos de textura das regiões dos dois olhos e do nariz, individualmente, conforme apresentado na Figura 11. Considerando que todas as faces são normalizadas, as regiões selecionadas são encontradas sob as mesmas condições de localização em, virtualmente, todas as imagens. Portanto, as diferenças entre elas são devidas às diferentes aparências dos usuários e não a aspectos como rotação e escala, por exemplo.



(a) (b) (c)

Figura 11 - Feições faciais selecionadas: (a) Olho esquerdo; (b) Olho direito; (c) Nariz.

4.2.3.2 Extração de atributos

De acordo com os trabalhos de Drimbarean e Whelan (2001) e Randen e Husoy (1999), a utilização de coeficientes da transformada discreta de co-senos (TDC) tem maior poder de discriminação de texturas que outros descritores comumente usados, como matrizes de co-ocorrência e filtros de Gabor. Além disso, são muito mais rápidas para calcular.

Sendo assim, neste trabalho, as texturas são representadas por descritores baseados em coeficientes TDC. Em cada feição facial (os dois olhos e o nariz), um conjunto de 9 máscaras TDC-2D (veja a equação 3.21) é aplicado sobre as componentes de cor RGB, separadamente, e a variância de cada matriz de coeficientes resultante é calculada por:

$$\begin{aligned} CM_i^C &= m_i \otimes FR_C \\ f_i^C &= \sigma^2\left(CM_i^C\right) \end{aligned} \tag{4.16}$$

onde, $i = 1,2,\text{K},9$, $C \in \{R,G,B\}$, e CM_i^C é a matriz de coeficientes resultante da convolução (denotada por $\otimes$) da máscara DCT m_i com a componente de cor C da feição facial FR, denominada FR_C.

Portanto, o vetor de energias resultante é dado por:

$$\overrightarrow{EV} = \left[f_1^R, f_1^G, f_1^B, f_2^R, \text{K}, f_9^B\right] \tag{4.17}$$

onde, $\overrightarrow{EV}$ denota vetor de energias 27-dimensional (27-D), e os elementos do vetor f_i^C são os desvios padrão elevados ao quadrado $\sigma^2\left(CM_i^C\right)$.

Lembrando que um vetor de energias 27-D $\overrightarrow{EV}$ é obtido para cada uma das 3 feições e todo $f_i^C = \sigma^2\left(CM_i^C\right)$, e $f_i^C \geq 0$. Desta forma, um vetor de atributos 3-D é obtido por meio do cálculo da magnitude dos vetores de energia 27-D. A norma Euclidiana quadrada de cada vetor de energias 27-D é calculada pela simples soma de todos os 27 elementos $EV(i)$:

$$\overrightarrow{FV} = \left[\sum_{i=1}^{27} EV_{le}(i), \sum_{i=1}^{27} EV_{re}(i), \sum_{i=1}^{27} EV_n(i)\right] \tag{4.18}$$

onde, *le* se refere ao olho esquerdo, *re* ao olho direito, e *n* à região do nariz[6].

Finalmente, cada face válida possui um vetor de atributos 3-D $\overrightarrow{FV}$ associado, provendo o espaço de feições 3-D usado para representar e discriminar faces de usuários, como será detalhado na próxima seção.

---

[6] Foram utilizados os dois olhos separadamente porque eles podem ser bastante diferentes no mesmo indivíduo. Isto ocorre devido, principalmente, à diferenças de cor da íris, modificações causadas por acidentes ou doenças.

4.3 AUTENTICAÇÃO MULTI-MODAL DE USUÁRIOS

A metodologia de gerenciamento de IDDs proposta é apresentada no restante do capítulo seguindo a ordem: (a) descrição do processo de registro dos usuários; (b) descrição do método para autenticação dos usuários; (c) descrição dos métodos utilizados para atualização dos modelos; e (d) apresentação de um exemplo ilustrativo.

Para evitar confusão entre os termos *usuário*, *aparência*, *classe* e *subclasse*, na seqüência do texto, os termos *usuário* e *aparência* são usados quando se está referindo ao contexto do mundo real. Quando a referência são os algoritmos no mundo digital, estes termos são substituídos por *classe* e *subclasse*, respectivamente.

4.3.1 Registro de novos usuários

O primeiro passo em um SGID é a inclusão de novos usuários na base de IDDs. Esta tarefa é chamada aqui de registro e é responsável pela criação de uma nova classe, associada ao novo usuário por meio de seu nome de usuário e senha. A classe é definida por um modelo construído a partir de 50 vetores de atributos que descrevem as imagens de face de 50 quadros de vídeo capturados durante o início da sessão. Estes vetores de atributos representam a aparência atual da face do usuário e formam uma nuvem de pontos no espaço de feições 3-D.

Um modelo estatístico é construído para esta nova classe, utilizando um método de aprendizagem não-supervisionado. Na verdade, o algoritmo de modelagem paramétrica utilizando misturas de Gaussianas CMM-MMG, proposto por Figueiredo e Jain (2002), é utilizado. Em outras palavras, uma distribuição complexa de pontos 3-D é *explicada* por uma mistura de M Gaussianas, MMG, para a qual se tem vasto conhecimento disponível:

$$MMG = \sum_{m=1}^{M} p_m \cdot G(\mu_m, \Sigma_m) \tag{4.19}$$

onde, p_m é o peso associado à componente Gaussiana $G(\mu_m, \Sigma_m)$ na mistura.

Com a inserção de usuários na base de IDDs, ocorre também aumento do número de classes e alguma sobreposição de classes no espaço de feições é esperada. Desta forma, um ponto no espaço de feições (i.e., uma aparência do usuário) é atribuído à classe mais provável, que é a classe na qual o MMG que *explica* este ponto tem maior probabilidade entre todas as classes.

Por exemplo, nas Figuras 12(a)-(b) uma distribuição de vetores de atributos típica é apresentada para o primeiro usuário, bem como a projeção 2-D do *cluster* formado pelo MMG da distribuição. As Figuras 12(c)-(d) apresentam o que acontece com a adição de um segundo usuário, e as Figuras 12(e)-(f) mostram os resultados após 10 usuários terem sido inseridos na base de IDDs. Para facilitar a visualização, somente as distâncias para o olho direito e esquerdo dos vetores de atributos foram *plotados* na Figura 12. Como pode ser verificado, a sobreposição de classes tende a aumentar quando mais usuários são registrados.

4.3.2 Autenticação de usuários

Quando um usuário registrado inicia uma nova sessão, ele informa um nome de usuário e uma senha, que são associadas a uma de classe na base de IDDs. Depois de verificadas estas informações, o processo de autenticação da IDD consiste em confirmar a identidade do usuário com base em uma amostra de vídeo S, ou seja, uma seqüência de vídeo com V_f quadros mostrando sua aparência facial atual, sendo que cada quadro gera um $\overrightarrow{FV}$. A probabilidade de esta amostra pertencer a cada uma das classes é calculada, se esperado que a classe mais provável seja aquela associada ao usuário atual.

Então, se o usuário C_j realiza inicia uma sessão, e a base de IDDs contêm J classes, surgem duas possibilidades:

1. *Se* $p(C_i) > p(C_j)$ para todo $j \neq i$, *então* o usuário é autenticado;
2. *Senão,* um supervisor é acionado para confirmar a identidade do usuário.

Neste trabalho, a abordagem Bayesiana é utilizada para calcular as probabilidades das classes e atribuir relevância a elas. Na forma geral, a probabilidade *a posteriori* de C_j, $j = 1,2,\mathrm{K}\ ,J$, é dada pelas equações 4.20-4.22:

$$p(C_j \mid S) = \frac{p(S \mid C_j) \cdot p(C_j)}{p(S)} \tag{4.20}$$

onde, a probabilidade *a priori* da classe C_j é:

$$p(C_j) = \frac{1}{J} \tag{4.21}$$

e a probabilidade *a priori* da amostra S é:

$$p(S) = \sum_{j=1}^{J} p(S \mid C_j) \cdot p(C_j) \tag{4.22}$$

onde, $p(S \mid C_j)$ é a probabilidade *a priori* da amostra S dada a classe C_j.

Para uma amostra genérica S, $p(S \mid C_j)$ é a densidade de probabilidade e, neste trabalho, esta densidade é estimada a partir da construção do MMG para a classe C_j, ou seja, modelando a distribuição da amostra $S = [f_1, f_2, f_3]$ na classe C_j.

Contudo, como mencionado anteriormente, a aparência facial do usuário pode mudar com o tempo. Para acomodar estas modificações, para cada nova aparência, uma nova subclasse é criada dentro da classe do usuário e, dessa forma, o cálculo da probabilidade *a priori* da subclasse deve ser reconsiderado. Neste trabalho, dois tipos de tratamento para as subclasses são propostos a seguir.

4.3.2.1 Subclasses com probabilidades a priori iguais

Lembrando que cada subclasse tem o mesmo número de amostras, ou seja, uma seqüência de vídeo com V_f quadros, então, uma abordagem possível é tratar as subclasses de todos os usuários de forma igualitária, ou seja, todas têm a mesma probabilidade *a priori*, pois todas são construídas a partir do mesmo número de quadros. Neste caso, classes que contêm mais subclasses têm, por conseqüência, maior probabilidade a priori e, desta forma, uma tendência é introduzida no processo de classificação. Para inserir K_{C_j} subclasses dentro de uma dada classe C_j, $j = 1,2,\text{K}\ ,J$, equações 4.21-4.22 são re-escritas como:

$$p(C_j \mid S) = \frac{\sum_{k=1}^{K_{C_j}} p(S \mid C_j^k) \cdot p(C_j^k)}{p(S)} \tag{4.23}$$

onde, as probabilidades *a priori* das subclasses são:

$$p(C_j^k) = \frac{1}{\sum_{j=1}^{J} K_{C_j}} \tag{4.24}$$

e a probabilidade *a priori* da amostra é:

$$p(S) = \sum_{j=1}^{J} \sum_{k=1}^{K_{C_j}} p\left(S \mid C_j^k\right) \cdot p\left(C_j^k\right) \tag{4.25}$$

onde, $p\left(S \mid C_j^k\right)$ é a probabilidade da amostra dada a subclasse k da classe C_j, e a densidade é estimada com base na amostras atribuídas a C_j^k, e um MMG específico para esta subclasse.

4.3.2.2 Classes com probabilidades a priori iguais

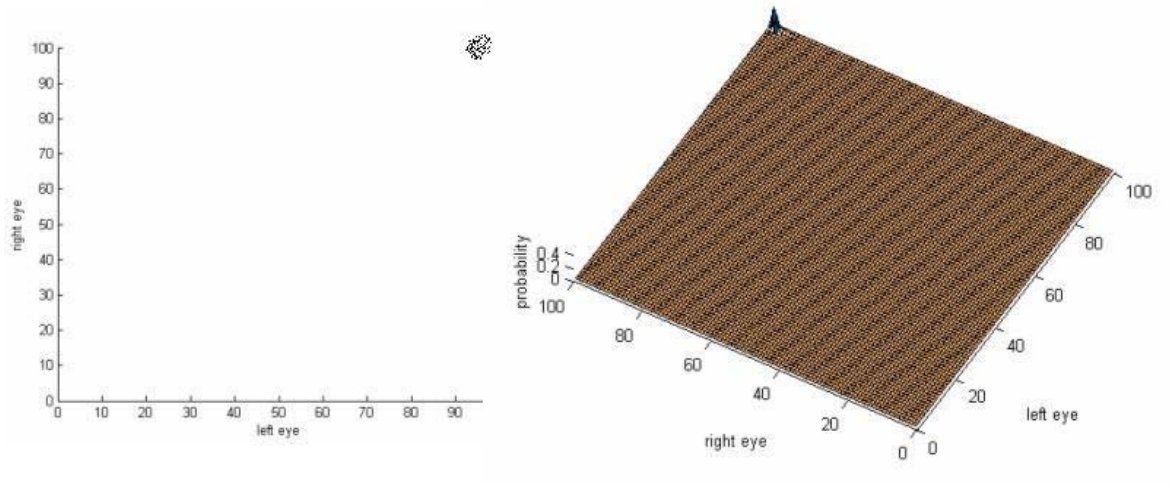
Nesta abordagem a mesma probabilidade *a priori* é atribuída a cada classe, independentemente do número de subclasses que ela contêm. Isto é possível porque as probabilidades *a priori* das subclasses $p\left(C_j^k\right)$ são inversamente proporcionais ao número de classes J e ao número de subclasses K_{C_j} contidas pela classe C_j. Da mesma forma que a abordagem anterior, a probabilidade *a posteriori* $p\left(C_j \mid S\right)$ é dada pela equação 4.23 e a probabilidade *a priori* da amostra, $p(S)$, pela equação 4.25. Contudo, as probabilidades *a priori* das subclasses são calculadas por:

$$p\left(C_j^k\right) = \frac{1}{J \cdot K_{C_j}} \tag{4.26}$$

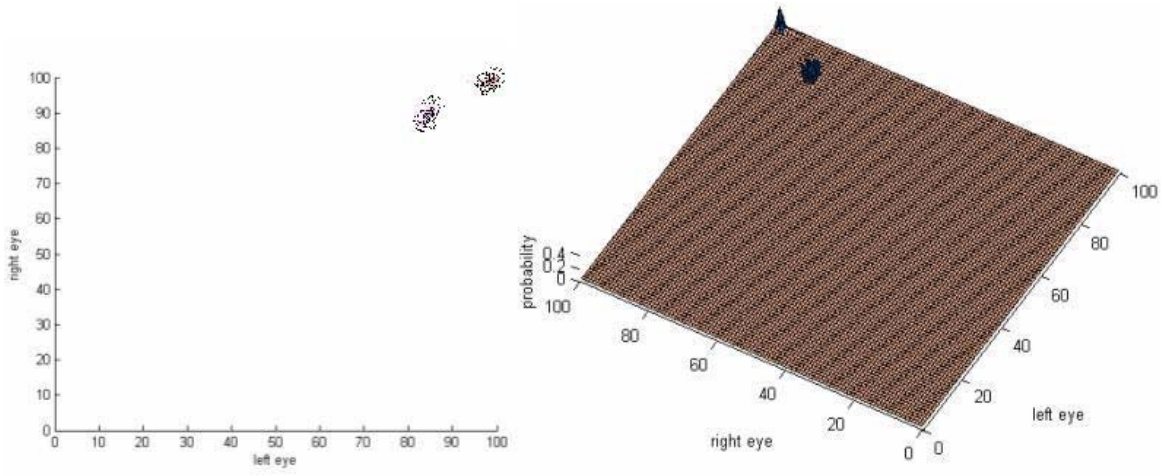
Sendo que todas as classes têm a mesma probabilidade *a priori*, então esta é uma abordagem que não adiciona tendência ao processo.

4.3.3 Atualização do modelo de aparência facial

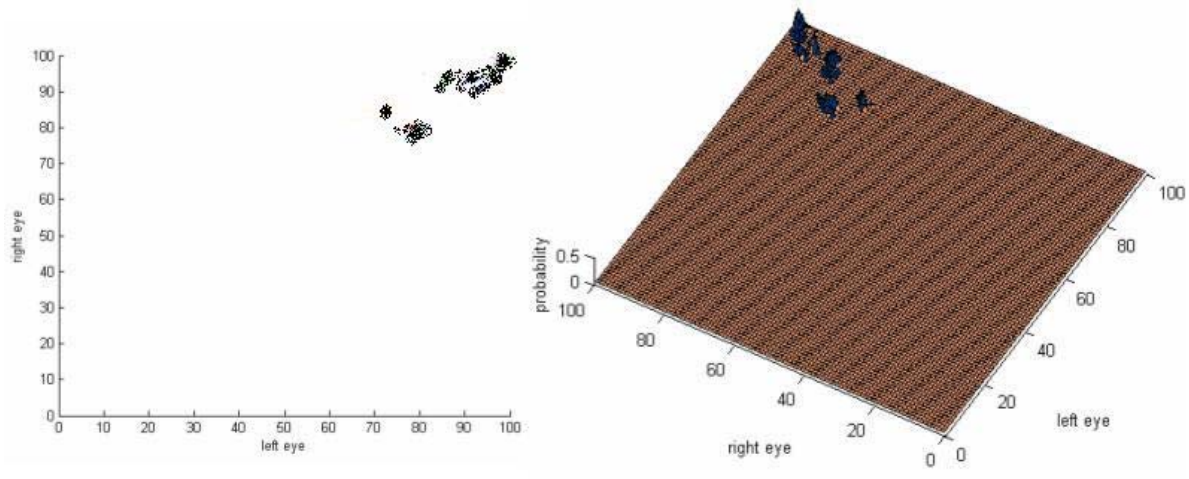
Para fazer com que o SGID esteja apto a *aprender* novas aparências dos usuários é necessário um algoritmo de atualização dos modelos, que atua quando a autenticação falha, mas o supervisor confirma a legitimidade do usuário. A idéia principal é que após a atualização do modelo da classe (e subclasse), o SGID possa autenticar o usuário mesmo com esta nova aparência. Como no processo de autenticação, duas alternativas para realizar a atualização dos modelos são discutidas a seguir.



(a)



(b)



(c)

Figura 12 - Evolução do processo de registro de usuários e formação de classes no espaço de feições direita: $\overrightarrow{FV}$ projetados em 2-D; esquerda: visão 3-D dos modelos das classes):
(a) 1 classe; (b) 2 classes; (c) 10 classes.

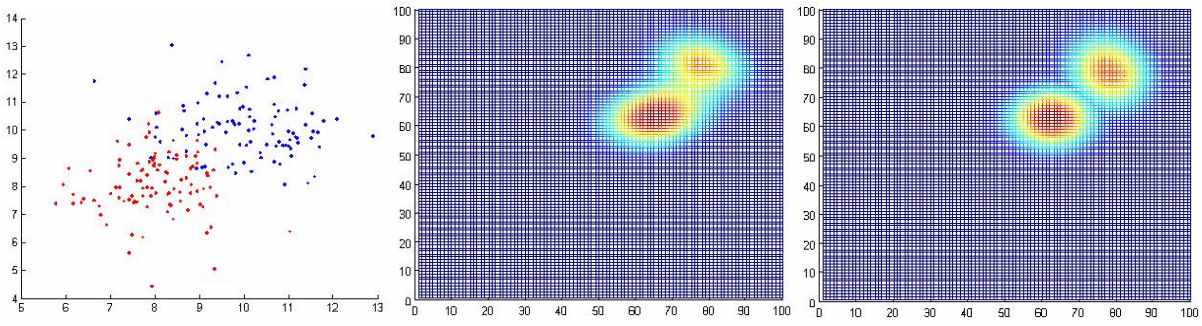
4.3.3.1 Classes com modelo adaptativo único

Segundo esta proposta, as novas amostras de aparência são usadas para atualizar o MMG da classe, gerando um MMG completamente novo. Não existem subclasses e a implementação é bastante simples. O revés é que pode ocorrer facilmente a saturação do espaço de feições, já que um MMG finito é forçado a acomodar novas aparências do usuário. Além disso, as classes podem ser construídas a partir de conjuntos com número de amostras diferentes, podendo incluir tendência na classificação. Conseqüentemente, esta abordagem é propensa à falhas em longo prazo. A situação de uma classe contendo duas subclasses (i.e., duas aparências do usuário) é ilustrada na Figura 13(b), onde a sobreposição das subclasses no espaço de feições fica clara.

4.3.3.2 Classes contendo múltiplos modelos

Outra abordagem trata cada aparência do usuário como uma subclasse independente, que tem o seu próprio modelo. Quando o processo de atualização do modelo da classe é iniciado, V_f quadros são obtidos e os respectivos vetores de atributos são calculados para construir uma nova subclasse e seu modelo. De acordo com esta abordagem, um usuário pode ter tantos modelos de aparência quanto forem necessários para descrever todas as suas aparências, já que esta abordagem tem capacidade de adaptação ilimitada. Esta é uma abordagem mais intuitiva e garante que todos os modelos de subclasses sejam construídos com o mesmo número de amostras e parâmetros do algoritmo $CMM - MMG$.

Em comparação com a abordagem descrita no item anterior, esta produz *clusters* melhor localizados no espaço de feições, levando a menor ocorrência de sobreposição dos modelos conforme ilustrado na Figura 13(c).



(a) (b) (c)

Figura 13 - Representação das classes: (a) nuvem de pontos formada por vetores de atributos 2-D de duas aparências de um mesmo usuário (vermelho: aparência 1; azul: aparência 2); (b) representação usando modelo adaptativo único; (c) representação usando múltiplos modelos.

4.3.4 Um exemplo de autenticação de usuário

Suponha que existam 3 usuários registrados na base de IDDs, denominados usuários 1, 2 e 3, e todos os modelos estejam adaptados às suas aparências conhecidas. Agora, suponha que o usuário 1 passe a usar óculos e tente iniciar uma sessão em uma estação de trabalho onde o sistema está operando. Nesta situação, uma nova aparência é detectada e as probabilidades do usuário pertencer a cada uma das 3 classes são $[0{,}40 \quad 0{,}10 \quad 0{,}50]$, respectivamente. Ocorre falha no acesso porque o nome se usuário e a senha estão associados ao usuário 1, mas a aparência é mais bem ajustada, embora com baixa probabilidade, ao usuário 3.

Então, o supervisor é acionado para confirmar a legitimidade do usuário. Se o supervisor autoriza o acesso, indicando que houve uma mudança de aparência, uma nova subclasse é criada automaticamente, com base nas informações atuais de aparência, e é associada à classe do usuário 1. Na próxima vez que este usuário tenta acessar a estação de trabalho, as probabilidades para as 3 classes passam a ser $[0{,}85 \quad 0{,}01 \quad 0{,}14]$, e a autenticação é realizada corretamente.

A idéia principal envolvida neste processo é que, em ambas as tentativas, o acesso do usuário foi garantido sem este saber que houve falha em um primeiro momento, já que, para ele, o processo de autenticação é transparente. Em outra situação, quando o usuário é realmente um impostor, o supervisor pode recusar o acesso sem bloquear o acesso do usuário legítimo e adotar as medidas de segurança adequadas.

5 RESULTADOS EXPERIMENTAIS

A fim de avaliar o desempenho do algoritmo proposto foi adotada uma metodologia de teste que consiste, basicamente, de duas etapas: (a) inclusão de usuários na base de IDDs e atualização dos modelos associados e; (b) verificação da eficácia dos modelos.

Para simular uma situação real de acesso de usuários a uma estação de trabalho, os conjuntos que compõem a BFIEE são acessados, uma única vez, de forma aleatória, seguindo a seqüência apresentada na Tabela 1. Lembrando que, cada um dos 50 conjunto da BFIEE representa uma seqüência de vídeo com 50 quadros, obtida em um início de sessão, e que cada um destes conjuntos representa, portanto, um *login* diferente. Esta metodologia foi adotada para fazer com que os resultados dos testes e processos de aprendizagem possam ser considerados válidos independentemente da seqüência de acesso dos usuários.

Na primeira parte do experimento é realizada a inclusão dos usuários ou a atualização dos modelos biométricos existentes na base de IDDs. Se o usuário não está registrado na base de IDDs, um modelo biométrico, construído a partir dos 50 vetores de feições extraídos da seqüência de vídeo, e uma senha são associados ao seu nome de usuário. Se ele já é um usuário registrado, é realizada a verificação biométrica. Em caso positivo de autenticação, o modelo é considerado válido e passa-se ao próximo conjunto da BFIEE. Em caso de falha na autenticação sendo que o usuário é legítimo, o modelo de aparência associado à classe é atualizado.

Após todos os 50 conjuntos da BFIEE terem sido utilizados e a fase de inclusão de usuários e adaptação dos modelos ter sido concluída, foi realizada a verificação da qualidade dos modelos. Amostras do BFIEE são escolhidas (cf. critérios apresentados na seqüência do capítulo) e a tentativa de autenticação facial é realizada. Com o aumento do número de usuários registrados na base de IDDs, alguma sobreposição inter-classes é esperada e, também, falhas na autenticação.

Tabela 1 - Seqüência para acesso à BFIEE.

Nº	Conjunto	Ação	Nº	Conjunto	Ação
1	usuário 1	cadastro	26	usuário 17	2º login e verificação
2	usuário 2	cadastro	27	usuário 13	2º login e verificação
3	usuário 3	cadastro	28	usuário 20	cadastro
4	usuário 4	cadastro	29	usuário 21	cadastro
5	usuário 5	cadastro	30	usuário 10	2º login e verificação
6	usuário 6	cadastro	31	usuário 13	3º login e verificação
7	usuário 7	cadastro	32	usuário 3	4º login e verificação
8	usuário 8	cadastro	33	usuário 12	2º login e verificação
9	usuário 9	cadastro	34	usuário 15	2º login e verificação
10	usuário 10	cadastro	35	usuário 22	cadastro
11	usuário 11	cadastro	36	usuário 7	2º login e verificação
12	usuário 12	cadastro	37	usuário 2	2º login e verificação
13	usuário 6	2º login e verificação	38	usuário 13	4º login e verificação
14	usuário 13	cadastro	39	usuário 7	3º login e verificação
15	usuário 11	2º login e verificação	40	usuário 1	3º login e verificação
16	usuário 14	cadastro	41	usuário 3	5º login e verificação
17	usuário 1	2º login e verificação	42	usuário 20	2º login e verificação
18	usuário 15	cadastro	43	usuário 4	2º login e verificação
19	usuário 16	cadastro	44	usuário 21	2º login e verificação
20	usuário 5	2º login e verificação	45	usuário 13	5º login e verificação
21	usuário 17	cadastro	46	usuário 5	3º login e verificação
22	usuário 18	cadastro	47	usuário 16	2º login e verificação
23	usuário 3	2º login e verificação	48	usuário 7	4º login e verificação
24	usuário 19	cadastro	49	usuário 19	2º login e verificação
25	usuário 3	3º login e verificação	50	usuário 16	3º login e verificação

5.1 CONFIGURAÇÕES DE TESTE

Os experimentos aqui apresentados são realizados com o intuito de avaliar os métodos apresentados ao longo do trabalho. Para tanto, são propostas algumas configurações de teste (CT) que se diferenciam nas etapas de construção de modelos e classificação das classes. Os resultados para cada CT são comparados com o teste de força bruta. Assim, as configurações de teste propostas são:

(a) **CT1**: Teste de força bruta, onde uma amostra é comparada a todas as outras amostras (exceto ela própria) de todas as classes por meio da busca exaustiva pela amostra mais próxima, utilizando distância Euclideana;

(b) **CT2**: Nesta CT cada classe é associada a um único modelo (i.e., não existem subclasses, (cf. 4.3.3.1) e o número de amostras associadas a uma classe pode variar;

(c) **CT3**: Nesta CT cada aparência do usuário é descrita por um modelo de subclasse (i.e., vários modelos podem ser associados a um usuário, (cf. 4.3.3.2) e todas as subclasses tem a mesma probabilidade *a priori* (cf. 4.3.2.1));

(d) **CT4**: Similar à CT3, utiliza o método de múltiplos modelos, porém todas as classes tem a mesma probabilidade *a priori* (cf. 4.3.3.2).

5.2 AVALIAÇÃO DO ALGORITMO DE APRENDIZAGEM

A primeira avaliação de desempenho realizada verifica do poder de adaptação dos modelos biométricos. Quando um modelo associado a um usuário necessita ser atualizado, devido a uma mudança de aparência, é necessário que a eficiência desta adaptação seja confirmada. Para tanto, a autenticação é refeita após cada atualização do modelo.

Para mensurar o quão efetivo o algoritmo é em *aprender* novas aparências faciais dos usuários, é realizada uma avaliação do algoritmo de aprendizagem. Esta medida é denominada taxa de sucesso de aprendizagem (TSA) e é dada por:

$$TSA = \frac{NAC}{NTA} \tag{5.1}$$

onde NAC é o número de atualizações corretas, ou seja, bem sucedidas, e NTA é o número total de atualizações realizadas.

Em nossos experimentos a TSA foi de 100%, ou seja, todos os modelos foram corretamente atualizados quando novas aparências de usuário foram apresentadas ao sistema[7]. Ainda, esta medida pode ser realizada *on-line*, dando um indicativo da capacidade de crescimento da base de IDDs.

5.3 AVALIAÇÃO DA CONFUSÃO INTER-CLASSES

Nesta segunda avaliação, é verificada a capacidade que o algoritmo tem para discriminar os usuários cadastrados. Para tanto, após a inserção de todos os usuários (e suas aparências) na base de IDDs, é refeita a autenticação dos usuários.

---

[7] Após a atualização, o modelo da classe correta, que não foi identificada como a mais provável, passou a ser a mais provável entre todas as classes da base de IDDs.

Uma medida quantitativa da confusão inter-classes é obtida por meio da construção de matrizes de confusão de classes (MCCs) (FUKUNAGA, 1990). São iniciadas sessões utilizando cada uma das 2500 imagens do BFIEE como amostras de autenticação independentes, sendo os resultados apresentados nas Tabelas 2-5, para cada uma das 4 configurações de teste propostas.

Como pode ser verificado na Tabela 5, que apresenta o melhor resultado, as confusões que ocorrem mais freqüentemente são: classe 2 classificada como classe 3 (14%), classe 11 como classe 15 (12%), classe 20 como classe 12 (12%), classe 20 como classe 10 (11%) e classe 15 como classe 11 (11%). Ainda, a classe com maior chance de ser confundida é a classe 2, com 41% de probabilidade. Estes 5 casos mais freqüentes de confusão são ilustrados na Figura 14, dando uma idéia da similaridade visual entre as classes.

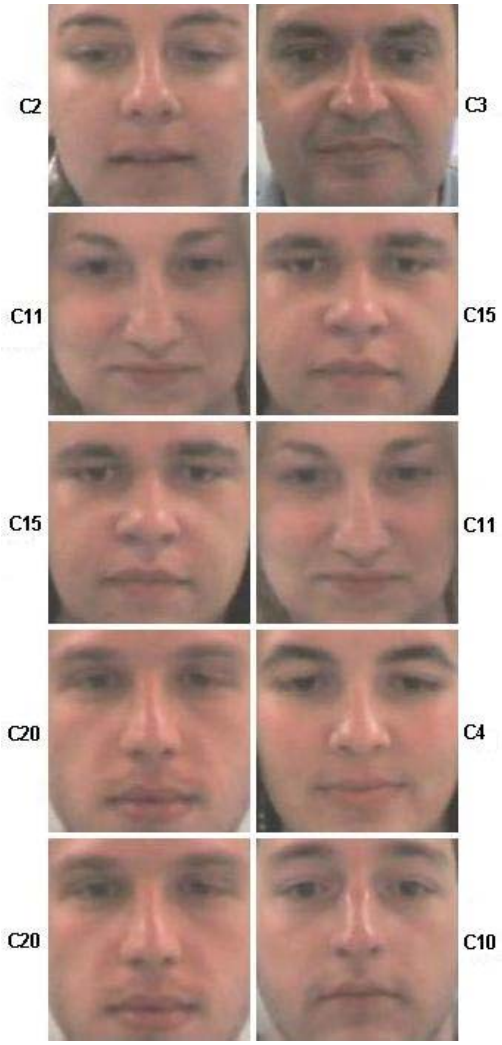


Figura 14 - Classes confundidas mais freqüentemente - esquerda: classe de entrada; direita: resultado da autenticação

.

Tabela 2 - Matriz de confusão de classes para a CT1.

Classe da amostra de entrada	**Resultados da classificação Classe mais provável**																					
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
C1	148	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C2	0	90	1	1	0	0	0	0	0	1	0	0	0	0	3	1	1	0	1	1	0	0
C3	1	2	225	0	8	0	7	0	0	1	2	1	3	0	0	0	0	0	0	0	0	0
C4	0	1	0	74	0	0	0	0	0	1	2	0	1	0	8	0	4	0	0	8	1	0
C5	3	0	13	0	110	0	2	0	4	1	0	0	17	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	95	0	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C7	0	0	1	0	1	0	198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	7	0	37	0	0	0	6	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	2	0	0	0	31	0	0	0	15	0	0	0	0	2	0	0	0	0
C10	0	1	1	0	0	0	0	0	0	75	0	0	8	0	0	0	4	1	7	3	0	0
C11	0	3	2	2	0	0	0	0	0	0	68	0	0	1	13	4	3	0	0	1	0	3
C12	0	0	0	0	0	4	0	8	0	0	0	88	0	0	0	0	0	0	0	0	0	0
C13	1	0	5	1	12	0	0	0	15	7	0	0	202	0	0	0	0	5	2	0	0	0
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	3
C15	0	4	0	8	0	0	0	0	0	0	9	0	4	1	65	4	5	0	0	0	0	0
C16	0	2	0	0	0	0	0	0	0	0	2	0	0	0	4	140	0	0	2	0	0	0
C17	0	4	0	6	0	0	0	0	0	3	2	0	0	0	4	0	75	0	1	5	0	0
C18	0	0	0	0	0	0	0	0	3	1	0	0	6	0	0	0	0	39	0	1	0	0
C19	0	6	2	1	0	0	0	0	0	10	0	0	2	0	0	1	1	0	77	0	0	0
C20	0	0	0	10	0	0	0	0	0	4	0	0	0	0	0	0	4	1	2	79	0	0
C21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0
C22	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	46

Tabela 3 - Matriz de confusão de classes para a CT2.

Classe da amostra de entrada	**Resultados da classificação Classe mais provável**																					
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
C1	133	0	0	0	15	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
C2	0	83	1	1	0	0	0	0	0	2	2	0	0	0	0	3	4	0	4	0	0	0
C3	3	1	217	0	10	0	3	1	0	0	0	6	7	0	0	0	0	0	2	0	0	0
C4	0	0	0	81	0	0	0	0	0	0	0	0	1	0	2	1	9	0	2	4	0	0
C5	5	0	13	0	95	0	3	0	1	0	0	0	33	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	86	0	5	0	0	0	9	0	0	0	0	0	0	0	0	0	0
C7	0	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	4	0	45	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	3	0	0	0	45	0	0	0	0	0	0	0	0	2	0	0	0	0
C10	0	0	0	2	0	0	0	0	0	80	0	0	8	0	0	0	3	1	3	3	0	0
C11	0	2	1	12	0	0	0	0	0	1	65	0	0	0	9	2	5	0	0	0	0	3
C12	0	0	0	0	0	2	0	6	0	0	0	92	0	0	0	0	0	0	0	0	0	0
C13	0	0	1	0	7	0	0	0	21	2	0	0	213	0	0	0	0	4	1	1	0	0
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	1
C15	0	3	0	4	0	0	0	0	0	2	10	0	1	0	74	5	1	0	0	0	0	0
C16	0	1	0	2	0	0	0	0	0	0	4	0	0	0	0	142	0	0	1	0	0	0
C17	0	6	0	3	0	0	0	0	0	6	0	0	0	0	2	0	74	0	2	7	0	0
C18	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	48	0	0	0	0
C19	0	17	3	0	0	0	0	0	0	4	0	0	1	0	0	1	2	0	72	0	0	0
C20	0	0	0	13	0	0	0	0	0	10	0	0	0	0	0	0	2	3	2	70	0	0
C21	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	0	0	0	8	0	84	0
C22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Tabela 4 - Matriz de confusão de classes para a CT3.

Classe da amostra de entrada	**Resultados da classificação Classe mais provável**																					
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
C1	147	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0
C2	0	78	4	1	0	0	0	0	0	3	6	0	0	0	1	0	2	0	5	0	0	0
C3	3	0	216	0	14	0	11	0	0	0	0	1	3	0	0	0	0	0	2	0	0	0
C4	0	0	0	78	0	0	0	0	0	0	0	0	1	0	5	0	8	1	2	5	0	0
C5	2	0	20	0	90	0	3	0	2	0	0	1	31	0	0	0	0	0	0	1	0	0
C6	0	0	0	0	0	86	0	5	0	0	0	9	0	0	0	0	0	0	0	0	0	0
C7	0	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	4	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	2	0	0	0	46	0	0	0	1	0	0	0	0	1	0	0	0	0
C10	0	0	0	2	1	0	0	0	0	84	0	0	6	0	0	0	3	1	3	0	0	0
C11	0	0	0	4	0	0	0	0	0	1	71	0	0	0	15	1	6	0	0	0	0	2
C12	0	0	0	0	0	2	0	10	0	0	0	88	0	0	0	0	0	0	0	0	0	0
C13	2	0	1	0	10	0	0	0	29	2	0	1	199	0	0	0	0	5	1	0	0	0
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	1
C15	0	2	0	1	0	0	0	0	0	2	9	0	1	0	79	4	2	0	0	0	0	0
C16	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	145	0	0	1	0	0	0
C17	0	4	0	3	0	0	0	0	0	7	0	0	0	0	4	0	70	0	4	8	0	0
C18	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	47	0	0	0	0
C19	0	4	8	0	0	0	0	0	0	8	0	0	1	0	0	1	2	0	76	0	0	0
C20	0	0	0	10	0	0	0	0	0	13	0	0	0	0	0	0	1	3	2	71	0	0
C21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0
C22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

Tabela 5 - Matriz de confusão de classes para a CT4.

Classe da amostra de entrada	**Resultados da classificação Classe mais provável**																					
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
C1	146	0	0	0	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0
C2	0	59	14	1	0	0	0	0	0	2	6	0	0	0	4	2	6	0	5	1	0	0
C3	1	0	207	0	25	0	3	0	0	0	0	6	6	0	0	0	0	0	2	0	0	0
C4	0	1	0	84	0	0	0	0	0	0	0	0	1	0	5	0	6	0	1	2	0	0
C5	0	0	2	0	135	0	1	0	3	0	0	0	9	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	86	0	5	0	0	0	9	0	0	0	0	0	0	0	0	0	0
C7	0	0	1	0	0	0	199	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	4	0	45	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	1	0	0	0	44	0	0	0	4	0	0	0	0	1	0	0	0	0
C10	0	1	0	1	0	0	0	0	0	82	0	0	6	0	0	0	5	1	4	0	0	0
C11	0	0	2	4	0	0	0	0	0	1	71	0	0	2	12	2	1	0	0	3	0	2
C12	0	0	0	0	0	2	0	8	0	0	0	90	0	0	0	0	0	0	0	0	0	0
C13	1	0	1	0	13	0	0	0	24	3	0	0	202	0	0	0	0	5	1	0	0	0
C14	0	0	0	0	0	0	0	0	0	0	1	0	0	48	0	0	0	0	0	0	0	1
C15	0	1	0	1	0	0	0	0	0	2	11	0	1	0	80	3	1	0	0	0	0	0
C16	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	144	0	0	1	0	1	0
C17	0	4	0	6	0	0	0	0	0	6	0	0	0	0	3	0	74	0	0	7	0	0
C18	0	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	46	0	0	0	0
C19	0	3	9	0	0	0	0	0	0	7	0	0	1	0	0	1	3	0	76	0	0	0
C20	0	0	0	12	0	0	0	0	0	11	0	0	0	0	0	0	2	3	0	72	0	0
C21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0
C22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

5.4 AVALIAÇÃO DO CLASSIFICADOR COMO PRÉ-FILTRO

Em uma situação ideal, a classe associada ao usuário atual (i.e., usuário em processo de autenticação) deve ser a mais provável entre todas as classes da base de IDDs. No entanto, devido à confusão inter-classes, esta premissa pode falhar e a classe correta ser classificada como uma classe menos provável.

Portanto, para cada tentativa de autenticação realizada, é construído um *ranking* de recuperação das classes no qual a classe mais provável aparece na 1ª posição e a menos provável, na última. Após este procedimento ter sido realizado para todas as 2500 amostras do BFIEE, foi construída um tabela na qual pode se verificar a freqüência com que as classes são recuperadas em cada posição do *ranking* (ver Tabela 6).

Tabela 6 - Ranking de recuperação de classes limitado nas 10 primeiras posições (utilizando 2500 amostras de autenticação).

Posição	**Percentuais Associados**			
	CT1	CT2	CT3	CT4
1	84,3	83,9	84,6	85,7
2	11,0	9,2	9,2	8,7
3	3,0	3,1	2,5	2,6
4	0,8	2,0	1,4	1,0
5	0,4	0,8	1,0	0,8
6	0,2	0,6	0,7	0,4
7	0,1	0,3	0,3	0,4
8	0,0	0,1	0,1	0,3
9	0,2	0,0	0,1	0,1
10	0,0	0,0	0,1	0,0

A Tabela 6 indica que, sob as condições da CT4, a classe correta (i.e., associada ao usuário sob autenticação) foi recuperada como a mais provável em 85,6% dos casos. Em 8,7% das tentativas foi recuperada como a segunda classe mais provável, em 2,6% dos casos como a terceira classe mais provável, e assim por diante.

Como alternativa, ao invés de tratar os $\overrightarrow{FV}$ de cada uma das 2500 imagens do banco como uma amostra de autenticação, foi utilizada o $\overrightarrow{FV}$ mediano entre os 50 $\overrightarrow{FV}$ obtidos para cada conjunto da BFIEE. A mediana foi utilizada por se tratar de um procedimento típico para

a eliminação de *outliers* (i.e., pontos "fora da curva"), aumentando a taxa de sucesso do autenticador conforme resultados apresentados na Tabela 7.

Tabela 7 - Ranking de recuperação de classes limitado nas 5 primeiras posições (utilizando as 50 amostras medianas).

Posição	**Percentuais Associados**			
	CT1	CT2	CT3	CT4
1	96,0	96,0	96,0	98,0
2	4,0	4,0	2,0	2,0
3	0,0	0,0	0,0	0,0
4	0,0	0,0	0,0	0,0
5	0,0	0,0	2,0	0,0

Esta alteração proporcionou que a classe correta fosse recuperada como a mais provável em 98% dos casos para a CT4. Além disso, a classe correta passou a estar dentro das 5 primeiras posições do *ranking* em todas as configurações de teste. Novamente, os melhores resultados são obtidos com a CT4.

5.5 COMPARAÇÃO DO MÉTODO PROPOSTO

Para melhor avaliar o método proposto neste trabalho, foi realizada uma comparação com o algoritmo proposto por Yang *et al.* (2004). Este algoritmo, que utiliza análise bi-dimensional de componentes principais (ACP-2D) para representação e reconhecimento de faces, é chamado aqui de MET_YANG. As comparações são realizadas utilizando a CT4, que apresentou os melhores resultados, e é chamada, a partir deste ponto, de MEU_MET. Enquanto que, em MEU_MET são utilizadas imagens coloridas para extrair feições, em MET_YANG são utilizadas imagens em tons de cinza. Todavia, como os tons de pele possuem cromaticidades semelhantes, pode-se considerar que as variações são devidas à intensidade e, portanto, os métodos podem ser comparados. Cabe ressaltar, ainda, que MET_YANG é utilizada a face completa para a extração atributos.

5.5.1 Métodos como autenticadores de identidades

Em um primeiro experimento procedemos a inserção dos usuários no na base de IDDs, sendo que, das 50 imagens de cada seqüência de vídeo (22 usuários), as 5 primeiras foram utilizadas para treinamento do classificador MET_YANG, e as 45 restantes foram utilizadas

para teste. Portanto, temos um conjunto de treinamento com 110 imagens e um conjunto de testes com 990 imagens. A taxa de classificações corretas (TCC) obtidas por MET_YANG (F1), neste experimento, foi de 98,79%.

Para verificar a eficiência de MET_YANG quando da inserção de novas aparências de usuários, foram utilizados as demais seqüências da BFIEE. Após todas as aparências terem sido inseridas, procedemos novamente com o classificador (MET_YANG F2)para este novo conjunto de testes, agora com 1400 imagens. Os resultados obtidos foram de 83,21% para TCC e demonstram que o método perde desempenho quando apresentadas novas aparências de usuários.

Em um terceiro momento, o classificador MET_YANG (F3) passa por uma etapa de re-treinamento utilizando amostras das novas aparências. Os resultados voltam a ser satisfatórios, atingindo uma TCC de 99,43%. Resultados para as 3 fases de tese são apresentadas na Tabela 8, juntamente com o resultado obtido por MEU_MET.

Tabela 8 - Ranking de recuperação de classes comparativo.

Posição	**Percentuais associados**			
	MET_YANG F1	MET_YANG F2	MET_YANG F3	MEU_MET
1	98,8	83,2	99,4	85,7
2	0,9	8,7	0,2	8,7
3	0,2	5,1	0,2	2,6
4	0,0	1,4	0,1	1,0
5	0,1	0,9	0,0	0,8
6	0,0	0,1	0,0	0,4
7	0,0	0,1	0,0	0,4
8	0,0	0,0	0,0	0,3
9	0,0	0,1	0,1	0,1
10	0,0	0,0	0,0	0,0
11	0,0	0,1	0,0	0,0
12	0,0	0,1	0,0	0,0
13	0,0	0,0	0,0	0,0
14	0,0	0,0	0,0	0,0
15	0,0	0,1	0,0	0,0
16	0,0	0,0	0,0	0,0
17	0,0	0,1	0,0	0,0
18	0,0	0,0	0,0	0,0
19	0,0	0,0	0,0	0,0
20	0,0	0,0	0,0	0,0
21	0,0	0,0	0,0	0,0
22	0,0	0,0	0,0	0,0

Percebe-se claramente que o método apresentado por Yang *et al.* é bastante adequado para realizar autenticações utilizando a BFIEE, desde que seja constantemente atualizado. No entanto, o re-treinamento do classificador MET_YANG é um processo de alto custo computacional, já que utiliza amostras de todas as aparências presentes no banco de IDDs. Ainda, se um novo usuário é inserido na base de IDDs, um re-treinamento completo deve ser realizado. Portanto, o aumento do número de aparências e usuários na base tornam o sistema impraticável em longo prazo.

Por outro lado, o classificador proposto em MEU_MET apresenta menor TCC quando comparado ao MET_YANG treinado para todas as aparências. No entanto, a atualização dos modelos pode ser realizada *on-line* e o custo computacional desta atualização permanece constante com a entrada de novas aparências ou usuários.

5.5.2 Métodos como pré-filtros

Como visto anteriormente, outra possibilidade de utilização de MEU_MET é como um pré-filtro para algoritmos mais especializados como os SRF. Neste caso, somente amostras significativas devem ser enviadas ao SRF para uma análise mais detalhada.

Para realizar um estudo comparativo dos métodos, utilizamos uma situação real de uso, obtendo o *ranking* de recuperação de usuários da base de IDDs para os melhores classificadores de cada abordagem. Para o método MET_YANG são utilizados o classificador totalmente treinado e as amostras de autenticação correspondem a primeira amostra da seqüência de vídeo obtida em cada *login*. Para o método MEU_MET utilizam-se as amostras medianas descritas anteriormente. Tendo os *rankings* de recuperação para os métodos (ver Tabela 9), podemos construir as curvas *recall-precision* de ambos, conforme apresentado na Figura 15.

Tabela 9 - Ranking de recuperação de classes comparativo limitado nas 5 primeiras posições.

Posição	**Percentuais Associados**	
	MET_YANG	MEU_MET
1	88,0	98,0
2	10,0	2,0
3	2,0	0,0
4	0,0	0,0
5	0,0	0,0

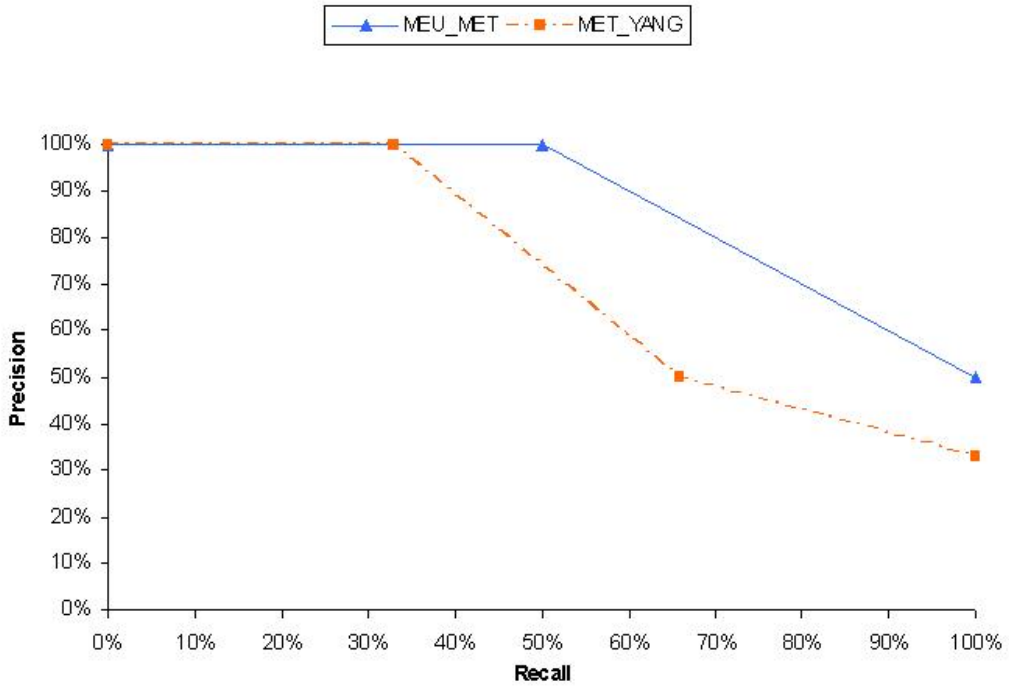


Figura 15 - Curvas *recall-precision* obtidas.

Os resultados mostram que, utilizando o método proposto neste trabalho, apenas 2 classes são enviadas ao SRF para garantir que a classe correta está presente. Utilizando o MET_YANG, é necessário enviar 3 classes ao SRF para ter esta garantia.

5.6 RESULTADOS EXPERIMENTAIS NA BASE DE DADOS EXTENDIDA

Para obter mais dados para avaliação do algoritmo proposto, foi realizado um segundo conjunto de testes. Para tanto, a base de dados foi extendida para 75 seqüências de vídeo e 40 usuários. Os testes realizados visaram verificar o desempenho do classificador CT4 sob outras condições que não as testadas no primeiro conjunto de testes.

Para tanto, foram utilizadas 90\% das imagens da base extendida para treinamento e os 10% restantes para teste. Os resultados obtidos nestes experimentos são apresentados na Tabela 10, sendo que:

(a) **A1:** autenticação utilizando as amostras do conjunto de treinamento como entrada;

(b) **A2:** autenticação utilizando as amostras do conjunto de teste como entrada;

(c) **A3:** autenticação utilizando as amostras medianas dos conjuntos de teste e treinamento como entrada.

Tabela 10 - Taxas de autenticação (TA) para o segundo conjunto de testes.

Posição	**Nº de amostras**	**TA %**	**FN**
A1	3375	89,6	352
A2	375	72,0	105
A3	75	96,0	3

Conforme pode ser verificado na Tabela 10, a utilização dos vetores de feições medianos apresentam resultado satisfatório mesmo com o aumento do número de classes. Verifica-se ainda que, para este caso, foi observada a ocorrência de apenas 3 falsos negativos (FN).

Ainda, para verificar a eficácia da modelagem, tentou-se realizar a autenticação utilizando distância Euclideana entre as amostras de entrada e as medianas dos vetores de feições das seqüências de vídeo, sendo que o resultado obtido foi de 74,9% de autenticações corretas.

6 CONCLUSÕES

Neste trabalho foi proposto um algoritmo para autenticação adaptativa de usuários e gerenciamento de identidades digitais com base em feições faciais. Este método foi projetado para proporcionar rápida autenticação de usuários, sendo bastante direto e simples de implementar.

Os resultados preliminares são encorajadores, mostrando que, em 85,6% dos casos, os usuários são autenticados corretamente com apenas uma amostra da face (cf. Tabela 6, CT4). Ainda, potencialmente, 98% dos usuários podem ser autenticados quando é utilizada uma amostra mediana obtida a partir de 50 amostras faciais. Cabe ressaltar que estes resultados incluem a autenticação de usuários cuja aparência mudou com o passar do tempo. Resultados similares foram obtidos quando a base de faces foi extendida de 22 para 40 indivíduos, sendo que a taxa de detecção, neste caso, foi de 96%.

A maior dificuldade fica por conta da forte confusão entre algumas classes. Isto pode ser confirmado pela análise do teste de força bruta, que se mostrou menos eficaz que os demais. A principal razão é o baixo poder de discriminação das feições utilizadas, apesar de que, para o banco de faces testado, um número reduzido de feições faciais se mostrou adequado. Contudo, é esperado que mais feições sejam necessárias para operar com bases de IDDs maiores.

Ainda, como alternativa para dar mais robustez ao processo de autenticação, o algoritmo pode, também, ser utilizado como um filtro para pré-selecionar faces representativas de usuários. Suponha a situação onde a autenticação de um usuário legítimo falha. Se o modelo de face do usuário é selecionado pelo classificador como um dos k mais prováveis para amostra de entrada, então o modelo pode ser confirmado por um SRF entre os k modelos pré-selecionados. Isto faz com que ocorra uma redução do espaço de busca do SRF para k casos, aumentando sua confiabilidade e, conseqüentemente, do autenticador. Os

experimentos indicam que, quando o algoritmo atua como um pré-filtro, este é capaz de selecionar apenas duas faces para enviar ao SRF (utilizando a CT4 e amostra mediana, cf. Tabela 7).

Verificou-se que, na comparação com um método proposto por Yang *et al.* (2004), a abordagem utilizada neste trabalho apresenta algumas vantagens importantes. Atuando como autenticador, a principal vantagem é a possibilidade de constante atualização dos modelos com baixo custo computacional. A desvantagem fica por conta da taxa de classificação que é inferior. Como pré-filtro, o método proposto apresentou melhores resultados que o método de Yang *et al*., selecionando apenas 2 classes para enviar a um SRF.

Em trabalhos futuros, planeja-se investigar descritores faciais adicionais para melhorar a discriminação entre faces, permitindo a utilização de bases de IDDs maiores. Comparações com outros métodos de autenticação facial também devem ser realizadas. Outro ponto importante a se estudar são alternativas para tornar o algoritmo robusto a variações de iluminação, pose, cenário, escala e outras condições de aquisição, sendo este um desafio para todas as áreas que envolvem processamento de faces.

REFERÊNCIAS

BEN-YACOUB, S.; ABDELJAOUED, Y.; MAYORAZ, E. Fusion of face and speech data for person identity verification. **IEEE Transactions on Neural Networks**, [S.l.], v.10, p.1065–1074, Sept. 1999.

BIGUN, J.; DUC, B.; SMERALDI, F.; FISCHER, S.; MAKAROV, A. Multimodal person authentication. In: FACE RECOGNITION: FROM THEORY TO APPLICATIONS, NATO ADVANCED STUDY INSTITUTE PROGRAMME, 1997, Stirling, UK. **Proceedings...**, UK: [s.n.], 1997.

BIGUN, J. et al. Multimodal biometric authentication using quality signal in mobile communications. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE ANALYSIS AND PROCESSING (ICIAP'03), 2003, Mantova, Italy. **Proceedings...**, Italy: IEEE, 2003. p.2–11.

CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.8, p.679–698, Nov. 1986.

DRIMBAREAN, A.; WHELAN, P. Experiments on colour texture analysis. **Pattern Recognition Letters**, [S.l.], v.22, p.1161–1167, Feb. 2001.

FALIPOU, F. **Face Detection and Reconstruction Based on the Active Shape Models, Features Extraction and Automatic Registration of Face Images**, 2006. Trabalho (Graduação), Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2006.

FIGUEIREDO, M.; JAIN, A. Unsupervised learning of finite mixture models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.24, p.381–396, Mar. 2002.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. 2.ed. San Diego: Academic Press, 1990.

GONZALEZ, R.; WOODS, R. **Digital Image Processing**. Reading: Addison-Wesley, 1992.

GOTH, G. Identity management access specs are rolling along. **IEEE Internet Computing**, [S.l.], v.9, p.9–11, Jan. 2005.

HJELMAS, E.; LOW, B. Face detection: a survey. **Computer Vision and Image Understanding**, [S.l.], v.83, p.236–274, 2001.

HSU, R.-L.; ABDEL-MOTTALEB, M.; JAIN, A. Face detection in color Images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.24, p.696–706, May 2002.

KAWAGUCHI, T.; HIDAKA, D.; RIZON, M. Robust extraction of eyes from face. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR'00), 2000, Barcelona, Spain. **Proceedings...**, Spain: IEEE, 2000. p.1109–1114.

KOREMAN, J. et al. Multi-modal biometric authentication on SecurePhone PDA. In: MULTI-MODAL USER AUTHENTICATION WORKSHOP, 2006, Toulouse, France. **Proceedings...**, France: [s.n.], 2006.

KOVAC, J.; PEER, P.; SOLINA, F. Human skin colour clustering for face detection. In: INTERNATIONAL CONFERENCE IN COMPUTER AS A TOOL (EUROCON'03), 2003, Ljubljana, Slovenia. **Proceedings...**, Slovenia: [s.n], 2003. p.144–148.

KRAWCZYK, S. **User Authentication Using On-line Signature and Speech**, 2005. Thesis (Master of Science), Michigan State University, Michigan, USA, 2005.

LAW, M.; FIGUEIREDO, M.; JAIN, A. Simultaneous feature selection and clustering using mixture models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.26, p.1154–1166, Sept. 2004.

LEE, T.; PARK, S.-K.; PARK, M. A new facial features and face detection method for human-robot interaction. In: INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2005, Barcelona, Spain. **Proceedings...**, Spain: IEEE, 2005. p.2063–2068.

LEUNG, T.; BURL, M.; PERONA, P. Finding faces in cluttered scenes using random labeled graph matching. In: INTERNATIONAL CONFERENCE IN COMPUTER VISION (ICCV'95), 1995, Cambridge, USA. **Proceedings...**, USA: IEEE, 1995. p.637–644.

MELEN, R.; PIGNOLO, M.; SIOLI, M. A multimodal authentication system for authorizing the access to NGN services. In: IEEE INTERNATIONAL CONFERENCE ON NETWORKING AND SERVICES (ICNS'06), 2006, Silicon Valley, USA. **Proceedings...**, USA: IEEE, 2006. p.47–51.

O'GORMAN, L. Comparing passwords, tokens, and biometrics for user authentication. **Proceedings of the IEEE**, [S.l.], v.91, p.2021–2040, Dec. 2003.

PENG, K. et al. A robust algorithm for eye detection on gray intensity face without spectacles. **Journal of Computer Science and Technology**, [S.l.], v.5, p.127–132, Oct. 2005.

PHIRI, J.; AGBINYA, J. Modelling and information fusion in digital identity management systems. In: INTERNATIONAL CONFERENCE ON NETWORKING, INTERNATIONAL CONFERENCE ON SYSTEMS AND INTERNATIONAL CONFERENCE ON MOBILE COMMUNICATIONS AND LEARNING TECHNOLOGIES (ICN/ICONS/MCL'06), 2006, Morne, Mauritius. **Proceedings...**, Mauritius: IEEE, 2006.

RANDEN, T.; HUSOY, J. Filtering for texture classification: a comparative study. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.21, p.291–310, Apr. 1999.

RíHA, Z.; MATYáS, V. **Biometric Authentication Systems**. Czech Republic: Faculty of Informatics of Masaryk University, 2000. FIMU-RS-2000-08.

ROWLEY, H.; BALUJA, S.; KANADE, T. Neural network-based face detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.20, p.23–38, Jan. 1998.

SAAD, E.-S. et al. Frontal-view face detection in the presence of skin-tone regions using a new symmetry approach. **Journal of Computer Science and Technology**, [S.l.], v.6, p.85–91, Oct. 2006.

SAKAI, T.; NAGAO, M.; KANADE, T. Computer analysis and classification of photographs of human faces. In: USA-JAPAN COMPUTER CONFERENCE, 1., 1972. **Proceedings...**, [S.l.: s.n.], 1972. p.55–62.

SHIM, S.; BHALLA, G.; PENDYALA, V. Federated identity management. **Computer**, [S.l.], v.38, p.120–122, Dec. 2005.

SHU, C.; DING, X. Multi-biometrics fusion for identity verification. In: IEEE INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR'06), 2006, Hong Kong. **Proceedings...**, Hong Kong: IEEE, 2006. p.493–496.

SNELICK, R. et al. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.27, p.450–455, Mar. 2005.

SUNG, K.-K.; POGGIO, T. Example-based learning for view-based human face detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.20, p.39–51, Jan. 1998.

TUCERYAN, M.; JAIN, A. Texture analysis. In: CHEN, C.; PAU, L.; WANG, P. (Ed.). **The Handbook of Pattern Recognition and Computer Vision**. 2.ed. New Jersey: World Scientific Publishing Co., 1998.

UNSER, M. Local linear transforms for texture measurements. **Signal Processing**, [S.l.], v.11, p.61–79, July 1986.

VEZHNEVETS, V.; SAZONOV, V.; ANDREEVA, A. A survey on pixel-based skin color detection techniques. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS & VISION (GRAPHICON'03), 2003, Moscow, Russia. **Proceedings...**, Russia: [s.n.], 2003. p.85–92.

VIOLA, P.; JONES, M. Robust real-time object detection. In: INTERNATIONAL WORKSHOP ON STATISTICAL AND COMPUTATIONAL THEORIES OF VISION - MODELLING, LEARNING, COMPUTING, AND SAMPLING, 2., 2001, Vancouver, Canada. **Proceedings...**, Canada: [s.n.], 2001.

WANG, P. et al. Combining face and iris biometrics for identity verification. In: COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05), 2005, San Diego, USA. **Proceedings...**, USA: IEEE, 2005. p.164–171.

WANG, Y.; TAN, T.; JAIN, A. Combining face and iris biometrics for identity verification. In: INTERNATIONAL CONFERENCE ON AUDIO- AND VIDEOBASED BIOMETRIC PERSON AUTHENTICATION (AVBPA'03), 2003, Guildford, UK. **Proceedings...**, UK: [s.n.], 2003. p.805–813.

WARING, C. et al. Face detection using spectral histogram and SVMs. **IEEE Transactions on Systems, Man, and Cybernetics - Part B:** cybernetics, [S.l.], v.35, p.467–476, Jun. 2005.

YANG, J. et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.26, p.131–137, Jan. 2004.

YANG, M.-H.; KRIEGMAN, D.; AHUJA, N. Detecting faces in images: a survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.24, p.34–58, Jan. 2002.

ZHOU, J. et al. A face and fingerprint identity authentication system based on multi-route detection. **Neurocomputing**, [S.l.], v.70, p.922–931, 2007.

ZIVKOVIC, Z.; HEIJDEN, F. Recursive unsupervised learning of finite mixture models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.26, p.651–656, May 2004.