Universidade Federal da Bahia
Instituto de Matemática

Programa de Pós-Graduação em Ciência da Computação

FUSÃO DE IMAGENS 2D, 3D E NIR PARA AUTENTICAÇÃO FACIAL CONTÍNUA UTILIZANDO SENSORES DE BAIXO CUSTO

Leone da Silva de Jesus

DISSERTAÇÃO DE MESTRADO

Salvador
15 de Setembro de 2017

LEONE DA SILVA DE JESUS

FUSÃO DE IMAGENS 2D, 3D E NIR PARA AUTENTICAÇÃO FACIAL CONTÍNUA UTILIZANDO SENSORES DE BAIXO CUSTO

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Mauricio Pamplona Segundo

Salvador
15 de Setembro de 2017

Sistema de Bibliotecas - UFBA

Jesus, Leone da Silva.

FUSÃO DE IMAGENS 2D, 3D E NIR PARA AUTENTICAÇÃO FACIAL CONTÍNUA UTILIZANDO SENSORES DE BAIXO CUSTO / Leone da Silva de Jesus – Salvador, 2017.

57p.: il.

Orientador: Prof. Dr. Mauricio Pamplona Segundo.

Dissertação (Mestrado) – Universidade Federal da Bahia, Instituto de Matemática, 2017.

1. Reconhecimento Facial. 2. Autenticação Contínua. 3. Biometria Multimodal. I. Pamplona Segundo, Mauricio. II. Universidade Federal da Bahia. Instituto de Matemática. III Título.

TERMO DE APROVAÇÃO

LEONE DA SILVA DE JESUS

FUSÃO DE IMAGENS 2D, 3D E NIR PARA AUTENTICAÇÃO FACIAL CONTÍNUA UTILIZANDO SENSORES DE BAIXO CUSTO

Esta Dissertação de Mestrado foi julgada adequada à obtenção do título de Mestre em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia.

Salvador, 15 de Setembro de 2017

Prof. Dr. Maurício Pamplona Segundo
Universidade Federal da Bahia

Prof. Dr. Rubisley de Paula Lemes
Universidade Federal da Bahia

Dr. Fillipe Dias Moreira de Souza
Intel/USA

AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar ao meu orientador, Maurício Pamplona Segundo, por ter sido peça fundamental para a conclusão desse trabalho e por ter me ajudado de diversas formas durante todo o meu mestrado. Sua orientação sempre foi atenciosa e paciente, e além de todos os ensinamentos que ele me passou, seus conselhos também contribuíram para meu crescimento pessoal. Gostaria também de agradecer aos meus colegas de laboratório, por todo o apoio, mesmo que indiretamente, e em especial a Gabriel Dahia Fernandes e Matheus Magalhães Batista dos Santos, que colaboraram de forma essencial para a conclusão desse trabalho. Finalmente, gostaria de agradecer aos meus pais, Jaimario Manoel de Jesus e Claudejane Gonçalves da Silva, à minha noiva, Alana Bonfim Goés, e à minha avó, Janira Alves Gonçalves, por todo o amor, suporte, motivação, conselhos, dentre outras contribuições e sentimentos positivos que me levaram a concluir essa jornada do mestrado.

RESUMO

O uso da autenticação contínua é mostrado como uma opção para aumentar a segurança de sistemas de controle de acesso, pois o método busca garantir que o acesso seja realizado somente pelo indivíduo autenticado previamente, reconhecendo-o constantemente. A precisão da autenticação é uma preocupação para tais sistemas, mas, além de seguros, estes também devem ser bem aceitos pelos usuários. O reconhecimento facial demonstra-se mais adequado nesse cenário por ser universal, distintivo, não intrusivo e não exigir esforço do usuário. Para prevenir tentativas de invasões devido às limitações do reconhecimento facial tradicional, o reconhecimento biométrico multimodal mostrou-se uma alternativa viável para alcançar um nível aceitável de segurança. Portanto, a proposta deste trabalho é um método de autenticação contínua baseado em múltiplas propriedades faciais obtidas através de imagens de textura, de profundidade e de infravermelho, que atualmente podem ser adquiridas simultaneamente por um único sensor de baixo custo. A extração de características faciais em todas as modalidades foi realizada por uma rede neural de convolução estado-da-arte, e múltiplos métodos de fusão em diferentes níveis do processo de autenticação foram investigados, tendo como melhor resultado um método pouco utilizado na literatura. Esse método foi então empregado na autenticação contínua multimodal de faces, obtendo resultados melhores do que os apresentados por soluções monomodais.

Palavras-chave: AUTENTICAÇÃO CONTÍNUA. RECONHECIMENTO FACIAL. BIOMETRIA MULTIMODAL. FUSÃO DE CARACTERÍSTICAS.

ABSTRACT

For a system to have proper access control, the use of continuous authentication has demonstrated to be a viable solution, as it will ensure that access remains being performed by the person previously authenticated by recognizing him constantly. Authentication accuracy is a concern for such systems, but, besides its security, these must also be well accepted by users. Facial recognition is more appropriate in this scenario, since it is universal, distinctive, non-intrusive and does not require effort from the user. To avoid the different ways of spoofing due to limitations of traditional face recognition, the multimodal biometric recognition has appeared as a viable alternative to achieve an acceptable safety level. Therefore, the purpose of this work is a continuous authentication method based on multiple facial properties using texture, depth and infrared images, which can be acquired simultaneously by a single low cost sensor nowadays. Facial description in all modalities was carried out by a state-of-the-art neural network, and multiple fusion methods at different levels of the authentication process were investigated, with the best result being an unusual method in the literature. This method was then employed in multimodal face continuous authentication, obtaining better results than those presented by monomodal solutions.

Keywords: CONTINUOUS AUTHENTICATION. FACE RECOGNITION. MULTIMODAL BIOMETRY. FEATURE FUSION.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

Capítulo 1

INTRODUÇÃO

Mesmo antes da era da informação, o controle de acesso já era um aspecto importante para o desenvolvimento da civilização. Há milhares de anos, paredes eram controladores de espaço, sendo portões e portas usadas como pontos de acesso em seus lados. Para os recursos mais valiosos, animais, guardas e palavras-chave foram inicialmente usados para melhorar a segurança. Posteriormente, as chaves e fechaduras mecânicas tornaram-se uma forma mais prática de controlar o acesso (*i.e.* apenas usuários autorizados tinham chaves para cadeados específicos). Embora chaves e cadeados ainda sejam amplamente utilizados hoje, a informatização trouxe opções mais seguras para controlar acessos, como *tokens*, cartões, senhas e dados biométricos. Estes são utilizados para proteger desde informações básicas do usuário, como dados pessoais em smartphones, até recursos perigosos, como maquinaria pesada e armas militares (SANDHU; SAMARATI, 1994; MCC, 2006; O'GORMAN, 2003).

O reconhecimento de dados biométricos garante o acesso a um recurso apenas à pessoa autorizada, e não a algo que ela possui ou que ela sabe, como acontece em métodos tradicionais de autenticação comercialmente utilizados (*e.g.* senhas, cartões de acesso) (BOLLE et al., 2003). Esta modalidade de reconhecimento analisa características humanas que sejam discriminativas o suficiente para distinguir indivíduos, estando em constante evolução por não haver um método adequado para qualquer contexto. Pesquisas na área de biometria baseiam-se em diferentes métodos para reconhecimento de pessoas, podendo utilizar biometrias físicas, comportamentais ou temporárias. O reconhecimento de biometrias físicas utiliza a descrição de caraterísticas de partes do corpo que são distintas entre indivíduos, como impressão digital, íris, DNA e faces. O reconhecimento biométrico comportamental baseia-se em ações realizadas pelo corpo que seguem um padrão capaz de diferenciar pessoas, como forma de falar, caminhar, assinar e até mesmo a frequência de batimentos cardíacos. Biometrias temporárias são características que fornecem informações sobre um indivíduo, mesmo não sendo altamente distintivas ou possuindo permanência suficiente para diferenciar qualquer indivíduo de um outro, tendo como exemplo cor de roupa, pele e cabelo, altura, peso e presença de pelos faciais (NIINUMA; PARK; JAIN, 2010; JAIN; ROSS; PRABHAKAR, 2004).

Segundo Jain, Ross e Prabhakar (2004), para que uma característica biométrica seja utilizada para autenticação, ela precisa: ser capaz de reconhecer qualquer pessoa (universalidade); ser capaz de distinguir qualquer pessoa de uma outra (distinguibilidade); garantir o reconhecimento de uma mesma pessoa independente do período de tempo que passe (permanência); ser coletável e mensurável (coletabilidade). Assim como um sistema capaz de reconhecer usuários por suas biometrias precisa: considerar fatores como ambiente e recursos que possam impactar na acurácia, tempo e custo do reconhecimento (desempenho); considerar questões de privacidade e requisitos legais que devem ser respeitadas, bem como a usabilidade do sistema e aceitação do usuário (aceitabilidade); e reconhecer as possibilidades do sistema ser fraudado (circunvenção).

Uma das possibilidades de autenticação biométrica que melhor abrange as características descritas é o reconhecimento facial. Faces possuem múltiplas características mensuráveis (*e.g.* tamanho do nariz, espaçamento entre olhos), são o meio mais utilizado pelos humanos para se reconhecerem e são universais. Além disso, sistemas capazes de reconhecer faces têm como vantagem o fato de não serem intrusivos e de não exigirem esforço por parte do usuário. Entretanto, a principal desvantagem desta biometria é a alta variabilidade que faces podem apresentar, seja por meio do uso de acessórios (*e.g.* óculos, maquiagem), mudança de componentes faciais (*e.g.* remoção de pelos faciais), envelhecimento ou variação em expressões faciais e na iluminação. Faces também são conhecidas por serem inferiores a outras biometrias físicas populares na literatura (*e.g.* íris e impressão digital) em termos de distinção, que é o critério principal procurado em características para serem utilizadas em sistemas de reconhecimento biométrico (JAIN; ROSS; PRABHAKAR, 2004; HASSABALLAH; ALY, 2015; SEGUNDO, 2013).

O reconhecimento facial pode ser realizado por diferentes propriedades, como textura (2D), formato (3D) e temperatura (infravermelho), e a melhoria e o barateamento de sensores multimodais tem permitido o uso de sistemas multimodais para superar as diversas fraquezas do reconhecimento facial (MIN; KOSE; DUGELAY, 2014; HASSABALLAH; ALY, 2015). Para exemplificar, imagens de face em infravermelho (IR- *infrared*) (Figura 1.1(a)) são menos variantes à iluminação que imagens de cor (Figura 1.1(c)), enquanto imagens de face 3D (Figura 1.1(b)) são mais robustas às variações de pose que imagens de cor e IR. Assim, a combinação entre as modalidades não só pode aprimorar a precisão do reconhecimento, mas também funcionará em cenários menos controlados.

Outro impacto recente nos sistemas de reconhecimento facial é o advento das Redes Neuronais Convolucionais (CNNs - *Convolutional Neural Networks*). Faces descritas com base em CNNs são melhor discriminadas, seja em condições controladas ou do mundo real, em comparação com outros descritores bem-sucedidos no estado-da-arte, como Padrões Binários Locais (LBP - *Local Binary Patterns*), Análise de Discriminantes Lineares (LDA - *Linear Discriminant Analysis*) e assim por diante (DING; TAO, 2015a). Em um dos *benchmarks* mais populares na literatura, o banco de imagens faciais *Labeled Faces in Wild* (LFW), pesquisas baseadas em CNN estão atingindo menos de 1% de erro, superando os resultados humanos, o que mostra a eficácia destes métodos em cenários não controlados (WU; HE; SUN, 2015).

Nós acreditamos que a combinação de informações multimodais baseados em descritores CNN pode tornar sistemas de reconhecimento facial mais confiáveis do que os

(a) IR

(b) 3D

(c) 2D

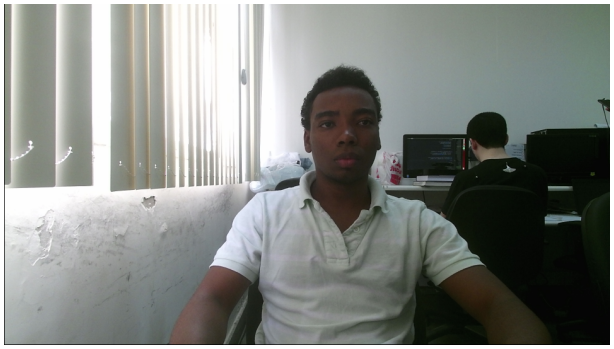


Figura 1.1 Exemplo de imagens de (a) IR, (b) 3D e (c) 2D adquiridas simultaneamente.

sistemas de reconhecimento de impressões digitais e íris, preservando ainda suas vantagens em termos de coletabilidade e aceitabilidade. Um sistema com tais características é mais adequado para o controle de acesso, que é um dos principais fins do reconhecimento biométrico. No entanto, tanto os métodos tradicionais quanto os biométricos realizam a autenticação apenas antes de permitir o acesso. Ou seja, os mesmos não poderão impedir que uma outra pessoa utilize o recurso protegido após a verificação inicial, seja este recurso um dado, informação ou equipamento sigiloso (JAIN; ROSS; PRABHAKAR, 2004). Uma possível solução para o problema descrito é a autenticação contínua. Sistemas baseados nesta ideia identificam o indivíduo autorizado, e o acesso só é permitido enquanto for possível garantir sua identidade. Para tanto, o sistema verifica constantemente a identidade do usuário, sendo uma opção de segurança mais adequada quando um controle de acesso mais rigoroso for necessário (ALTINOK; TURK, 2003; SEGUNDO, 2013). Contudo, nem todas as biometrias podem ser utilizadas para autenticação contínua, e dentre as que são possíveis de serem utilizadas, a relação entre custo, desempenho e usabilidade faz o reconhecimento de faces mais adequado que as demais (SEGUNDO, 2013).

Neste contexto, este trabalho propõe um processo de autenticação contínua baseado em imagens faciais, sob a hipótese de que o uso de múltiplas propriedades da face, como textura, geometria e infravermelho próximo (NIR - *near infrared*), aperfeiçoará a precisão do reconhecimento facial e prevenirá acessos indevidos. Para mostrar a eficácia do

reconhecimento multibiométrico proposto, investigamos vários métodos para combinar as propriedades em diferentes níveis de fusão. As imagens utilizadas nesta investigação foram adquiridas em uma aplicação do mundo real usando descritores extraídos por um modelo CNN no estado-da-arte, o que nos permitiu também apresentar um método de fusão incomum com base em nossas observações sobre este tipo de descrição. Para a autenticação contínua, utilizamos o método de fusão com melhor desempenho e demonstramos suas vantagens em comparação com desempenho das modalidades individualmente. Até onde sabemos, este é o primeiro trabalho a combinar essas três modalidades adquiridas simultaneamente por um único sensor de baixo custo.

Além deste capítulo introdutório, este trabalho apresenta outros cinco capítulos. O Capítulo 2 tem foco em mostrar conceitos e definições que permitam o devido entendimento do projeto proposto e de cada uma de suas etapas. Em seguida, o Capítulo 3 aborda os demais trabalhos que também utilizam autenticação contínua, reconhecimento facial e/ou biometria multimodal. No Capítulo 4 é apresentada a metodologia que este trabalho seguiu para seus experimentos e os resultados destes. Para concluir, o Capítulo 5 discute sobre as considerações finais do trabalho e resultados alcançados com a conclusão do mesmo.

Capítulo

2

REFERENCIAL TEÓRICO

Apesar de ser uma área ativa há décadas, as possibilidades para realização de reconhecimento biométrico continuam a crescer. Isso ocorre devido à aparição de novos métodos para coletar informações dessas biometrias e às diversas possibilidades para reconhecer seus padrões (JAIN et al., 2004). Além disso, não há um reconhecimento biométrico ideal para atender a todas as necessidades, e as possibilidades de ataques maliciosos também tendem a crescer. Ainda assim, métodos tradicionais de autenticação baseados em itens ou senhas tornaram-se opções de segurança menos confiáveis, exigindo que os métodos de reconhecimento biométrico se tornem cada vez mais seguros.

Quando biometrias como impressão digital e íris são comparadas com faces por capacidade de distinção e precisão, reconhecimento facial não demonstra ser o método mais adequado. Por outro lado, a necessidade de contato com um sensor no caso de impressões digitais e o esforço para manter o olho visível no caso de reconhecimento de íris exigem uma maior cooperação por parte do usuário. Comparando as características do reconhecimento facial, como mostra a Tabela 2.1, é possível observar que outras biometrias podem ser mais viáveis para diversos tipos de sistemas. Já para a autenticação contínua, as demais biometrias demandam do usuário mais esforço para realizar autenticações sucessivas, enquanto o reconhecimento facial provê melhores condições de usabilidade, permitindo que o usuário realize outras atividades. Logo, para manter o nível de aceitabilidade e buscar um maior nível de segurança por meio da multibiometria, o reconhecimento contínuo de múltiplas características faciais é proposto nesse trabalho.

Para sistemas que realizam tanto verificação quanto identificação, de forma contínua ou não, o processo de reconhecimento facial precisa ser realizado adequadamente para atingir resultados satisfatórios. Para cada uma das propriedades faciais utilizadas, deve-se considerar as etapas de detecção facial, normalização, extração de características e correspondência de características, como mostra a Figura 2.1. Mais detalhes sobre as etapas do reconhecimento facial são apresentados nas Seções 2.1 a 2.4.

Tabela 2.1 Comparação entre biometrias, nivelando em alto (A), médio (M) e baixo (B) as características (universalidade, distinção, coletabilidade, desempenho, aceitabilidade e circunvenção) de cada uma para sistemas não contínuos (adaptado de Jain, Ross e Prabhakar (2004)).

Característica/ Biometria	Univ.	Dist.	Perm.	Colet.	Desem.	Aceit.	Circ.
Face	A	B	M	A	M	A	M
DNA	A	A	A	B	A	B	B
Impressão Digital	M	A	A	M	A	M	M
Íris	A	A	A	M	A	B	B
Retina	A	A	M	B	A	B	B
Pulso cardíaco	A	A	B	M	M	B	B

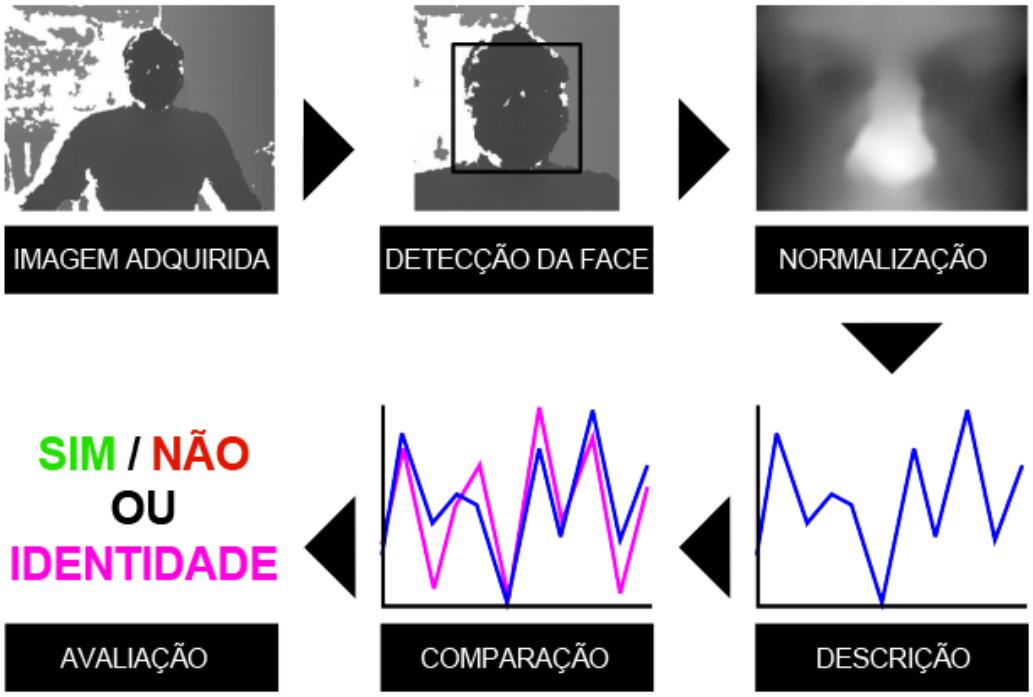


Figura 2.1 Sequência de passos básicos para o reconhecimento facial (adaptada de Segundo (2013)).

2.1 DETECÇÃO FACIAL

Uma vez que uma imagem é adquirida, esta etapa é a base para qualquer reconhecimento facial, pois para o funcionamento apropriado de todas as etapas seguintes é necessário localizar a face na imagem (ZHAO et al., 2003). Entre os possíveis métodos para detectar faces, o mais abordado, principalmente para imagens 2D, é a detecção facial em tempo real proposta por Viola e Jones (2004). Este método realiza uma varredura em cada região da imagem, chamada de sub-janela. Cada sub-janela é analisada por uma sequência de retângulos com diferentes dimensões, chamados de características de *Haar*, exemplificadas na Figura 2.2 [1]. Essas características são posicionadas na sub-janela e, quando combinadas, podem representar os contrastes presentes em uma face. Assim, a sub-janela segue linha por linha até varrer a imagem completamente, e diferentes tamanhos de sub-janelas são utilizados para garantir que faces em diferentes distâncias da câmera sejam detectadas.

Figura 2.2 Exemplo de diferentes tipos de características de *Haar* analisados em cada sub-janela

Todas as sub-janelas passam por diversos classificadores, que são formados por uma combinação de características de *Haar* e organizados em cascata. Apenas sub-janelas com resultados positivos para todos os classificadores são consideradas faces, como exemplifica a Figura 2.3. Se, em qualquer estágio desta cascata de classificadores, uma sub-janela for rejeitada, o classificador entende que ali não há uma face.

Um dos problemas do método proposto por Viola e Jones (2004) é o custo computacional de varrer determinadas imagens, como as de alta resolução, para procurar faces

[1] Imagem retirada de www.flickr.com/photos/unavoidablegrain/6884354772

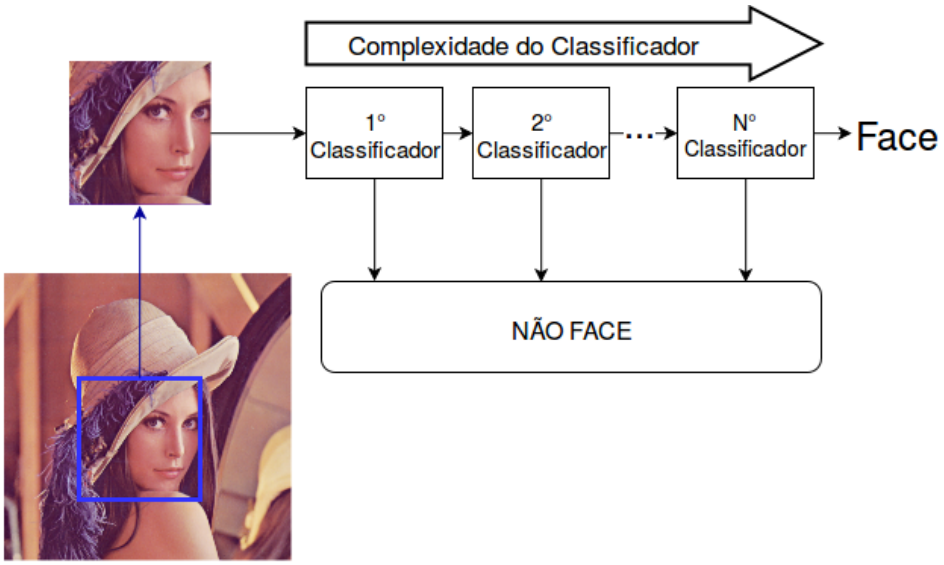


Figura 2.3 Representação do funcionamento da classificação em cascata, onde uma sub-janela representa uma face, se ela for aceita por todos os níveis da cascata.

em diferentes escalas. Uma solução para este problema foi proposto em Segundo (2013), adaptando o detector descrito para processar faces 3D em tempo real utilizando um sensor de baixo custo. A proposta foi o treino e uso do detector em imagens de projeção ortogonal (uma representação 2D das coordenadas 3D do mundo real) que mantêm as faces sempre em uma mesma escala, permitindo varrer as imagens utilizando apenas um tamanho de sub-janela.

Detectores de face no estado-da-arte propõem o uso de CNN para esta tarefa. O trabalho de Zhang et al. (2016) utiliza a combinação de 3 diferentes CNNs, criando um processo em cascata para detecção e normalização facial. A primeira é mais rápida e superficial sobre a imagem para achar janelas candidatas a face. A segunda é mais complexa e utilizada para refinar as janelas, eliminando as que não são faces. Por fim, a terceira faz um novo refinamento para encontrar a face, sendo mais poderosa que as outras. A proposta de Zhang et al. (2016), além de detectar a face, apresenta 5 pontos dela (*i.e.* centro dos olhos, ponta do nariz e cantos da boca), que são utilizados para o alinhamento facial na etapa de normalização, explicada na próxima seção. Tanto o método proposto por Segundo (2013) quanto o proposto por Zhang et al. (2016) serviram de base para as detecções faciais de nossos experimentos.

2.2 NORMALIZAÇÃO

Esta etapa do reconhecimento facial visa a padronização de imagens das faces a serem reconhecidas, tanto com cadastro quanto para identificação. Essa padronização pode ocorrer por meio do alinhamento da posição facial, redução de ruído, regularização de iluminação (para reconhecimento baseado em luz visível) e outras técnicas para melhorar a resolução ou dar realce às informações mais discriminantes da face, como exemplifica a Figura 2.4.

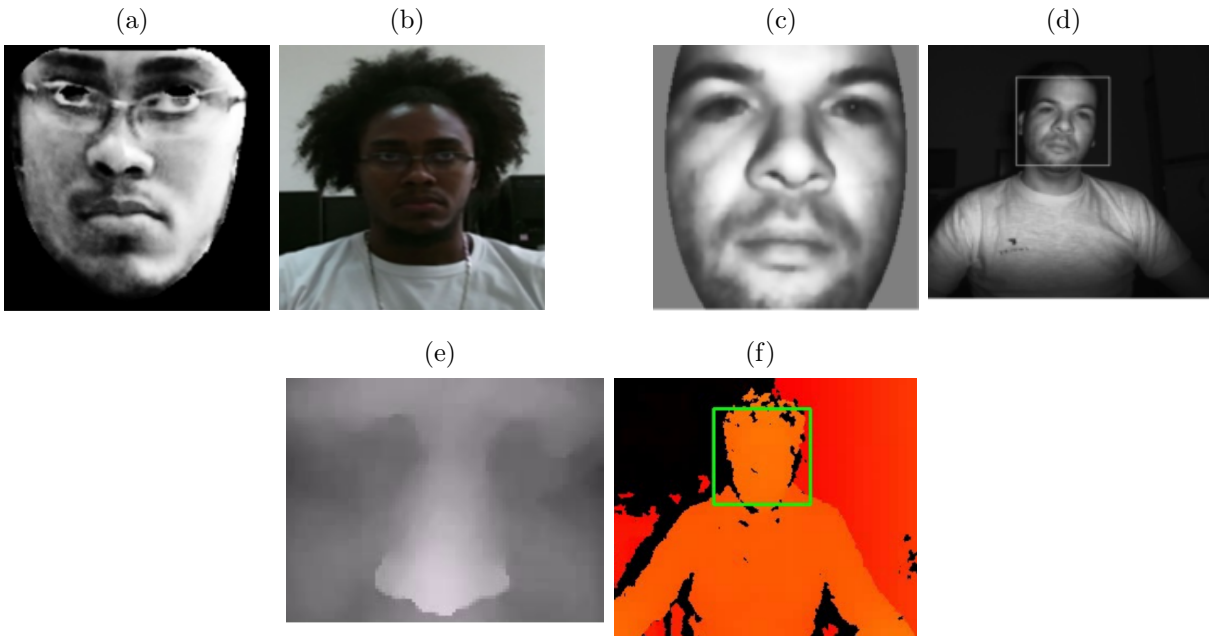


Figura 2.4 Exemplos de normalizações em imagens. (a) Ajuste de pose e iluminação em (b) imagem de textura (imagens adaptadas de Silva e Segundo (2015)). (c) Ajuste de pose em (d) imagem de infravermelho (imagens adaptadas de Magalhaes e Segundo (2015)). (e) Segmentação de região de interesse em (f) imagem de profundidade (imagens adaptadas de Segundo (2013)).

Os principais métodos de regularização da pose da face são: posicionamento padrão de pontos específicos da face (*e.g.* canto de olhos, ponta de nariz) (LU; JAIN, 2005; MIAN; BENNAMOUN; OWENS, 2007), transformação de um modelo de face deformável para representar uma face (KAKADIARIS et al., 2007) e mapeamento para um modelo genérico de face (CHANG; BOWYER; FLYNN, 2006; SEGUNDO, 2013). Os métodos de pré-processamento de imagens, como filtros para redução e ruídos (*e.g.* filtro gaussiano, filtro bilateral) e equalização de contraste (*e.g.* equalização de histograma, *CLAHE*), podem ser utilizados para alterar a qualidade da imagem de forma a melhorar as condições do reconhecimento.

2.3 EXTRAÇÃO DE CARACTERÍSTICAS

Realizar a comparação diretamente entre as imagens normalizadas é uma tarefa não recomendada por diversos trabalhos na literatura (ZHAO et al., 2003; BELHUMEUR; HESPANHA; KRIEGMAN, 1997) devido ao alto custo computacional que é exigido para tal, pois imagens as possuem alta dimensionalidade. Métodos descritores podem ser criados baseados nas imagens de forma a criar um novo modelo discriminante com menor dimensionalidade. Dentre os métodos de descrição mais utilizados na literatura, podemos destacar os métodos de transformação linear: Análise de Componentes Principais (PCA - *Principal components analysis*), proposto por Pearson (1901), sendo popularizado posteriormente para reconhecimento de faces com *eigenfaces* (TURK; PENTLAND, 1991), e

LDA, apresentada por Fisher (1936) (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). Para reconhecer faces, o uso do LDA tem vantagem sobre o uso do PCA por comprimir as informações de forma a maximizar a distinção entre classes, já que o PCA não foca nas informações mais discriminantes da imagem. Apesar de possibilitar uma melhor classificação, o LDA precisa de mais tempo e mais informações para o treino em relação ao PCA.

O PCA e LDA, além de poderem ser combinados, também podem ser aplicados para reduzir a dimensionalidade de outros descritores. Outras formas de descrição bem conhecidas, não só para reconhecimento por textura, mas também para imagens de profundidade, incluem o LBP (AHONEN; HADID; PIETIKAINEN, 2006), que é robusto a altas variações de iluminação, e Histograma de Gradientes Orientados (HOG - *Histogram Oriented Gradient*) (DALAL; TRIGGS, 2005), que analisa partes da imagem por meio de gradientes de intensidade e descreve formato e aparência das faces.

Contudo, foi com a evolução dos métodos de *Deep Learning* que a distância diminuiu drasticamente entre a capacidade humana e a capacidade de sistemas computacionais para o reconhecimento facial 2D, possibilitando o desenvolvimento de CNNs melhores que humanos para essa tarefa. Os trabalhos de reconhecimento facial 2D no estado-da-arte utilizam CNNs treinados com bases de imagens consideradas muito grandes (*e.g.* 500 mil para OpenFace (AMOS; LUDWICZUK; SATYANARAYANAN, 2016), 200 milhões para Facenet (SCHROFF; KALENICHENKO; PHILBIN, 2015)), permitindo a geração de descritores mais discriminativos e robustos a variações de pose e iluminação que os demais métodos de descrição citados nesta seção.

2.4 CORRESPONDÊNCIA DE CARACTERÍSTICAS

Uma vez que as características foram extraídas, é possível fazer uma correspondência entre os elementos do descritor gerado da face normalizada e dos descritores previamente armazenados de indivíduos que podem ser reconhecidos. No caso da verificação, o descritor é comparado com o descritor do usuário que ele diz ser, no caso da identificação, essa comparação é feita com os descritores de todos os usuários da base. O resultado dessas comparações deve ser um valor de similaridade entre os modelos. Quanto mais próximo o modelo adquirido for do modelo da base, maior a confiança de que aquelas informações são da mesma pessoa. Para sistemas que realizam verificação, o valor de similaridade deve superar um determinado limiar, que pode variar para estabelecer o nível de segurança do reconhecimento.

Para encontrar a similaridade entre os modelos biométricos, as abordagens mais utilizadas são medidas de distância, como Distância Euclidiana, Distância de Manhattan e Distância de Mahalanobis (DEZA; DEZA, 2009). As medidas de distância retornam o quão próximo um modelo está do outro se baseando em vetores de informação, histogramas ou imagens normalizadas que representem as faces. Há abordagens que correspondem diretamente os elementos da imagem facial para descobrir o quão se parecem (LU; JAIN, 2005; CHANG; BOWYER; FLYNN, 2006), mas esse é um processo demorado e custoso para comparações entre um usuário com todos os demais em uma base de dados, e também inviável para o reconhecimento contínuo.

2.5 RECONHECIMENTO MULTIMODAL

O reconhecimento biométrico multimodal é uma proposta para superar os pontos fracos que um modelo biométrico tem quando utilizado individualmente. Logo, para sistemas que sofrem com aquisição insuficiente de dados ou com características pouco discriminantes, integrar múltiplos modelos biométricos pode ser uma alternativa. O objetivo é integrar uma característica biométrica a outra para cobrir uma determinada limitação, como características que nem todos possuem ou características que permitam muitas possibilidades de sofrer ataques (JAIN; FLYNN; ROSS, 2007).

Como a Figura 2.5 demonstra, sistemas de reconhecimento multimodal podem ser elaborados a partir da integração em diferentes níveis:

- Múltiplas capturas: para assegurar que não houve falhas na primeira captura e na análise da mesma, mais amostras da mesma característica biométrica são coletadas para o reconhecimento.

- Múltiplas unidades: características biométricas que apresentam padrões diferentes em mais de uma parte similar do corpo (*e.g.* impressão digital, íris) podem ser utilizadas para o mesmo reconhecimento biométrico.

- Múltiplos métodos de classificação: combinar a classificação obtida por mais de um método de reconhecimento para um mesmo dado biométrico adquirido.

- Múltiplos dispositivos de entrada: o uso de mais de um sensor para capturar diferentes propriedades de uma mesma característica biométrica (*e.g.* reconhecimento facial por textura e forma da face).

- Múltiplas características biométricas: utiliza-se mais de um sensor para capturar dados de biometrias distintas simultaneamente.

Apesar de alguns sistemas multimodais também abrangerem métodos não-biométricos, levá-los em consideração retornaria ao problema de reconhecimento baseado em possuir ou saber algo (HONG; JAIN; PANKANTI, 1999). Para sistemas que utilizam reconhecimento biométrico multimodal, suas informações podem ser combinadas a nível de:

- Extração de características: as representações de cada biometria podem ser unidas para gerar uma característica nova, que deve ser utilizada no processo de reconhecimento e pode ser mais discriminante que as características biométricas individuais inicialmente obtidas.

- Correspondência de características: os níveis de similaridade gerados por cada característica utilizada são combinados, gerando uma nova similaridade para a classificação.

- Decisão: as decisões geradas por cada modalidade são combinadas de modo a gerar uma decisão mais segura.

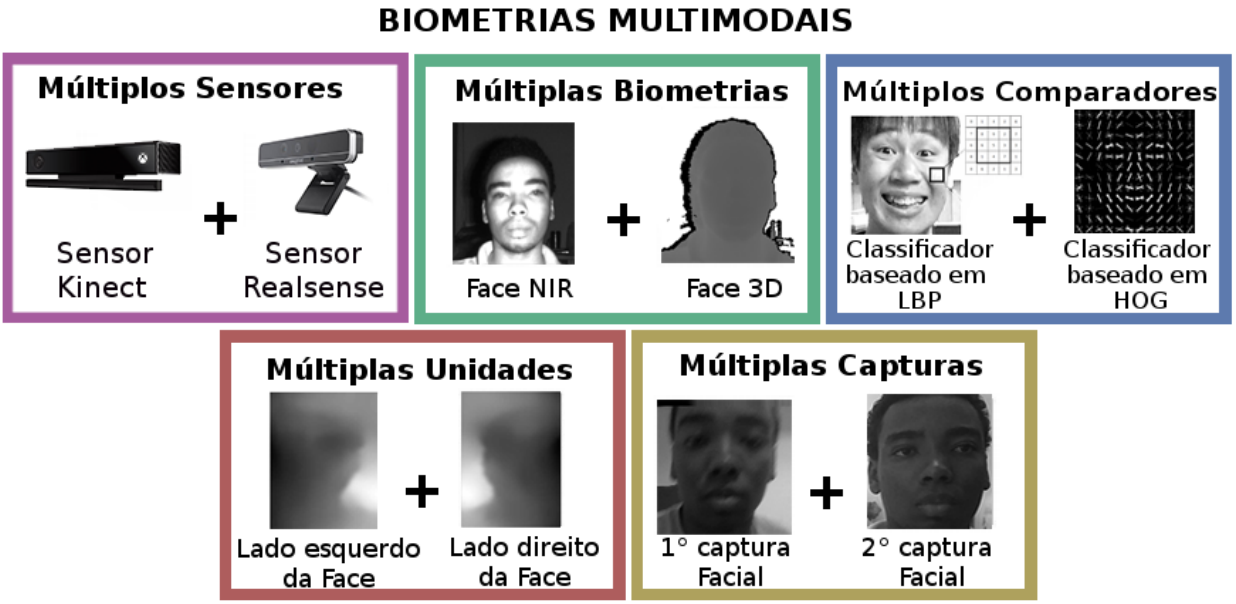


Figura 2.5 O uso de reconhecimento de biometria multimodal é realizado por meio da fusão de diferentes métodos de reconhecimento individuais.

Como podemos analisar na Figura 2.6, o reconhecimento de faces por meio de múltiplas características pode utilizar os três modelos de fusão de reconhecimento, principalmente por utilizarem técnicas de extração e comparação de características similares (ABATE et al., 2007).

2.6 AUTENTICAÇÃO CONTÍNUA

Sistemas de autenticação contínua verificam ou identificam o usuário continuamente. Para ambos os casos, o desenvolvimento costuma seguir uma sequência de passos:

- Registro: o sistema armazena os modelos biométricos obtidos de usuários autorizados juntamente com suas respectivas identidades para futuras comparações.
- Aquisição contínua: é fundamental coletar dados biométricos em uma frequência mínima a garantir que, entre duas coletas, seja possível detectar a mudança ou ausência do usuário autenticado antes de permitir um ataque. Se a aquisição falhar, o sistema pode ficar vulnerável a ataques ou interromper um acesso válido.
- Verificação contínua: para um indivíduo acessar sistemas com base em verificação contínua, como mostra a Figura 2.7, é necessário uma etapa de autenticação inicial. Utiliza-se a referência de quem este diz ser na busca dos dados do usuário na galeria, para depois realizar o reconhecimento e verificar se o acesso deve ser permitido. Após a autenticação inicial, o sistema deve utilizar os dados adquiridos continuamente para verificar se o usuário autorizado ainda está acessando-o. A verificação pode ser realizada com os dados biométricos do usuário na galeria ou com o modelo

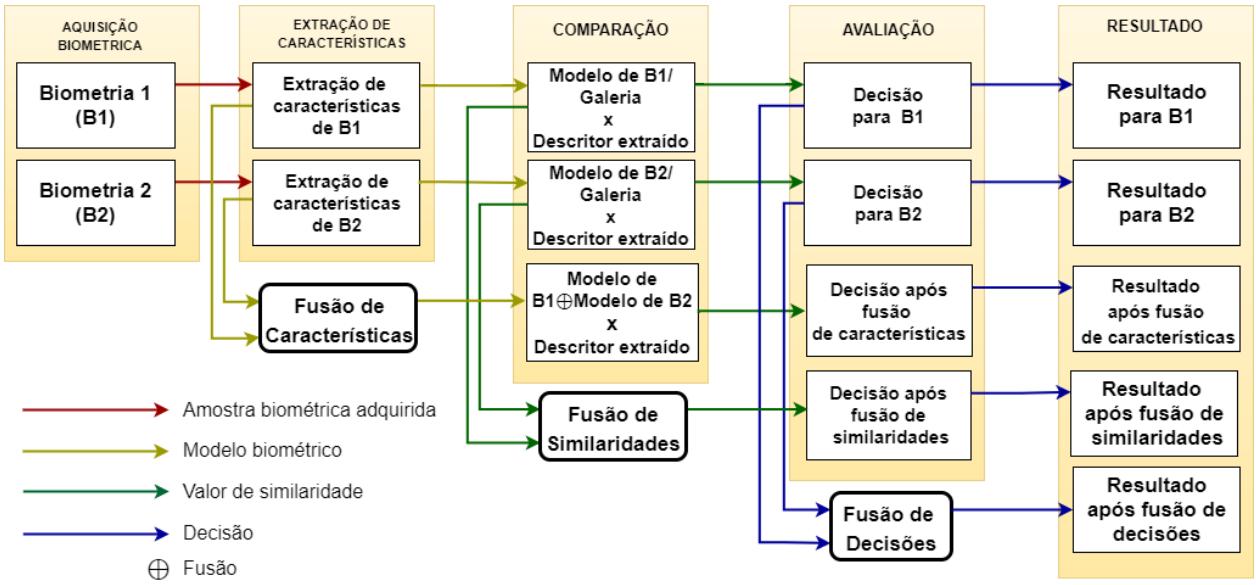


Figura 2.6 Fluxograma demonstrando possíveis fusões entre os diferentes módulos para um sistema baseado em múltiplas biometrias.

biométrico gerado durante a autenticação inicial. No entanto, nesta segunda opção a confiabilidade da verificação contínua será a mesma da autenticação inicial.

- Identificação contínua: o sistema, a cada aquisição contínua, compara a amostra biométrica com toda a base de dados, como demonstrado na Figura 2.8. A identificação contínua tende a ser mais custosa computacionalmente, mas é mais segura por não depender da autenticação inicial.

O uso de faces para autenticação contínua possui mais vantagens quando se compara o custo, a usabilidade e o desempenho das demais biometrias passíveis de serem utilizadas de forma contínua. Há sistemas e dispositivos de entrada que são adaptados de forma a permitir o uso de outras biometrias, como o mouse com sensor de impressão digital integrado utilizado por Sim et al. (2007). Mesmo assim, o módulo responsável pelo reconhecimento da impressão digital só pode ser reconhecido enquanto o mouse estiver em uso, o que não ocorreria quando o usuário estivesse digitando ou utilizando as duas mãos para outras atividades. Tais limitações fazem com que a maioria das biometrias seja utilizada como complemento em sistemas multimodais.

Biometrias comportamentais, mesmo para autenticação não contínua, precisam ser coletadas em um intervalo de tempo considerável, que pode levar segundos (*e.g.* ECG) ou até minutos (*e.g.* estilometria). Apesar do aspecto temporal da captura destas biometrias, utilizá-las para autenticação contínua também pode exigir muito esforço por parte do usuário (*e.g.* permanecer digitando para captura do ritmo de digitação), ser facilmente prejudicada por condições ambientes (*e.g.* ruídos na captura de reconhecimento de voz) ou ser prejudicado por alterações comportamentais devido ao estado físico e/ou emocional do indivíduo.

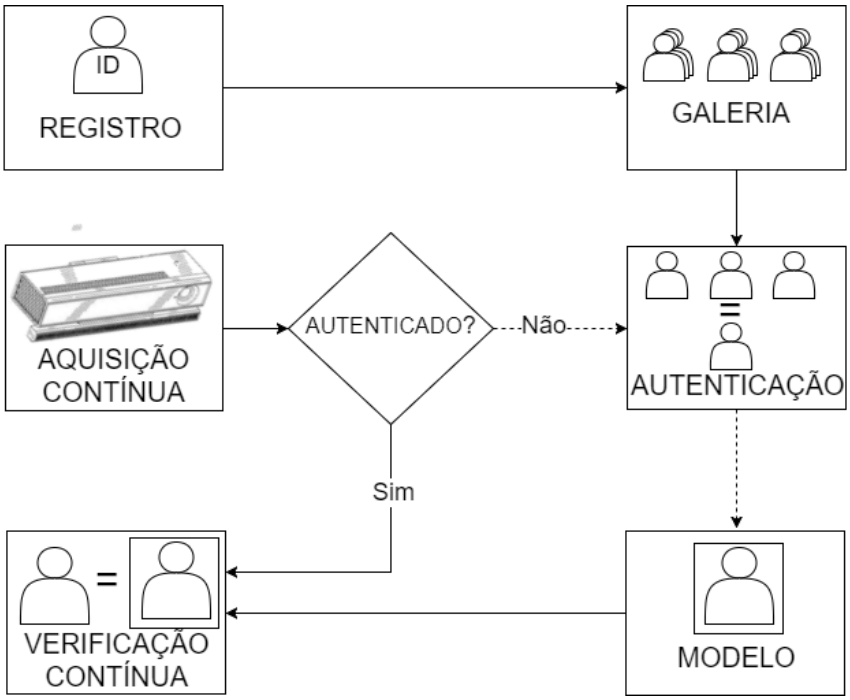


Figura 2.7 Para sistemas que utilizam verificação contínua, uma fase de autenticação inicial identificará o usuário com base nas imagens registradas na galeria. Após a autenticação, que pode ser realizada por meios tradicionais ou até outro tipo de reconhecimento biométrico, o usuário será constantemente verificado.

Ainda que a face seja a biometria mais adequada para sistemas de autenticação contínua, o método utilizado para o reconhecimento é que precisa ser rápido o suficiente para garantir o aspecto contínuo. Além do tempo de execução, o desempenho e segurança do sistema é determinado também por sua acurácia, que pode ser mensurado por:

- Taxa de falsa aceitação (*False accept rate* (FAR) ou *False match rate* (FMR)): representa a probabilidade de uma pessoa não autorizada ser autenticada. Essa taxa representa a vulnerabilidade do sistema, ou seja, para sistemas de alta segurança, ela precisa estar próxima de zero.

- Taxa de falsa rejeição (*False rejection rate* (FRR) ou *False non-match rate* (FNMR)): a probabilidade do sistema falhar em autenticar um usuário válido. Um alto valor de FRR pode representar que um usuário será frequentemente interrompido indevidamente durante o acesso.

- Taxa de erro igual (*Equal error rate* (EER)): representa a igualdade entre FRR e FAR. O sistema tende a ser mais preciso quanto menor for essa taxa.

- Taxa de falha de captura (*Fail to capture rate* (FTC)): a aquisição contínua é a base para sistemas de autenticação contínua. Logo, devido à importância desta etapa

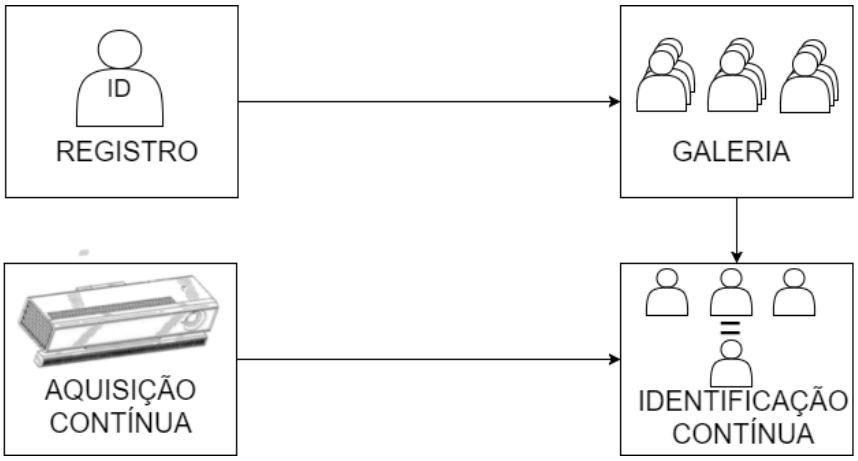


Figura 2.8 Em sistemas que utilizam identificação contínua, para cada captura da aquisição contínua, o usuário será autenticado se for reconhecido após a comparação com todos os usuários da galeria.

para o bom desempenho do sistema, o FTC deve ser o mais próximo possível de zero.

Capítulo 3

TRABALHOS RELACIONADOS

Apesar de ser uma área recente, existem diversos trabalhos sobre autenticação contínua baseados em diferentes características do usuário. A primeira pesquisa realizada dessa área foi a análise da forma de digitar do usuário (LEGGETT et al., 1991; MONACO et al., 2012), em que o ritmo da digitação é utilizado para verificar continuamente a identidade da pessoa. Posteriormente, muitos trabalhos utilizaram experimentos com biometrias comportamentais, por serem analisadas ao longo do tempo, como se fossem para autenticação contínua, mas não realizaram experimentos longos o suficiente para analisar a real viabilidade do sistema ou só faziam autenticação periodicamente (BOURS; MONDAL, 2015). Outros trabalhos utilizaram biometrias físicas para autenticação contínua, possuindo um tempo de aquisição de amostras mais rápido e maiores taxas de precisão, porém são mais vulneráveis a fraudes que biometrias comportamentais, pois biometrias físicas são consideradas mais fáceis de imitar (*e.g.* uso de máscaras, lentes ou impressões digitais falsas) (CAO; JAIN, 2016; LOUIS; KOMEILI; HATZINAKOS, 2016; ERDOGMUS; MARCEL, 2014).

3.1 AUTENTICAÇÃO CONTÍNUA

Quando biometrias são usadas em sistemas contínuos, os dois requisitos mais críticos dentre os descritos por Jain, Ross e Prabhakar (2004) são desempenho e coletabilidade. Os problemas de precisão podem levar a enganos com usuários genuínos ou impostores, enquanto um grande intervalo de tempo entre cada autenticação permite mais possibilidades de fraudes, e ambos os problemas anulam a eficácia de segurança de autenticação contínua. Como uma biometria precisa ser adquirida com frequência durante todo o processo de autenticação contínua, algumas biometrias podem apresentar problemas de coletabilidade, ao não estarem disponíveis ou exigindo um nível de cooperação do usuário que atrapalhe o uso do sistema.

Por meio da Tabela 3.1 podemos ver diversos trabalhos para cada biometria já utilizada individualmente para autenticação contínua, com exceção de faces, selecionados

Tabela 3.1 Trabalhos de autenticação contínua na literatura com base nos resultados e experimentos mais relevantes para diferentes biometrias. Além da biometria e referência do trabalho, são listados detalhes dos experimentos (tamanho de cada amostra contínua, frequência de cada autenticação e quantidade de usuários) e os principais resultados alcançados.

Biometria	**Autores**	**Tamanho**	**Frequência**	**Usuários**	**Resultado**
Impressão digital	Sim et al. (2007)	30min	$\sim 0,6\times$ por segundo, se disponível	11	∼12% FAR e ∼4% FRR em 3,1s
Olhar	Sui et al. (2012)	20 quadros		20	∼0,5% EER
PPG	Bonissi et al. (2013)	15 min	0.025 × por segundo	14	9% EER
ECG	Agrafioti, Bui e Hatzinakos (2012)	20-45 min	0,2 × por segundo	43	3,96% EER
EEG	Matsuyama, Shozawa e Yokote (2015)	400s	0,1 × por segundo	10	0,03% EER
Voz	Feng, Fawaz e Shin (2017)	30 comandos de voz	$> 0,1\times$ por segundo	18	0,09% FPR
Caminhar	Xu et al. (2017)	5 min × 2 sessões	0,2 × por segundo	20	8,4% EER
Digitação	Xi, Tang e Hu (2011)	1-15 amostras × 700-900 caracteres	0,09 × por segundo	205	1.65% FAR 2,75% FRR
Mouse	Feher et al. (2012)	entre 1 e 100 ações	A cada 30 ações (∼0,24 × por segundo)	25	8,53% EER
Toque	Shen et al. (2015)	800 ações com *touchscreen*	0,28 × segundo	51	6,17% FAR 3,385% FRR
Estilometria	Brocardo (2015)	500 caracteres	A cada 50 blocos de caracteres	76	8,21% EER
Bioimpedância	Martinovic et al. (2017)	A cada 20 amostras	100 bins de frequência	30	2% EER

com base na relação entre a relevância de seus experimentos e resultados encontrados em comparação a outros trabalhos para a mesma biometria.

Ao analisarmos experimentos de trabalhos para autenticação contínua utilizando biometrias comportamentais, é possível perceber que aplicações onde esses sistemas poderiam ser implantados são bem limitadas a ambientes onde o usuário já teria que utilizá-las com frequência. Para exemplificar, o trabalho de Feng, Fawaz e Shin (2017) propôs um sistema de autenticação contínua baseado em comandos para um assistente de voz para smartphones, utilizando um conjunto de comandos predefinidos para os testes. O trabalho de Xu et al. (2017) para autenticação contínua baseada na forma de caminhar também limita as ações dos usuários, realizando experimentos em terrenos diferentes, mas com uma rota predefinida. Outros trabalhos até permitem ações livres dos usuários, mas simulando cenários onde o usuário teria que utilizar a biometria continuamente. Nos experimentos de Xi, Tang e Hu (2011) para autenticação contínua utilizando padrões de digitação, os usuários estavam livres para digitar o que quisessem em um campo de até 780 caracteres, simulando a ação de enviar um email. O envio de emails também foi escolhido como ação para os experimentos de Brocardo (2015), com textos de 500 caracteres. No trabalho de Feher et al. (2012) os usuários tiveram que simular o uso do mouse durante um ambiente de trabalho. Nos experimentos de Shen et al. (2015) os usuários simulavam interações contínuas com o touchscreen até alcançar 800 ações.

A autenticação contínua baseada no comportamento natural e contínuo do corpo, como utilizando ECG, EEG e PPG não exige nenhuma colaboração específica do usuário. Além disso, por só depender do usuário estar vivo, estas características podem ser capturadas a qualquer momento, além de serem discretas e de possuir baixa circunvenção. Por outro lado, estas biometrias podem revelar mais informações do que apenas as necessárias para o reconhecimento biométrico, e o uso de sensores de captura, como exemplificado na Figura 3.1, pode ser desconfortável para o usuário. Os experimentos realizados nos trabalhos de ECG, PPG e EEG citados na Tabela 3.1 utilizam sensores ligados à testa, um cinto sensor preso ao peito e um oxímetro de pulso ligado à ponta do dedo respectivamente.

(a) ECG (b) PPG (c) EEG

Figura 3.1 Apesar de suas vantagens para segurança, sensores comuns de (a) ECG, (b) PPG and (c) EEG podem ser intrusivos ou inapropriados para cenários diários.

A resposta do corpo ou parte dele a uma corrente elétrica, chamada de bioimpedância,

pode ser medida como uma característica física e também já foi utilizada para autenticação contínua, apesar de haverem poucos trabalhos do tipo. Essa corrente e sua resposta devem ser medidas por sensores em contato com o corpo e podem ser afetadas por fatores externos (*e.g.* usuário estar molhado ou cortado próximo ao local onde o sinal é aplicado). No trabalho de Martinovic et al. (2017) experimentos foram realizados com usuários tocando em dois eletrodos: um para emitir o pulso elétrico e outro para receber, após passar pelo torso do usuário.

A autenticação contínua baseada em algumas características físicas também pode causar um desconforto no usuário por restringir seus movimentos. O reconhecimento por impressões digitais é uma biometria altamente distintiva (ALTINOK; TURK, 2003), mas tem como desvantagem a permanência do dedo do usuário sobre o sensor de captura. Logo, apesar de ser um método altamente discriminativo, não leva em consideração a usabilidade do processo. Sim et al. (2007) propuseram o uso de um sensor incorporado em um mouse (exemplificado na Figura 3.2) para capturar amostras de forma contínua sem prejudicar a usabilidade em termos de cooperação do usuário, e mesmo com o auxílio do reconhecimento facial, as impressões digitais foram cruciais para o resultado final. Eles foram capazes de manter os usuários genuínos autenticados durante aproximadamente 88% do tempo, enquanto detectaram cerca de 96% dos impostores em 3.1 segundos ou menos. No entanto, quando o mouse não estava em uso, nenhuma impressão digital podia ser capturada e o sistema tinha que confiar na outra biometria.

Figura 3.2 Demonstração de um sensor de impressão digital embarcado em um mouse.

A aquisição de íris não requer contato físico com o dispositivo de captura, mas requer uma certa quantidade de cooperação do usuário para obter uma precisão de reconhecimento razoável. Sui et al. (2012) simularam um sistema de autenticação de íris contínua, utilizando um conjunto de imagens de íris, usando-as como uma sequência contínua de amostras. Embora a precisão relatada tenha sido alta (aproximadamente 0,5 EER), eles não avaliaram o quão rápido impostores poderiam ser detectados, e por não usarem sequências reais de imagens, os desafios de uma aplicação real não estavam bem representados em seus experimentos.

3.2 RECONHECIMENTO FACIAL PARA AUTENTICAÇÃO CONTÍNUA

Através de reconhecimento facial é possível identificar um usuário sem contato físico, sem ser intrusivo e sem obrigá-lo a agir de forma diferente durante o acesso (ZHAO et al., 2003; SEGUNDO, 2013). Além disso, faces possuem múltiplas características distintivas que podem ser consideradas para a autenticação contínua.

O método mais comum para reconhecimento facial é o baseado em textura. Utilizar esta modalidade é interessante devido à quantidade de algoritmos consolidados, tanto para detecção quanto para reconhecimento (VIOLA; JONES, 2004; ZHAO et al., 2003). Outra vantagem é a facilidade para aquisição de imagens 2D, pois câmeras possuem um baixo custo e atualmente podem ser encontradas em grande parte dos computadores pessoais. Em contrapartida, este tipo de reconhecimento é facilmente afetado por iluminação e pose (ZHAO et al., 2003; ABATE et al., 2007). Janakiraman et al. (2005) é um exemplo da eficácia desta modalidade, demonstrando usabilidade de um sistema protegido por autenticação contínua baseado em reconhecimento facial 2D. O trabalho de Silva e Segundo (2015) é outro exemplo, incluindo testes com fraudes após a autenticação inicial. As imagens do trabalho foram obtidas através do sensor *Microsoft Kinect for Windows v2*[1], com 1920×1080 pixels de resolução. O algoritmo utilizado para o reconhecimento foi o LBP com o qual obtiveram cerca de 8% de EER.

Outro possível método para o reconhecimento facial faz uso de imagens em infravermelho, que tem como principal vantagem ser invariante a iluminação ambiente (LI et al., 2007; MAGALHAES; SEGUNDO, 2015). Além disso, é possível encontrar sensores capazes de capturar imagens em infravermelho com preços acessíveis. Devido a fatores como este, o uso de câmeras infravermelho tem crescido na área de reconhecimento facial (HASSABALLAH; ALY, 2015). Magalhaes e Segundo (2015) desenvolveram um sistema de autenticação contínua baseado em reconhecimento facial por infravermelho, também utilizando o sensor Kinect v2, alcançando aproximadamente 7% de EER em seus experimentos.

Por fim, com sensores de profundidade, é possível utilizar a forma do rosto para reconhecimento facial, sofrendo pouca influência de iluminação e pose da face (SEGUNDO et al., 2013; ABATE et al., 2007; BOWYER; CHANG; FLYNN, 2006). Um sistema de autenticação contínua baseado em reconhecimento facial 3D foi desenvolvido por Segundo et al. (2013), alcançando 0,8% EER nos testes realizados. O sensor utilizado foi o *Microsoft Kinect for Windows v1*[2], e o descritor para o reconhecimento foi o HOG.

Outros trabalhos na literatura já propuseram a fusão de faces e outras biometrias para autenticação contínua, como listado na Tabela 3.2, no entanto, nenhum combinou múltiplas modalidades faciais.

O trabalho de Sim et al. (2007) realiza uma combinação da verificação de impressão digital e face para tornar a autenticação mais precisa, apesar do usuário precisar manter o dedo no sensor para reconhecer a impressão digital. Altinok e Turk (2003) foi ainda mais além, incluindo voz, impressão digital e face para aumentar a segurança. No entanto, além da limitação da impressão digital, o reconhecimento de voz só é possível enquanto

[1] https://developer.microsoft.com/en-us/windows/kinect/hardware
[2] http://www.xbox.com/pt-BR/Kinect/Home-new

Tabela 3.2 Lista de trabalhos de autenticação contínua fundindo faces com outras biometrias, citando os autores, ano de publicação e as biometrias fundidas a face.

Autores	**Biometrias fundidas**
Altinok e Turk (2003)	Voz e impressão digital
Sim *et al.* (2007)	Impressão digital
Azzini *et al.* (2008)	Impressão digital
Niinuma *et al.* (2010)	Biometrias suaves
Raja *et.al* (2014)	Impressão digital e cor de roupa
Crouse*et.al* (2015)	Unidade de Medição Inercial
Wang *et.al* (2016)	EEG
Srivastava e Sudhish (2016)	Dinâmica de digitação
Schiavone *et.al* (2016)	Dinâmica de digitação e impressão digital
Shen *et.al* (2016)	Dinâmica de digitação e cor de pele

o usuário estiver falando e ainda pode ser prejudicada por diversos fatores (*e.g.* ruído ambiente, condições físicas do usuário). Niinuma, Park e Jain (2010) desenvolveram um sistema de autenticação contínua multimodal integrando reconhecimento facial e cor da roupa, mas, além de sofrer com iluminação, biometrias suaves como a cor da roupa são mais vulneráveis a ataques.

Alguns trabalhos na literatura propuseram o uso de reconhecimento facial combinando informações faciais adquiridas por diferentes tipos de sensores, mas não utilizam autenticação contínua. Muitos destes trabalhos demonstraram um ganho significativo para suas fusões em relação às modalidades separadas. Chang, Bowyer e Flynn (2003) demonstraram que integrar imagens de faces 3D e 2D pode gerar um reconhecimento mais preciso que utilizando as modalidades separadas. Os testes realizados para o reconhecimento de 275 pessoas alcançaram uma taxa de 89.5% para imagens de textura, 92.8% para imagens de profundidade e 98.8% combinando os métodos. Bowyer et al. (2006) usou soma ponderada para combinar as similaridades de imagens de faces 2D, 3D e térmicas, obtendo o melhor resultado ao fundir as três modalidades. Ding e Tao (2015b) melhoraram a precisão de um sistema de reconhecimento de face 2D baseado em CNN usando a concatenação de características extraídas de diferentes regiões da imagem de face de entrada e de uma face frontal reconstruída. Para lidar com um problema da alta dimensionalidade, eles usaram redes neurais de *autoencoder* para reduzir o tamanho do vetor de características resultantes. Logo, mesmo que ainda não avaliada para a autenticação contínua, a fusão de modalidades faciais é demonstrada como uma opção melhor que o uso das modalidades separadamente. Para demonstrar essa viabilidade, realizaremos comparações entre experimentos de autenticação contínua utilizando 2D, 3D e NIR separadamente e a fusão entre as modalidades.

3.3 DISCUSSÃO

Como mencionado anteriormente, o uso de multibiometria pode aperfeiçoar a acurácia do sistema de autenticação, além de evitar fraudes mediante às limitações de cada biometria. Apesar dos resultados positivos demonstrados na Seção 3.2 pelos sistemas de reconhecimento utilizando uma propriedade facial, desenvolver um sistema de reconhecimento contínuo integrando múltiplas propriedades faciais podem ser mais eficaz e eficiente (ABATE et al., 2007; BOWYER; CHANG; FLYNN, 2006).

A maioria dos trabalhos que estudam a eficácia da autenticação estática focam na variabilidade entre pessoas, ao invés da diversidade de imagens faciais que uma mesma pessoa pode apresentar, razão que leva a criação de diversas bases de imagens onde as faces foram capturadas sob condições controladas. Para a autenticação contínua, além de garantir a segurança contra fraudes, os testes devem ser realizados com usuários agindo de forma natural, o que resulta em variações consideráveis nas imagens faciais de uma mesma pessoa.

Nenhum trabalho no estado da arte lida com todas as variações que faces podem gerar. Para tentar lidar com isso, alguns trabalhos integram múltiplas propriedades faciais para reconhecimento, mostrando que as modalidades de reconhecimento combinadas são mais robustas que as individuais, sendo mais viáveis para aplicações no mundo real. Dentre estes trabalhos, a maioria integra informações 2D e 3D, mesclando os resultados finais de cada abordagem para a classificação. Até onde sabemos, não há nenhum trabalho que utilize imagens de cor, profundidade e infravermelho para a realização da autenticação contínua.

Capítulo

4

AUTENTICAÇÃO CONTÍNUA MULTIMODAL

4.1 AQUISIÇÃO, NORMALIZAÇÃO E DESCRIÇÃO

Muitos trabalhos na literatura utilizam sensores RGB-D de baixo custo para diferentes experimentos de reconhecimento facial (MIN; KOSE; DUGELAY, 2014; SEGUNDO, 2013; ERDOGMUS; MARCEL, 2014). Para nosso sistema, as capturas contínuas de textura, 3D e NIR utilizando um sensor RGB-D são feitas simultaneamente. Cada imagem facial adquirida nessas três modalidades é normalizada girando a imagem de modo que os centros dos olhos se alinhem horizontalmente, para então serem cortadas a 128×128 pixels, deixando 48 pixels entre os olhos e a boca, e 40 pixels entre os olhos e a borda superior da imagem, como descrito por Wu, He e Sun (2015).

Então, as faces normalizadas de cada modalidade são representadas através de um descritor de 256 dimensões, usando a CNN (modelo C) disponibilizada publicamente por Wu, He e Sun (2015). Este modelo de CNN foi escolhido com base nos experimentos de Dahia, Santos e Segundo (2017), onde 3 CNNs com resultados comparáveis aos do estado-da-arte para reconhecimento facial 2D e disponibilizadas publicamente foram comparadas para o reconhecimento de imagens NIR e 3D. Este modelo foi treinado usando mais de 5 milhões de imagens da base de dados de CASIA-WebFace (YI et al., 2014) e MS-Celeb-1M (GUO et al., 2016). Como em seus experimentos, nós comparamos descritores usando a distância cosseno para obter pontuações correspondentes. Finalmente, estas pontuações são usadas para decidir se duas imagens pertencem ao mesmo sujeito ou não (*e.g.* coincidindo com genuíno ou impostor). Wu, He e Sun (2015) reportaram uma precisão de 98,8% no LFW (somente imagens 2D) com esta CNN, e também mostraram que sua abordagem poderia atingir resultados razoáveis em comparações de domínios cruzados (*e.g.* cor versus infravermelho). Também utilizando este modelo, Dahia, Santos e Segundo (2017) reportaram cerca de 7% de EER para 3D utilizando uma base controlada com 4,950 imagens e cerca de 2% para NIR em uma base semi-controlada com 17,580 imagens.

Até onde sabemos, não há modelos CNN publicamente disponíveis para imagens 3D ou NIR, nem grandes bases de dados que poderiam permitir criar tais CNNs do princípio.

Portanto, imagens 3D e NIR foram convertidas para imagens com uma escala de cinza de 8-bit e foram descritas usando o mesmo CNN usado paras as de cor, o que pode ser considerado um processo de transferência de aprendizagem.

Para a autenticação contínua, todo o processo, desde a aquisição das três modalidades de imagens até a comparação entre os descritores, é feito continuamente desde que o usuário esteja diante do sensor. A primeira imagem capturada de cada modalidade é utilizada como modelo para as comparações seguintes. Uma vez definido qual o melhor método para combinar as modalidades, esse processo é feito de forma multimodal, e o resultado final da comparação é utilizado para medir o quão seguro o sistema está em um dado momento após a autenticação inicial.

4.2 FUSÃO DAS MODALIDADES

Após adquirirmos os descritores faciais de cada modalidade, estas podem ser fundidas em diferentes etapas. Para definirmos qual abordagem é a mais adequada para fusão, realizamos experimentos utilizando métodos bem conhecidos na literatura em diferentes níveis. Para combinar características, a concatenação de caraterísticas é o método mais utilizado para este nível de fusão, que consiste em unir os vetores de descrição, gerando um vetor com maior dimensionalidade e mais discriminativo (JAIN; NANDAKUMAR; ROSS, 2005).

Para a fusão de similaridades, as distâncias entre os descritores de cada modalidade podem ser fundidas por meio da soma, produto, mínimo e máximo dessas similaridades, gerando assim uma nova similaridade que represente a fusão. A soma de similaridades é o método mais comum na literatura para esse nível, sendo mais utilizado de forma ponderada (JAIN; NANDAKUMAR; ROSS, 2005).

Para a etapa de fusão de decisões, o uso dos operadores AND e OR e da regra da Maioria Absoluta são bem utilizados em experimentos de verificação, cuja resposta é verdadeiro (genuíno) ou falso (impostor). A maior diferença entre os métodos está no impacto que as modalidades possuem na decisão final.

4.2.1 Soma de Características

Como usamos o mesmo CNN para descrever imagens 2D, 3D e NIR, os descritores obtidos têm as mesmas dimensões e uma faixa similar de valores. Além disso, o uso da distância cosseno e o treinamento do CNN baseado no *softmax loss* sugere que o espaço de características para cada modalidade se aproximam de uma hiperesfera. Com estas observações em foco, desenvolvemos uma hipótese de que somar vetores poderia gerar resultados satisfatórios para fundir biometrias no nível de características por duas razões:

1. O vetor gerado vai resultar em um espaço de característica estendido, porém a proporção entre a variação intraclasse e a variação do espaço de característica será reduzido. Isto pode ser demonstrado ao assumir que nós temos dois espaços de característica para um mesmo sujeito representado como duas variáveis aleatórias X e Y com uma distribuição normal, o que é plausível dado que o *softmax loss* foi usado durante o treinamento. Se eles fossem independentes, a soma dessas distribuições

Z teriam desvio padrão $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$, que é estritamente menor que $\sigma_x + \sigma_y$ (EISENBERG; SULLIVAN, 2008). Em outras palavras, o espaço de característica resultante seria sempre menor que a soma da variação nos espaços de característica originais. Todavia, como X e Y não são necessariamente independentes no nosso caso, a distribuição resultante talvez não seja normal, mas ainda esperamos que tenha menos variação intraclasse.

2. O vetor resultante da soma de vetores tem o mesmo tamanho que o vetor de característica original, como ilustrado na Figura 4.1(a), então não teremos problemas com alta dimensionalidade.

Além disso, nós podemos explorar a soma de vetores para mais do que dois vetores de características como mostrado na Figura 4.1(b), assim como a soma de vetores ponderada, como sugerido por Reddy et al. (2016) para características de face e impressão digital descritas por LBP.

(a) Soma de dois vetores. (b) Soma de múltiplos vetores.

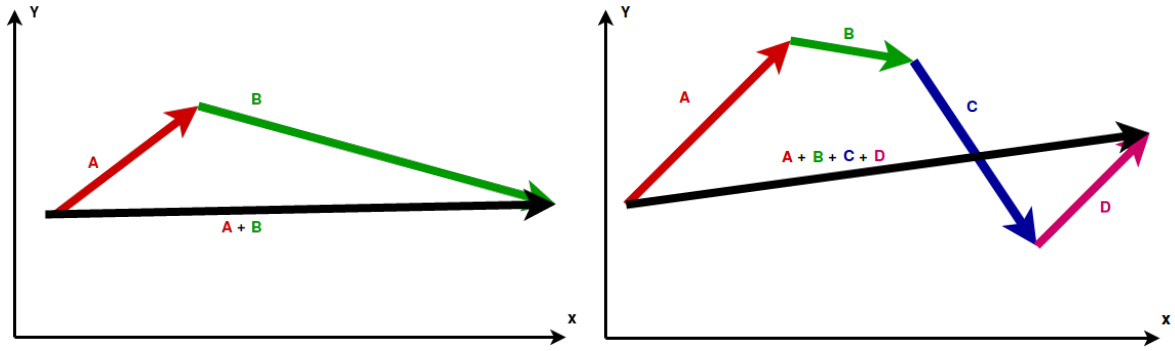


Figura 4.1 Representação do tamanho de vetores resultantes da soma entre vetores.

4.3 AUTENTICAÇÃO CONTÍNUA DO USUÁRIO

Para avaliar a segurança do sistema ao longo do tempo, utilizamos o P_{safe} sugerido por Segundo (2013), baseado em um modelo proposto por Sim et al. (2007), onde determina-se a probabilidade do sistema estar seguro no tempo t, considerando um histórico de observações Z_t. Cada observação $z_i \in Z_t$ corresponde à similaridade entre as faces adquiridas e o modelo de faces do usuário no instante i. Para calcular o P_{safe}, só utilizamos as probabilidades mais recentes do sistema estar seguro para computar a atual, não utilizando diretamente as observações mais antigas. Para calcular o P_{safe} utilizamos a Equação 4.1:

$$P_{safe} = \frac{2^{\frac{-\Delta t}{k}} \times P(safe|Z_t))}{\sum_{x \in X} P(x|Z_t))} \tag{4.1}$$

onde k é a taxa de decaimento que define quão rápido o sistema esquece antigas observações, Δt é o tempo decorrido desde a última observação Z_t, $X = safe, \neg safe$ e u

é o tempo da última observação antes de t. Assim, o $P(safe|Z_t)$ é obtido pela Equação 4.2:

$$P(safe|Z_t) = P(z_t|x) + 2^{\frac{(u-t)}{k}} \times P(x|Z_u) \tag{4.2}$$

$P(z_i|safe)$ e $P(z_i|\neg safe)$ são dados pelas Equações 4.3 e 4.4:

$$P(z_i|safe) = 1 - \frac{1}{2}\left[1 + erf(\frac{similaridade - \mu_{seguro}}{\sigma_{seguro} \times \sqrt{2}})\right] \tag{4.3}$$

$$P(z_i|\neg safe) = 1 - \frac{1}{2}\left[1 + erf(\frac{similaridade - \mu_{\neg seguro}}{\sigma_{\neg seguro} \times \sqrt{2}})\right] \tag{4.4}$$

sendo que para a primeira autenticação o sistema assume que está totalmente seguro, então $P(safe|Z_0) = 1$ e $P(\neg safe|Z_0) = 0$.

Os valores de $\mu_{seguro}, \mu_{\neg seguro}, \sigma_{seguro}$ e $\sigma_{\neg seguro}$ foram gerados por meio de experimentos com 11 indivíduos, possuindo 1111 imagens faciais de cada, para cada modalidade. Por vez, cada imagem, ou a fusão delas, era selecionada como modelo do indivíduo e comparada a todas as outras. Quando a comparação era da mesma pessoa, a similaridade gerada ia para o grupo de seguras, e caso o contrário ia para não-seguro. Assim foi definida a média e o desvio padrão das similaridades, atribuídas a μ e σ, para os estados seguro e não-seguro.

Como o foco dos experimentos é demonstrar o desempenho das modalidades para a autenticação contínua, os quadros contínuos para cada usuário não consideram a situação em que ele não estivesse sendo capturado, o que exigiria uma investigação mais aprofundada sobre os parâmetros para a análise temporal que melhor se encaixasse aos nossos experimentos. Assim, definimos nosso tempo de decaimento como 0.999, já que não há decaimento por falta de aquisições faciais.

Capítulo 5

EXPERIMENTOS E RESULTADOS

5.1 RECONHECIMENTO FACIAL MULTIMODAL

Neste experimento, para definirmos qual a melhor abordagem para fundir as modalidades de imagens faciais para a autenticação contínua, imagens de faces 2D, 3D e NIR, com as respectivas resoluções de 1280×720, 640×480 e 640×480 pixels, foram adquiridas pelo Intel Realsense SR300 [1], que pode ser considerado um dispositivo de baixo custo (*e.g.* o preço de uma unidade custa por volta de US$150,00). Embora o alcance da profundidade seja maior que um metro, faces a partir desta distância dificilmente produzem bons resultados de reconhecimento.

Após todo o processamento das amostras adquiridas, nós terminamos com três vetores de características de 256 dimensões que representam cada modalidade da face de entrada. Esses vetores de características ou a pontuação obtida após combiná-las com outras imagens podem ser fundidas de múltiplas formas, então nós exploramos os métodos de fusão mais comuns na literatura para encontrar as melhores relações entre eficiência e eficácia.

5.1.1 Base de Dados de Face Multimodal

Nossa base de dados contém 1,845 imagens de face em textura, profundidade e NIR simultaneamente capturadas de 96 indivíduos diferentes. A aquisição foi realizada usando uma aplicação no mundo real para acompanhar a ausência de alunos em diversos cursos, durante um período de quatro meses. Isso ocorreu em tempos diferentes do dia e em diferentes locais, de forma que a luz ambiente não fora controlada. Nenhuma instrução sobre expressões faciais ou acessórios para a cabeça foram dados aos sujeitos. Eles apenas foram instruídos a sentar em frente à câmera com uma distância de um metro, no máximo. Em alguns casos, entretanto, os sujeitos estavam longe o suficiente para apresentar dados de profundidade altamente ruidosos, como exemplificado na Figura 5.1. Outros fatores que dificultaram o reconhecimento na nossa base de dados foram as expressões faciais diferentes (ver Figura 5.2), pequenas variações de pose e variações ao longo do tempo (ver Figura

[1] https://software.intel.com/en-us/realsense/sr300camera

Figura 5.1 Exemplos de problemas em imagem de profundidade causados pela distância, que resultam em ruídos ou dados perdidos.

5.3). Para prevenir fraudes usando fotos impressas ou telas de smartphones/tablets, as detecções de face foram feitas usando somente imagens de profundidade.

5.1.2 Resultados Monomodais

Todos os experimentos foram feitos usando uma comparação de todos-contra-todos entre as imagens da base de dados acima, totalizando 1,701,090 combinações. Para ter como base, os resultados do reconhecimento monomodal são mostrados na Figura 5.4. Como esperado, já que as imagens foram treinadas para textura, a comparação desta modalidade tem o melhor desempenho, alcançando 1.12% EER, mostrando o quão bem o método de Wu, He e Sun (2015) se desempenha no nosso cenário de aquisição. Imagens NIR atingiram 5.08% EER, que é inferior do que os resultados de textura, mas é aceitável, considerando que a resolução de imagens NIR é inferior e a CNN de Wu, He e Sun (2015) não foi treinada para elas. Imagens de profundidade obtiveram o inferior resultado, principalmente por causa das limitações do sensor, com 35,33% EER. Outra possível razão é um menor poder discriminativo das características de profundidade em relação ao descritor usado, já que a correlação entre textura e profundidade parece menor que entre textura e NIR. Ainda assim, essas características com menor poder discriminativo podem vir a complementar os resultados multimodais. Embora os resultados de profundidade possam ser desencorajadores, profundidade ainda é importante para prevenir fraudes, para estimação/normalização da pose, e pode ainda auxiliar o reconhecimento.

5.1.3 Resultados da Fusão de Características

Os experimentos para a fusão em um nível de característica entre diferentes combinações de modalidades foram realizados usando a soma de características e a concatenação de caraterísticas. Os resultados são mostrados nas respectivas Figuras 5.5(a) e Figura 5.5(b).

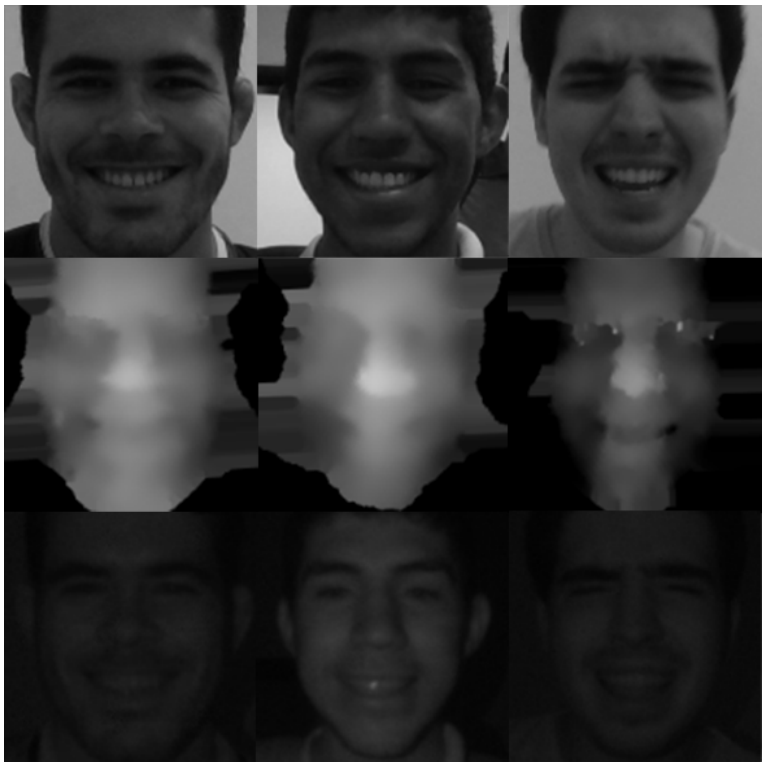


Figura 5.2 Exemplos de diferentes expressões faciais em nossa base de imagens para as três modalidades.

Figura 5.3 Exemplos de mudança em imagens de faces ao longo do tempo, como o crescimento da barba e acessórios.

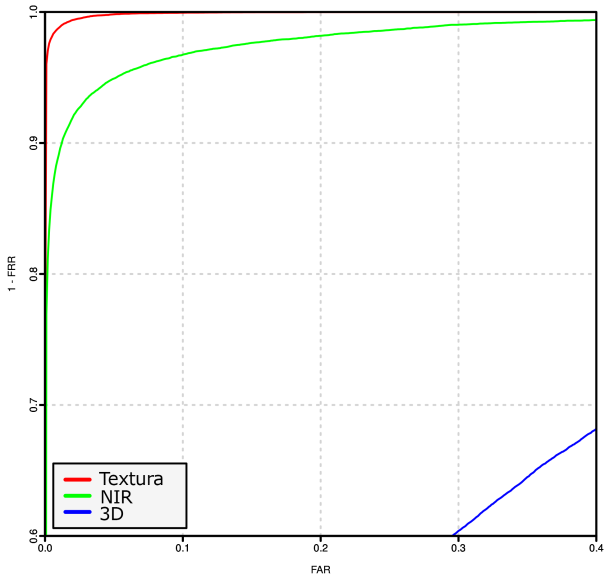


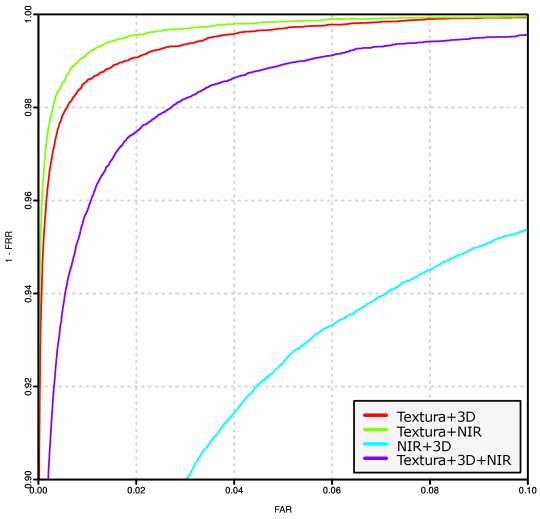
Figura 5.4 Curvas ROC para reconhecimento facial monomodal usando imagens de textura, 3D ou NIR.

Em ambos os experimentos, a fusão entre as características de textura e NIR alcançaram os melhores valores de similaridades (*i.e.* melhor do que somente textura na Figura 5.4), com 0,92% EER para concatenação e 0,78% EER para soma. A principal diferença entre estes métodos é na fusão de todas as modalidades, com a soma de característica sendo muito menos afetada por profundidade do que por concatenação, e, consequentemente, mais perto do resultado da fusão entre textura e NIR. Por este motivo, nós investigamos mais a fundo uma soma ponderada de características tentando todas as ponderações possíveis na faixa de [0,3] com um passo de 0,1, e conseguimos uma precisão de 0,67% EER para os pesos de 1,8, 0,8 e 1 para características 2D, 3D e NIR, respectivamente. As ROCs resultantes para todas as somas ponderadas de características são mostradas na Figura 5.5(c). Como pode ser observado, todas essas curvas ROC (*Receiver Operating Characteristic*) superam o resultado da concatenação de característica, mostrando que a soma ponderada de características é uma opção razoável de fusão, já que ela obtém resultados mais precisos enquanto preserva a dimensionalidade original.

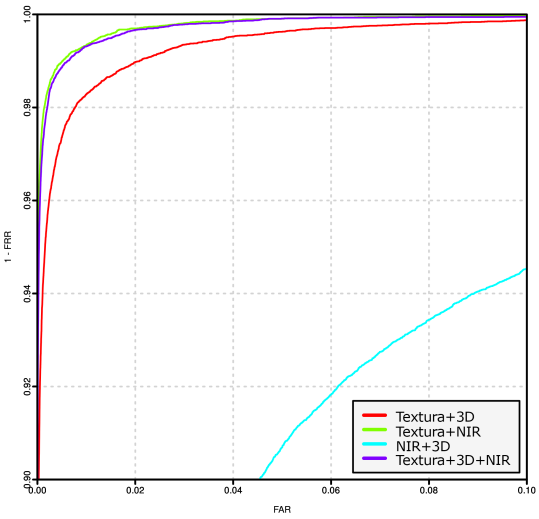
5.1.4 Resultados da Fusão de Similaridade

Para a fusão de similaridade, as regras bem conhecidas de soma, produto, mínimo e máximo foram aplicadas às pontuações de correspondência monomodais, e os resultados são mostrados nas Figuras 5.6(a), 5.6(b), 5.6(c) e 5.6(d). Entretanto, nenhuma destas técnicas de fusão de similaridades foram capazes de superar o melhor resultado do monomodal. Novamente, a combinação das modalidades de textura e NIR alcançaram melhores resultados do que outras combinações para todos os métodos de fusão avaliados, mas sua EER no melhor caso é por volta de 1% pior que a modalidade de textura na

(a) Concatenação de características



(b) Soma de características



(c) Soma ponderada de características

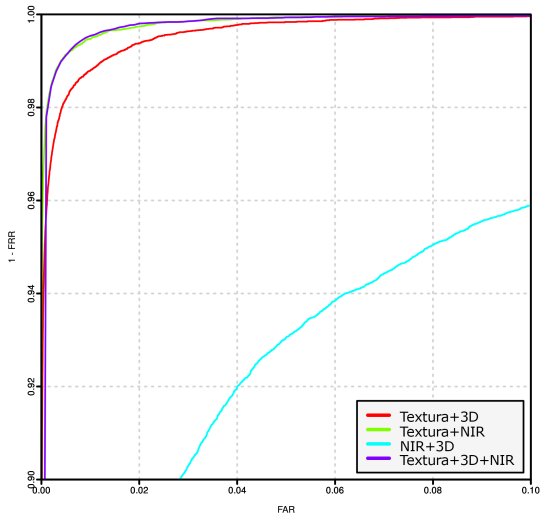


Figura 5.5 Resultados de fusão a nível de características para (a) concatenação de características, (b) soma de características e (c) soma ponderada de características.

Figura 5.4.

Como a soma de similaridades obteve os melhores resultados, nós investigamos o método de soma ponderada de similaridades, e o resultado é apresentado na Figura 5.7. Como nos nossos experimentos ponderados anteriores, nós tentamos todas as possíveis ponderações na faixa de [0,3] com um passo de 0,1, e a melhor precisão alcançada foi 0,91% de EER para as modalidades de 2D e NIR, com pesos de 0,3 e 1, respectivamente. Com a soma de similaridades ponderada, a fusão em nível de similaridade foi capaz de superar a melhor performance monomodal, mas ainda é menos efetiva do que a fusão no nível de característica.

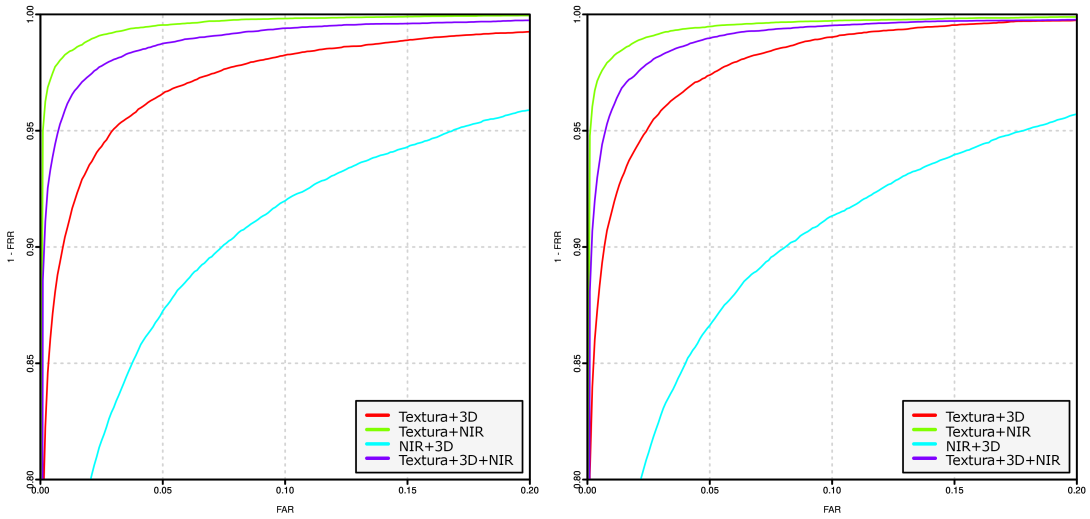
5.1.5 Resultados da Fusão de Decisões

Para a combinação de resultados no nível de decisão, os métodos para fusão escolhidos foram os operadores OR e AND e a regra da Maioria Absoluta. O operador OR alcançou os melhores resultados, como apresentado na Figura 5.8(a). Os resultados para o operador AND e a regra da Maioria Absoluta são mostrados nas Figuras 5.8(b) e 5.8(c), respectivamente, e ambos não foram capazes de superar o operador OR em nenhuma combinação de diferentes modalidades. O operador OR também atingiu um EER de 6,14% para a combinação das modalidades de 3D e NIR, que é a melhor fusão de resultados para esta combinação até agora, embora ainda seja inferior às outras modalidades de combinação. Entretanto, em termos de maior precisão, a fusão no nível de decisão não foi capaz de obter melhor performance que a fusão no nível de similaridade ou características.

5.1.6 Discussão

Os métodos de fusão avaliados em nossos experimentos não demandam um alto custo computacional, o que nos permite combinar múltiplas modalidades para aprimorar os resultados finais sem consumir demais os recursos computacionais disponíveis. Além do mais, tais métodos são menos propensos a *overfitting* quando comparados com outras abordagens baseadas em treinamento, como *Support Vector Machine* (SMV) e redes neurais. Ao ponderar a soma de características, foi possível diminuir os impactos negativos das características de profundidade e combiná-la com textura e NIR para criar uma característica mais discriminativa. Os resultados das seções anteriores foram compilados na Tabela, 5.1 onde é possível ver que a melhor EER e taxas de reconhecimento em 1% e 0,1% de FAR foram alcançadas ponderando a soma de características, mostrando que nossa suposição inicial estava certa e que somar características é, de fato, uma opção de fusão viável. Ela atinge uma taxa de reconhecimento de 97,9% em 0,1 de FAR, o que é próximo aos padrões comerciais para os sistemas de reconhecimento de impressão digital (99% em 0,1% de FAR) e aproximadamente 2% acima da melhor performance monomodal na Tabela 5.1. A soma ponderada de similaridade e o operador OR, os melhores métodos de fusão nos níveis de similaridade e decisão, não foram capazes superar a soma ponderada de característica, corroborando a afirmação de que a fusão no nível de característica retém mais informações discriminativas e é mais efetiva do que outros níveis de fusão.

(a) Regra da soma (b) Regra do produto



(c) Regra do mínimo (d) Regra do máximo

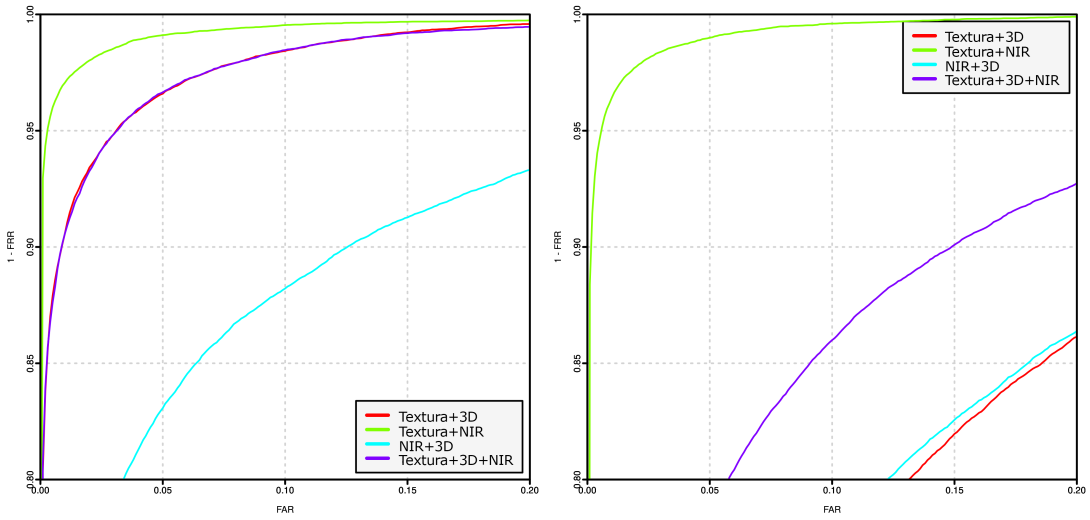


Figura 5.6 Resultado de fusões a nível de similaridade para (a) regra da soma, (b) regra do produto, (c) regra do mínimo e (d) regra do máximo.

Tabela 5.1 FARs and EERs dos experimentos de fusão. Melhores resultados estão em negrito.

Método	Modalidade	FAR 1%	FAR 0.1%	FAR 0.01%	EER %
Monomodal	2D	0.9878773	0.9602555	0.9076967	1.1187
	NIR	0.8911399	0.7692621	0.6144740	5.081
	3D	0.1201285	0.0392564	0.0000814	35.3267
Soma de Características	2D+3D	0.9828330	0.9379221	0.8607924	1.2611
	2D+NIR	0.9933691	0.9764055	0.9375153	0.9234
	3D+NIR	0.7835001	0.5700919	0.3623790	6.4315
	2D+3D+NIR	0.9931251	0.9705476	0.9241315	2.2944
Soma Ponderada de Características	2D+3D	0.9878773	0.9602555	0.9076967	1.1187
	2D+NIR	0.9947523	**0.9786836**	**0.9407697**	0.70377
	3D+NIR	0.8911399	0.7692621	0.6144740	5.081
	2D+3D+NIR	**0.9954031**	0.9779107	0.9366203	**0.6712**
Concatenação de Características	2D+3D	0.9857213	0.9502481	0.8849972	1.3913
	2D+NIR	0.9915792	0.9651778	0.9193719	0.7811
	3D+NIR	0.8311773	0.6622325	0.4867789	7.1597
	2D+3D+NIR	0.9897893	0.9600521	0.9100561	0.8177
Regra da Soma de Similaridade	2D+3D	0.9042389	0.7830933	0.6436824	4.0762
	2D+NIR	0.9826296	0.9498820	0.8916687	1.4767
	3D+NIR	0.7370434	0.5467415	0.3644130	8.8845
	2D+3D+NIR	0.9588317	0.8856887	0.7733708	2.3635
Regra da Soma Ponderada de Similaridade	2D+3D	0.9878773	0.9602555	0.9076967	1.1187
	2D+NIR	0.991783	0.963551	0.923114	0.895
	3D+NIR	0.8911399	0.7692621	0.6144740	5.081
	2D+3D+NIR	0.991783	0.963551	0.923114	0.895
Regra do Produto de Similaridade	2D+3D	0.9154259	0.8022130	0.6652022	3.649
	2D+NIR	0.9812058	0.9478073	0.8905297	1.5418
	3D+NIR	0.7346839	0.5456025	0.3671386	9.2669
	2D+3D+NIR	0.9590351	0.8808478	0.7702384	2.335
Regra da Similaridade Mínima	2D+3D	0.905825	0.803596	0.679888	4.1168
	2D+NIR	0.970629	0.929623	0.873200	2.0503
	3D+NIR	0.693922	0.497396	0.315800	11.1179
	2D+3D+NIR	0.905744	0.807013	0.699333	4.0477
Regra da Similaridade Máxima	2D+3D	0.450452	0.242657	0.000081	16.6423
	2D+NIR	0.964242	0.883817	0.747864	2.152
	3D+NIR	0.469002	0.242901	0.000081	16.4307
	2D+3D+NIR	0.590391	0.350256	0.000081	12.0291
Regra do *OR*	2D+3D	0.981450	0.947116	0.884428	1.5214
	2D+NIR	0.990928	0.961557	0.902937	0.9845
	3D+NIR	0.862542	0.729111	0.554918	6.1386
	2D+3D+NIR	0.988081	0.953950	0.889472	1.1024
Regra do *AND*	2D+3D	0.315393	0.162843	0.090920	25.6244
	2D+NIR	0.940932	0.873037	0.775608	3.4375
	3D+NIR	0.308803	0.158409	0.088520	33.8134
	2D+3D+NIR	0.391709	0.225165	0.135221	22.427
Maioria Absoluta	2D+3D	0.315393	0.162843	0.090920	25.6244
	2D+NIR	0.940932	0.873037	0.775608	3.4375
	3D+NIR	0.308803	0.158409	0.088520	33.8134
	2D+3D+NIR	0.935278	0.865715	0.770523	3.6327

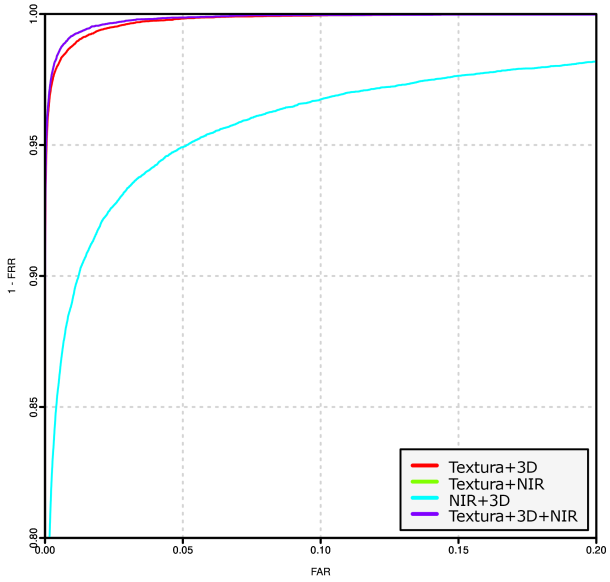


Figura 5.7 Resultado da fusão a nível de similaridade para regra da soma ponderada.

5.2 AUTENTICAÇÃO CONTÍNUA MULTIMODAL

5.2.1 Reconhecimento Contínuo Multimodal

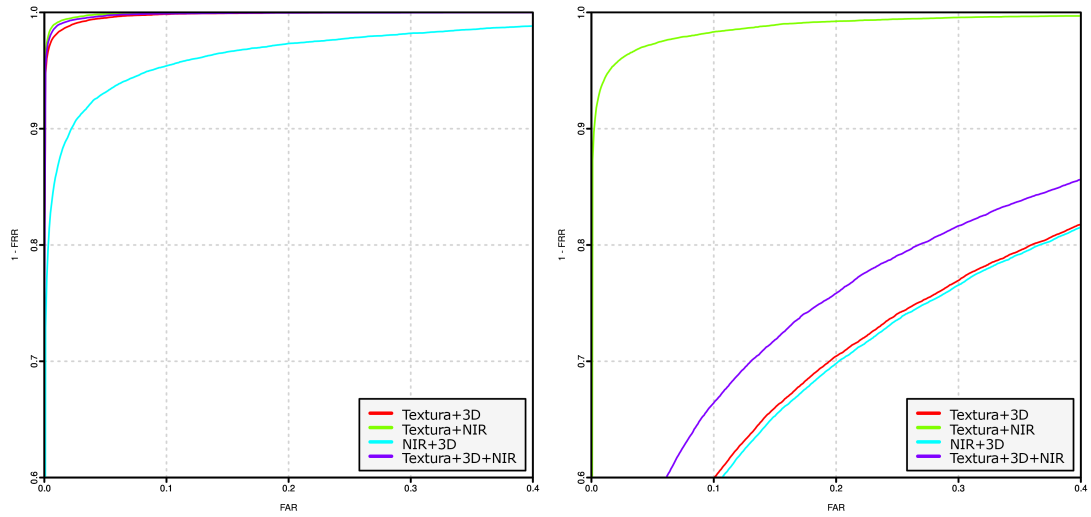
Para os experimentos de autenticação contínua, 8 indivíduos foram gravados por um *Kinect v2* enquanto utilizavam um computador para fins diversos. Eles podiam agir livremente e a única restrição que lhes foi dada foi para permanecer em frente ao sensor durante a gravação das 3 modalidades simultaneamente. As gravações foram realizadas a cerca de 13 quadros por segundo, e como todo o processo do reconhecimento, desde a aquisição até a decisão, leva cerca de 0.6 segundos, nem todos os quadros seriam processados se o tempo real em que foram adquiridos fosse considerado. Assim, para aumentar o número de amostras para o experimento, todos os quadros foram reconhecidos de forma contínua e sequencialmente. Além disso, os quadros que não puderam ser normalizados foram descartados. Então, a quantidade de quadros variou de 20771 a 34502 selecionados por pessoa para cada modalidade, que representavam cerca de 27 a 44 minutos por gravação, foram simulados como aquisições contínuas de 208 a 345 minutos aproximadamente.

Cada quadro do processo de autenticação contínua foi processado pelo detector proposto por Zhang et al. (2016) para imagens de cor. Os pontos da face retornados pelo detector permitiram realizar a normalização facial, rotacionando a imagem e alinhando-a às informações de 3D e NIR. A face 3D foi em seguida frontalizada, e os buracos gerados por oclusão ou ruído do sensor foram preenchidos por meio da propagação de pixels válidos vizinhos. Uma vez que as informações estavam alinhadas, as imagens de 2D e NIR foram modificadas de forma a simular uma frontalização, como exemplifica a Figura 5.9.

As imagens normalizadas de cada modalidade foram então descritas pela CNN su-

(a) Operator OR

(b) Operator AND



(c) Maioria absoluta

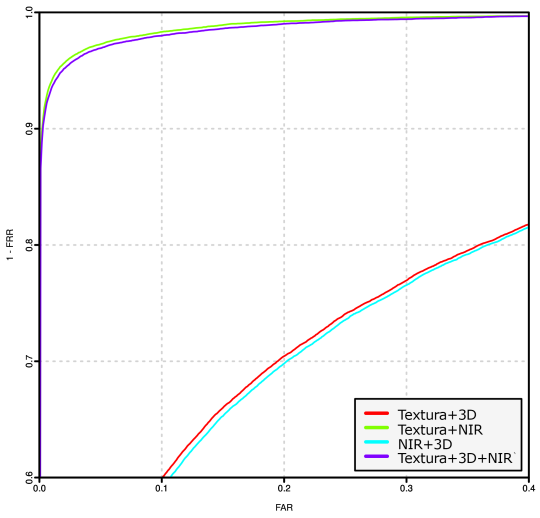


Figura 5.8 Resultados de fusões a nível de decisão para (a) o operador OR, (b) o operador AND e (c) Maioria Absoluta.

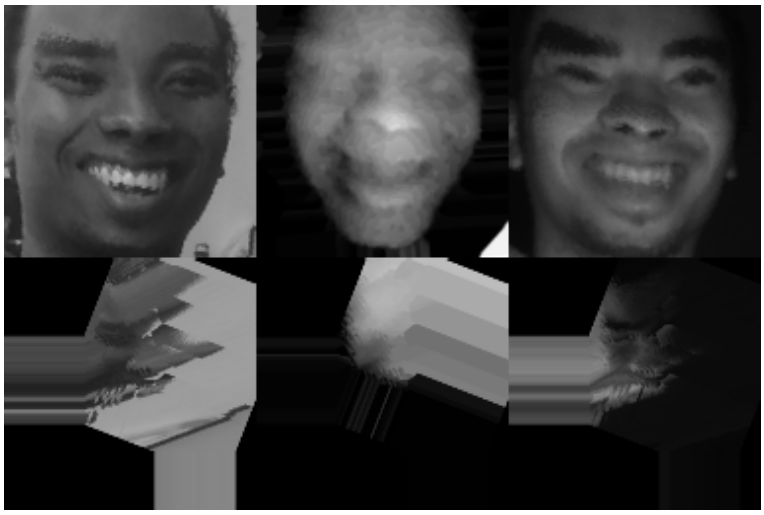


Figura 5.9 Na primeira linha, exemplos de imagens normalizadas corretamente para as 3 modalidades, na segunda, exemplos de imagens não normalizadas de forma adequada.

gerida por Wu, He e Sun (2015), e o descritor gerado para a primeira imagem facial foi utilizado como modelo de comparação com os descritores gerados posteriormente. A primeira imagem facial foi selecionada como em um processo de autenticação inicial, ou seja, o usuário agiu da forma que achou melhor para ser reconhecido inicialmente. As similaridades resultantes das comparações feitas para as autenticações contínuas monomodais de 3D, 2D e NIR e a multimodal baseada na soma dos descritores, como descrito na Seção 4.1, foram avaliadas considerando observações ao longo do tempo.

5.2.2 Resultados Experimentais

Seguindo os experimentos realizados por Segundo (2013), para analisar os resultados desse experimento, levamos em consideração dois aspectos: 1) o desempenho de cada modalidade para manter a face de um usuário continuamente reconhecida, sendo esta face a primeira capturada e utilizada como modelo para comparação ao longo da seção, simulando assim uma autenticação inicial e a verificação contínua; 2) a capacidade de cada modalidade para lidar com invasões, por meio da concatenação de cada sequência genuína de quadros às sequências de invasores (*i.e.* sequência dos outros 7 indivíduos gravados), simulando assim 56 tentativas de ataque no total, para as 4 modalidades.

Para os experimentos monomodais, podemos ver pelas Figuras 5.10 e 5.11 que cor e NIR monomodal mantiveram o P_{safe} acima de 0.95 na maior parte do tempo e se comportaram de forma similar para o reconhecimento do usuário genuíno. As pequenas diferenças entre estes reconhecimentos demonstram a vantagem no uso de cor na maioria dos casos, como esperado ao considerarmos os resultados apresentados na Seção 4.1. Por meio das figuras também é possível perceber decréscimos mais significativos nos valores do P_{safe} em momentos parecidos para ambas as modalidades. Esse comportamento representa um aumento da FRR e também pode ser observado nas demais modalidades, incluindo a fusão, pois está relacionada aos momentos em que os usuários aproximaram-se

Tabela 5.2 TRR médio para os experimentos de cada modalidade variando limiares de P_{safe} entre 0.9 e 1. Os valores mostram que mesmo com as distorções nas imagens da face causadas pela proximidade do usuário diante do sensor, ao definirmos um limiar de 0.9 conseguiríamos manter o usuário genuíno autenticado quase sempre.

Limiar / Modalidade	2D	NIR	3D	Multimodal
0.9	99,196576286	98,685672286	78,831594286	99,031972429
0.91	99,101121714	98,504305786	76,32248	98,907992286
0.92	98,993189714	98,332188714	73,058852857	98,778406714
0.93	98,888744571	98,146237286	69,340638571	98,622779143
0.94	98,751244286	97,877863286	65,306184286	98,485699429
0.95	98,505058143	97,666845143	60,338005714	98,265601571
0.96	98,173268571	97,288161143	54,555664286	97,926498571
0.97	97,663480601	96,550718571	46,63464	97,303940429
0.98	96,488319486	95,246621714	36,070684286	96,379384857
0.99	92,540538571	92,354245571	16,687171429	92,358793714
1	70,360297143	57,776527143		67,614371429

demais do sensor durante o experimento, causando distorção na imagem de profundidade e, consequentemente, nas demais modalidades devido à normalização. A FRR para cada modalidade é apresentada na Tabela 5.2, onde os resultados são demonstrados para P_{safe} acima de 0.9, pois, sem contar com 3D, as quedas abaixo desse limiar estão somente relacionadas ao problema da proximidade.

Para as simulações de ataques, dentre as 56 tentativas de invasões, 4 tentativas para cor e 7 para NIR ultrapassam o limiar de 0.8 de P_{safe}, ocorrendo para 4 usuários dos experimentos de cor e 5 de NIR dentre os 8 para cada modalidade. Estes valores demonstram a superioridade da modalidade de cor em relação a NIR para a quantidade de tentativas de ataques, porém, as invasões na modalidade de cor que superam 0.8 de P_{safe} possuem, no geral, valores superiores aos de NIR, bem como permanecem por mais tempo com altos valores.

Os resultados para a modalidade de 3D monomodal, como mostrado na Figura 5.12, também foram previsíveis se considerarmos a Seção 4.1, com valores de P_{safe} mais baixos que cor e NIR para os genuínos, porém, estes resultados não foram tão inferiores às outras modalidades quanto no experimento passado, visto que o P_{safe} foi mantido acima de 0.8 de P_{safe} durante de 0.6% a 6.5% do tempo em 6 experimentos, ainda considerando o problema da proximidade inadequada dos usuários, explicado anteriormente. Esse fato enfatiza nossa premissa de que a baixa qualidade nas imagens 3D do experimento de reconhecimento multimodal foi o real motivo para os resultados menos satisfatórios, e que o descritor proposto por (WU; HE; SUN, 2015), apesar de ter sido treinado para faces 2D, pode ser utilizado para faces 3D. Para compararmos com os dados analisados no parágrafo passado, dentre as 56 tentativas de ataque para 3D, 11 ultrapassaram 0.8 de P_{safe}, sendo só para 5 usuários, assim como NIR, mas não exatamente para os mesmos ataques.

Na Figura 5.13 podemos ver o resultado do uso da fusão multimodal para a auten-

ticação contínua, que demonstrou resultados similares aos de cor e NIR para os usuários genuínos. É possível perceber em algumas gravações que a modalidade de cor consegue manter o P_{safe} melhor que a fusão multimodal, o que deve estar relacionado a quedas muito mais significativas nas modalidades de NIR e 3D. Ao definirmos um limiar de 95 de P_{safe}, a fusão multimodal mantem o usuário autenticado em 98,3% dos 255128 quadros de acesso genuíno que foram analisados, considerando o problema da proximidade inadequada descrito anteriormente. Com este mesmo limiar, a modalidade de cor mantém o usuário genuíno autenticado em 98,5% e NIR em 97,8% dos quadros.

O ganho significativo de acurácia para a fusão multimodal em relação às demais modalidades mais uma vez não pode ser percebido ao analisarmos o EER médio dos resultados, que foi de 0,1% para fusão e 0,3%, 0,3% e 6% para cor, NIR e 3D respectivamente. Contudo, a eficácia do método de fusão é destacada durante as simulações de tentativas de invasões ao sistema, onde podemos observar o maior poder discriminativo da soma ponderada de características para detectar os invasores mais rápido que as outras modalidades. Há também uma redução significativa do P_{safe} para alguns ataques que mantinham valores mais próximos do genuíno para os métodos monomodais. Além disso, para a fusão multimodal, apenas um dos ataques demonstrou uma chance expressiva de elevar o nível do P_{safe} após a queda, sendo ainda assim inferior a 0.8 de P_{safe}, ao contrário das modalidades monomodais. Este único ataque também é o único a cruzar o limiar de 0.6 de P_{safe}, enquanto 10 ataques para cor e 11 para NIR conseguem superar essa marca e quase todos permanecem por mais tempo acima dela.

Como apresentado na Tabela 5.3, em comparação com os resultados de cor e NIR, que mantiveram suas taxas de P_{safe} para genuíno acima de 0.9 a maior parte do tempo, a fusão multimodal teve uma queda mais rápida do P_{safe} no início da invasão. Esses resultados demonstram mais uma vez que as características geradas pela fusão multimodal possuem maior poder discriminativo que as das demais modalidades. Contudo, para os experimentos realizados, se o limiar fosse definido em 0.99 de P_{safe} apenas um caso levaria mais de 5 segundos para o intruso ser detectado (*i.e.* 10 segundos), mas isso aumentaria a FRR. Com um limiar de 0.9 de P_{safe}, a FRR reduziria, mas no pior caso do experimento um intruso levaria 46.8 segundos para ser detectado, o que pode ser um tempo significativo para um ataque a um sistema de alto risco. Assim, considerando os melhores resultados da fusão, esta é a opção mais viável para um sistema de autenticação contínua que as demais modalidades, mas uma relação adequada entre o limiar do P_{safe} e o tempo para detectar um intruso precisa ser definido.

5.2.3 Discussão

Se compararmos os experimentos que realizamos com os de demais trabalhos de autenticação contínua na Tabela 3.1, a quantidade de amostras contínuas e a frequência com que estas são reconhecidas em nossos testes são mais próximos de cenários reais. Além disso, os resultados que alcançamos, mesmo para as modalidades individualmente, foram superiores ao dos demais trabalhos. Com isso podemos enfatizar o fato de que o reconhecimento facial para autenticação contínua é mais viável que o reconhecimento das demais biometrias já estudadas até agora, considerando as devidas limitações (*e.g.* é necessário

que haja um sensor capturando a face continuamente), sem ser intrusivo e exigindo pouco esforço do usuário.

Podemos também fazer uma comparação com o trabalho de Segundo (2013), que possui experimentos similares ao nosso e é o único outro trabalho a utilizar um sensor de baixo custo para autenticação contínua, mas utilizando apenas imagens 3D. Nosso trabalho alcançou uma melhor taxa de EER, apesar de seus experimentos possuírem cerca de 70000 quadros para cada um dos 4 indivíduos gravados. Por outro lado, em termos de P_{safe} os experimentos de Segundo (2013) conseguem detectar intrusos em cerca de 1 segundo para quase todos os casos com um limiar de 0.8 de P_{safe}, enquanto em nossos experimentos os intrusos podem ser detectados em menos de 30 segundos, mas com um limiar de 0.95 de P_{safe}. Ainda assim, vale salientar que na simulação da invasão o rosto do intruso é substituído em menos de um segundo, o que dificilmente aconteceria em um cenário real. Além disso, como nossos experimentos não focaram na investigação de métodos de análise temporal para autenticação contínua, alterar a equação do P_{safe} ou as constantes utilizadas nesta, de forma a se adaptarem ao nosso problema, pode fazer com que haja um decaimento mais rápido para intrusões com limiares menores de P_{safe}, bem como para a ausência de faces. Outra possibilidade seria aumentar o limiar do P_{safe}, mas, como demonstrado previamente, isso aumentaria a FRR.

Tabela 5.3 Quantidade vezes em que ataques não são detectados utilizando determinados limiares para determinados limites de tempo, para cada modalidade. Por meio desses valores podemos observar a maior eficácia da fusão multimodal para todos os casos. Apesar de limiares mais altos representarem detecções mais rápidas de ataques, ao relacionarmos essa tabela com a Tabela 5.2 é possível ver que tais limiares também representam menores TRR. Logo, para um sistema que utilize autenticação contínua é necessário definir qual critério é mais relevante.

Modalidade	Limiar/Segundos	5	10	15	20	25	30	35	40	45	50
2D	0.9	46	27	20	15	13	10	5	5	4	3
	0.91	41	24	20	15	10	7	5	5	4	3
	0.92	37	22	18	15	10	6	5	5	4	3
	0.93	34	21	17	12	9	5	5	4	4	3
	0.94	28	20	15	10	6	5	5	4	3	3
	0.95	25	18	12	9	5	5	4	4	3	3
	0.96	22	15	10	5	5	5	4	4	2	2
	0.97	20	12	6	5	5	4	2	1	1	1
	0.98	15	6	5	5	3	2	1	1	1	1
	0.99	6	5	3	2	1	1	1	1	1	1
NIR	0.9	49	29	22	17	13	9	9	6	3	2
	0.91	46	26	20	15	11	9	8	5	3	2
	0.92	42	24	17	13	9	9	7	5	2	1
	0.93	37	23	17	12	9	8	7	3	2	1
	0.94	30	21	15	9	9	7	5	3	2	1
	0.95	28	17	13	9	8	7	3	2	1	1
	0.96	22	15	9	9	7	5	3	1	1	0
	0.97	21	13	9	7	5	3	1	1	1	0
	0.98	17	9	7	3	3	1	1	0	0	0
	0.99	9	5	2	1	1	0	0	0	0	0
Multimodal	0.9	37	18	12	7	4	4	2	1	1	0
	0.91	33	16	11	4	4	3	1	1	1	0
	0.92	31	15	11	4	4	2	1	1	0	0
	0.93	27	13	5	4	4	2	1	1	0	0
	0.94	24	11	4	4	2	1	1	0	0	0
	0.95	19	8	4	4	2	1	1	0	0	0
	0.96	15	5	4	2	1	1	0	0	0	0
	0.97	10	4	3	1	1	0	0	0	0	0
	0.98	6	4	1	1	0	0	0	0	0	0
	0.99	1	1	0	0	0	0	0	0	0	0

Figura 5.10 Resultados da autenticação contínua monomodal para imagens de faces em textura para os genuínos (em azul) e para cada simulação de fraude.

Figura 5.11 Resultados da autenticação contínua monomodal para imagens de faces em NIR para os genuínos (em azul) e para cada simulação de fraude.

Figura 5.12 Resultados da autenticação contínua monomodal para imagens de faces em 3D para os genuínos (em azul) e para cada simulação de fraude.

Figura 5.13 Resultados da autenticação contínua multimodal para a fusão de imagens de faces em textura, NIR e 3D, para os genuínos (em azul) e para cada simulação de fraude.

Capítulo 6

CONCLUSÃO

Por meio de informações de 3D, 2D e infravermelho é possível reconhecer uma face utilizando diferentes propriedades, como textura, formato e emissões termais. Com as três modalidades combinadas, a precisão do reconhecimento pode ser aprimorada, e as chances da autenticação contínua ser burlada são reduzidas.

Para encontrar a melhor forma de fundir as modalidades, avaliamos o desempenho de vários métodos de fusão bem conhecidos em diferentes estágios do nosso sistema de reconhecimento. As faces em diferentes modalidades foram descritas pelo mesmo descritor CNN no estado-da-arte, que foi originalmente treinado apenas para imagens em cores, mostrando que os descritores profundamente aprendidos são transferíveis para outras modalidades.

6.1 RESULTADOS ALCANÇADOS

Para os experimentos de reconhecimento facial multimodal, criamos uma base de dados de faces multimodais com 1.845 imagens de 96 pessoas, coletadas ao longo de quatro meses em diferentes ambientes. Assim, 1.701.090 pares de imagens foram combinados um contra o outro para cada método avaliado.

Apesar de que o método de reconhecimento 2D no estado-da-arte que utilizamos atingiu uma impressionante marca de 1.12 % EER em nosso banco de dados, usando apenas imagens de textura, os resultados para a fusão multimodal mostram que é possível aumentar a precisão ao combinar diferentes propriedades faciais. Mesmo não sendo usual, a soma ponderada para a fusão no nível de características atingiu 0,67 % EER e foi a melhor opção para integrar as três modalidades. Além da precisão, a dimensionalidade original também é mantida, o que é vantajoso em termos de armazenamento e velocidade das comparações.

Embora as modalidades de profundidade e NIR não sejam tão precisas como cor, como esperado, elas contribuíram para um ganho de desempenho e contribuirão ainda mais contra tentativas de falsificação em um sistema real, pois é consideravelmente mais difícil de falsificar múltiplas modalidades simultaneamente. Assim, os resultados obtidos

mostram que o método de reconhecimento apresentado é uma opção viável para sistemas de segurança baseados no reconhecimento de rosto, uma vez que utiliza dispositivos de baixo custo, obtém maior precisão em comparação com sistemas monomodais no estado-da-arte, mantém a custo computacional original e é mais robusto a fraudes.

Para os experimentos de autenticação contínua, 255128 quadros para cada modalidade foram adquiridos de gravações de 8 indivíduos enquanto agiam livremente em frente a um sensor de baixo custo. Dessas gravações, os quadros onde as faces foram detectadas e normalizadas simularam uma aquisição contínua sem considerar o tempo em que foram adquiridos. Como as gravações estavam sendo feitas a uma taxa de captura mais rápida do que o sistema realizava todo o processo do reconhecimento, os quadros foram ordenados sequencialmente com base em suas capturas. Para aplicações no mundo real, restrições relacionadas às limitações dos sensores precisam ser devidamente definidas, pois, como demonstrado, o alto poder discriminativo da CNN utilizada, ao fazer comparações com faces deformadas, pode gerar valores de similaridades discrepantes o suficiente para ter um impacto negativo no sistema.

O desempenho do reconhecimento contínuo foi satisfatório para as modalidades de cor, NIR e para a fusão multimodal de forma similar. A principal diferença entre os métodos foi demonstrada ao concatenarmos as gravações para simular ataques fraudulentos, onde a fusão demonstrou uma maior capacidade para rejeitar os invasores. Contudo, apesar da EER de 0,1% alcançada, que é muito mais significativa que os resultados de demais trabalhos de autenticação contínua utilizando faces, uma relação entre o limiar de P_{safe} e o tempo para a detecção de um intruso precisa ser devidamente estabelecido diante do nível de segurança exigido pelo sistema. Além disso, pela viabilidade do método de análise temporal utilizado já sido comprovada em diferentes trabalhos de autenticação contínua, testes mais intensivos podem ser feitos para encontrar parâmetros mais adequados para o decaimento mais rápido da segurança após a mudança de usuário.

Contudo, a conclusão do projeto resultou no primeiro trabalho que combina as três propriedades faciais e que as aplica para a autenticação contínua utilizando um sensor de baixo custo, sendo uma opção e de baixo custo que auxiliará em problemas de autenticação de usuários, como o acesso indevido a recursos que requerem um alto nível de segurança. O modelo de autenticação contínua desenvolvido, após os devidos ajustes para o decaimento do P_{safe}, deve ser adaptado para se tornar ou ser embutido em um produto real.

6.2 TRABALHOS FUTUROS

A integração entre informações 2D, NIR e 3D para o reconhecimento facial previne ataques fraudulentos utilizando fotos, por meio da captura de profundidade, e máscaras, por meio da captura de cor. Contudo, a tecnologia de impressão 3D tem se tornado mais acessível e a partir dela é possível criar máscaras coloridas. Alguns trabalhos na literatura propõem métodos *anti-spoofing* utilizando a intensidade da reflexão sob a textura, mas não possuem experimentos suficientes para garantir que não há um material que burle as propriedades de captura de um determinado sensor. Então, para prevenir qualquer tipo de fraude em sistemas de autenticação contínua baseado em faces que utilizem sensores

RGB-D, a combinação de fotopletismografia remota, que é a análise da variação sanguínea por meio da face, utilizando NIR pode ser realizado. Dependendo do quão distintiva essa característica puder ser, visto que PPG já é utilizado para autenticação contínua, este método pode ser fundido às demais modalidades e, possivelmente como resultado, aumentar a acurácia e melhorar a detecção de fraudes.

REFERÊNCIAS BIBLIOGRÁFICAS

ABATE, A. F. et al. 2d and 3d face recognition: A survey. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 28, n. 14, p. 1885–1906, out. 2007. ISSN 0167-8655. Disponível em: ⟨http://dx.doi.org/10.1016/j.patrec.2006.12.018⟩.

AHONEN, T.; HADID, A.; PIETIKAINEN, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 28, n. 12, p. 2037–2041, dez. 2006. ISSN 0162-8828. Disponível em: ⟨http://dx.doi.org/10.1109/TPAMI.2006.244⟩.

ALTINOK, A.; TURK, M. Temporal integration for continuous multimodal biometrics. In: *Multimodal User Authentication*. [S.l.: s.n.], 2003. p. 11–12.

AMOS, B.; LUDWICZUK, B.; SATYANARAYANAN, M. *OpenFace: A general-purpose face recognition library with mobile applications*. [S.l.], 2016.

BELHUMEUR, P. N.; HESPANHA, J. a. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 19, n. 7, p. 711–720, jul. 1997. ISSN 0162-8828. Disponível em: ⟨http://dx.doi.org/10.1109/34.598228⟩.

BOLLE, R. et al. *Guide to Biometrics*. [S.l.]: SpringerVerlag, 2003. ISBN 0387400893.

BOURS, P.; MONDAL, S. Performance evaluation of continuous authentication systems. *IET Biometrics*, v. 4, n. 4, p. 220–226, 2015. ISSN 2047-4938.

BOWYER, K. W.; CHANG, K.; FLYNN, P. *A Survey of Approaches and Challenges in 3D and Multi-modal 3D + 2D Face Recognition*. New York, NY, USA: Elsevier Science Inc., 2006. 1–15 p. Disponível em: ⟨http://dx.doi.org/10.1016/j.cviu.2005.05.005⟩.

BOWYER, K. W. et al. Face recognition using 2-d, 3-d, and infrared: Is multimodal better than multisample? *Proceedings of the IEEE*, v. 94, n. 11, p. 2000–2012, Nov 2006. ISSN 0018-9219.

BROCARDO, M. L. *Continuous Authentication using Stylometry*. Tese (Doutorado) — Department of Electrical and Computer Engineering. University of Victoria - UVIC, Victoria, British Columbia, Canada, 2015.

CAO, K.; JAIN, A. K. *Hacking Mobile Phones Using 2D Printed Fingerprints*. East Lansing, Michigan, 2016. 3 p.

CHANG, K. I.; BOWYER, K. W.; FLYNN, P. J. Face recognition using 2d and 3d facial data. In: *ACM Workshop on Multimodal User Authentication*. [S.l.: s.n.], 2003. p. 25–32.

CHANG, K. I.; BOWYER, K. W.; FLYNN, P. J. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 28, n. 10, p. 1695–1700, Oct 2006. ISSN 0162-8828.

DAHIA, G.; SANTOS, M.; SEGUNDO, M. P. A study of cnn outside of training conditions. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. [S.l.: s.n.], 2005. v. 1, p. 886–893 vol. 1. ISSN 1063-6919.

DEZA, M. M.; DEZA, E. *Encyclopedia of Distances*. [S.l.]: Springer Berlin Heidelberg, 2009.

DING, C.; TAO, D. Robust face recognition via multimodal deep face representation. *CoRR*, abs/1509.00244, 2015. Disponível em: ⟨http://arxiv.org/abs/1509.00244⟩.

DING, C.; TAO, D. Robust face recognition via multimodal deep face representation. *CoRR*, abs/1509.00244, 2015. Disponível em: ⟨http://arxiv.org/abs/1509.00244⟩.

EISENBERG, B.; SULLIVAN, R. Why is the sum of independent normal random variables normal? *Mathematics Magazine*, Mathematical Association of America, v. 81, n. 5, p. 362–366, 2008. ISSN 0025570X, 19300980. Disponível em: ⟨http://www.jstor.org/stable/27643141⟩.

ERDOGMUS, N.; MARCEL, S. Spoofing face recognition with 3d masks. *IEEE Transactions on Information Forensics and Security*, v. 9, n. 7, p. 1084–1097, July 2014. ISSN 1556-6013.

FEHER, C. et al. User identity verification via mouse dynamics. *Inf. Sci.*, Elsevier Science Inc., New York, NY, USA, v. 201, p. 19–36, out. 2012. ISSN 0020-0255. Disponível em: ⟨http://dx.doi.org/10.1016/j.ins.2012.02.066⟩.

FENG, H.; FAWAZ, K.; SHIN, K. G. Continuous authentication for voice assistants. *CoRR*, abs/1701.04507, 2017. Disponível em: ⟨http://arxiv.org/abs/1701.04507⟩.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals Eugen.*, v. 7, p. 179–188, 1936.

GUO, Y. et al. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. *CoRR*, abs/1607.08221, 2016. Disponível em: ⟨http://arxiv.org/abs/1607.08221⟩.

HASSABALLAH, M.; ALY, S. Face recognition: challenges, achievements and future directions. *IET Computer Vision*, v. 9, n. 4, p. 614–626, 2015. ISSN 1751-9632.

HONG, L.; JAIN, A.; PANKANTI, S. Can Multibiometrics Improve Performance. In: *Proc. AUTOID*. [S.l.: s.n.], 1999.

JAIN, A.; NANDAKUMAR, K.; ROSS, A. Score normalization in multimodal biometric systems. *Pattern Recogn.*, Elsevier Science Inc., New York, NY, USA, v. 38, n. 12, p. 2270–2285, dez. 2005. ISSN 0031-3203. Disponível em: ⟨http://dx.doi.org/10.1016/j.patcog.2005.01.012⟩.

JAIN, A. K.; FLYNN, P.; ROSS, A. A. *Handbook of Biometrics*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007. ISBN 038771040X.

JAIN, A. K. et al. Biometrics: a grand challenge. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. [S.l.: s.n.], 2004. v. 2, p. 935–942 Vol.2. ISSN 1051-4651.

JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, v. 14, p. 4–20, 2004.

JANAKIRAMAN, R. et al. Using continuous face verification to improve desktop security. In: *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*. [S.l.: s.n.], 2005. v. 1, p. 501–507.

KAKADIARIS, I. A. et al. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 29, n. 4, p. 640–649, April 2007. ISSN 0162-8828.

LEGGETT, J. et al. Dynamic identity verification via keystroke characteristics. *Int. J. Man-Mach. Stud.*, Academic Press Ltd., London, UK, UK, v. 35, n. 6, p. 859–870, nov. 1991. ISSN 0020-7373. Disponível em: ⟨http://dx.doi.org/10.1016/S0020-7373(05)80165-8⟩.

LI, S. et al. Illumination invariant face recognition using near-infrared images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 29, n. 4, p. 627–639, April 2007. ISSN 0162-8828.

LOUIS, W.; KOMEILI, M.; HATZINAKOS, D. Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ecg biometrics. *IEEE Transactions on Information Forensics and Security*, v. 11, n. 12, p. 2818–2832, Dec 2016. ISSN 1556-6013.

LU, X.; JAIN, A. K. *Multimodal facial feature extraction for automatic 3D face recognition*. [S.l.], 2005.

MAGALHAES, M. B. S.; SEGUNDO, M. P. Autenticação facial contínua usando imagens de infravermelho. *CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 28. (SIBGRAPI)*, Aug 2015.

MARTINOVIC, I. et al. Pulse-response: Exploring human body impedance for biometric recognition. In: . [s.n.], 2017. v. 20, n. 2, p. 6:1–6:31. Disponível em: ⟨http://dblp.uni-trier.de/db/journals/tissec/tissec20.html\#MartinovicRRT17⟩.

MCC, R. A history of security. In: GILL, M. (Ed.). *Handbook of Security*. London: Palgrave Macmillan: [s.n.], 2006.

MIAN, A.; BENNAMOUN, M.; OWENS, R. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 29, n. 11, p. 1927–1943, Nov 2007. ISSN 0162-8828.

MIN, R.; KOSE, N.; DUGELAY, J. L. Kinectfacedb: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, v. 44, n. 11, p. 1534–1548, Nov 2014. ISSN 2168-2216.

MONACO, J. et al. Developing a keystroke biometric system for continual authentication of computer users. In: *Intelligence and Security Informatics Conference (EISIC), 2012 European*. [S.l.: s.n.], 2012. p. 210–216.

NIINUMA, K.; PARK, U.; JAIN, A. K. Soft biometric traits for continuous user authentication. *Trans. Info. For. Sec.*, IEEE Press, Piscataway, NJ, USA, v. 5, n. 4, p. 771–780, dez. 2010. ISSN 1556-6013. Disponível em: ⟨http://dx.doi.org/10.1109/TIFS.2010.2075927⟩.

O'GORMAN, L. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, v. 91, n. 12, p. 2021–2040, Dec 2003. ISSN 0018-9219.

PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, v. 2, n. 6, p. 559–572, 1901.

REDDY, C. V. R. et al. Person identification system using feature level fusion of multibiometrics. In: *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. [S.l.: s.n.], 2016. p. 1–6.

SANDHU, R. S.; SAMARATI, P. Access control: principle and practice. *IEEE Communications Magazine*, v. 32, n. 9, p. 40–48, Sept 1994. ISSN 0163-6804.

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832, 2015. Disponível em: ⟨http://arxiv.org/abs/1503.03832⟩.

SEGUNDO, M. P. *Real-time 3D face recognition using low-cost acquisition devices*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DO PARANA, Curitiba, 2013.

SEGUNDO, M. P. et al. Continuous 3d face authentication using rgb-d cameras. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. [S.l.: s.n.], 2013. p. 64–69.

SHEN, C. et al. Touch-interaction behavior for continuous user authentication on smartphones. In: *2015 International Conference on Biometrics (ICB)*. [S.l.: s.n.], 2015. p. 157–162. ISSN 2376-4201.

SILVA, A. A. D.; SEGUNDO, M. P. Reconhecimento facial 2d para autenticação contínua. *CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 28. (SIBGRAPI)*, aug 2015.

SIM, T. et al. Continuous verification using multimodal biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 29, n. 4, p. 687–700, 2007.

SUI, Y. et al. Secure and privacy-preserving biometrics based active authentication. In: *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.: s.n.], 2012. p. 1291–1296. ISSN 1062-922X.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. *J. Cognitive Neuroscience*, MIT Press, Cambridge, MA, USA, v. 3, n. 1, p. 71–86, jan. 1991. ISSN 0898-929X. Disponível em: ⟨http://dx.doi.org/10.1162/jocn.1991.3.1.71⟩.

VIOLA, P.; JONES, M. J. Robust real-time face detection. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 57, n. 2, p. 137–154, maio 2004. ISSN 0920-5691. Disponível em: ⟨http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb⟩.

WU, X.; HE, R.; SUN, Z. A lightened CNN for deep face representation. *CoRR*, abs/1511.02683, 2015. Disponível em: ⟨http://arxiv.org/abs/1511.02683⟩.

XI, K.; TANG, Y.; HU, J. Correlation keystroke verification scheme for user access control in cloud computing environment. *Comput. J.*, Oxford University Press, Oxford, UK, v. 54, n. 10, p. 1632–1644, out. 2011. ISSN 0010-4620. Disponível em: ⟨http://dx.doi.org/10.1093/comjnl/bxr064⟩.

XU, W. et al. Keh-gait: Towards a mobile healthcare user authentication system by kinetic energy harvesting. In: *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*. [S.l.]: Proceedings of NDSS, 2017.

YI, D. et al. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014. Disponível em: ⟨http://arxiv.org/abs/1411.7923⟩.

ZHANG, K. et al. Joint face detection and alignment using multi-task cascaded convolutional networks. *CoRR*, abs/1604.02878, 2016. Disponível em: ⟨http://arxiv.org/abs/1604.02878⟩.

ZHAO, W. et al. Face recognition: A literature survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 35, n. 4, p. 399–458, dez. 2003. ISSN 0360-0300. Disponível em: ⟨http://doi.acm.org/10.1145/954339.954342⟩.