



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Bauru

Marcelo Vilela Vizoni

Reconhecimento da Região Ocular para a Identificação Biométrica de Pessoas Utilizando Aprendizado em Profundidade

Bauru, São Paulo, Brasil

30 de Agosto, 2019

Marcelo Vilela Vizoni

Reconhecimento da Região Ocular para a Identificação Biométrica de Pessoas Utilizando Aprendizado em Profundidade

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração - Computação Aplicada, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Bauru.
Financiadora: CAPES

Orientador: Prof. Associado Aparecido Nilceu Marana

Bauru, São Paulo, Brasil

30 de Agosto, 2019

Vizoni, Marcelo Vilela.

Reconhecimento da Região Ocular para a
Identificação Biométrica de Pessoas Utilizando
Aprendizado em Profundidade / Marcelo Vilela Vizoni,
2019.

75 f.: il.; 30cm

Orientador: Aparecido Nilceu Marana

Dissertação (Mestrado)-Universidade Estadual
Paulista "Júlio de Mesquita Filho". Faculdade de
Ciências, Bauru, 2019

1. Identificação de Pessoas. 2. Biometria. 3.
Reconhecimento da Região Ocular. 4. Aprendizado em
Profundidade. 5. Redes Neurais Convolucionais I.
Universidade Estadual Paulista "Júlio de Mesquita
Filho". Faculdade de Ciências. II. Dissertação de
Mestrado.

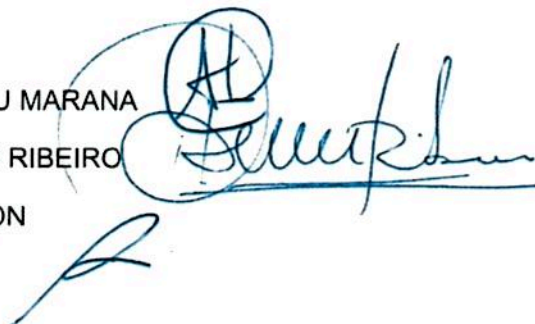
ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE MARCELO VILELA VIZONI, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DO INSTITUTO DE BIOCÊNCIAS, LETRAS E CIÊNCIAS EXATAS - CÂMPUS DE SÃO JOSÉ DO RIO PRETO.

Aos 30 dias do mês de agosto do ano de 2019, às 14:00 horas, no(a) FC/Bauru, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. APARECIDO NILCEU MARANA - Orientador(a) do(a) Departamento de Computação / UNESP/Câmpus de Bauru, Profa. Dra. PATRICIA BELLIN RIBEIRO do(a) Departamento de Computação / Faculdade de Tecnologia de Bauru, Profa. Dra. ROBERTA SPOLON do(a) Departamento de Computação / UNESP- Câmpus de Bauru, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de MARCELO VILELA VIZONI, intitulada **Reconhecimento da Região Ocular para a Identificação Biométrica de Pessoas Utilizando Aprendizado em Profundidade**. Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: Aprovado. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. APARECIDO NILCEU MARANA

Profa. Dra. PATRICIA BELLIN RIBEIRO

Profa. Dra. ROBERTA SPOLON



Marcelo Vilela Vizoni

Reconhecimento da Região Ocular para a Identificação Biométrica de Pessoas Utilizando Aprendizado em Profundidade

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, Área de Concentração - Computação Aplicada, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Bauru.
Financiadora: CAPES

Comissão Examinadora

Prof. Associado Aparecido Nilceu Marana

UNESP - Câmpus de Bauru

Orientador

Prof^ª. Associada Roberta Spolon

UNESP - Câmpus de Bauru

Prof^ª. Dr^a. Patricia Bellin Ribeiro

Fatec - Bauru

Bauru, São Paulo, Brasil

30 de Agosto, 2019

Este trabalho é dedicado aos meus pais, Braz Aparecido Vizoni e Ana Benedita Vilela Vizoni.

Agradecimentos

Em primeiro lugar, não posso deixar de agradecer aos meus pais, Braz e Ana, que sempre me ajudam e apoiam em todos desafios que me proponho a enfrentar.

Agradeço à minha namorada, Ana Julia, que está ao meu lado há tantos anos, e continua me fortalecendo para que eu atinja meus objetivos.

Agradeço ao meu orientador, professor Nilceu, por toda paciência e empenho que sempre me orientou buscando sempre o melhor de mim.

Agradeço aos meus amigos de mestrado, Juliana, Nathalia, Felix e Remo, que dividiram não só conhecimento comigo, mas também me permitiram dividir momentos de descontração e companheirismo.

Agradeço, também, à CAPES pelo apoio financeiro, à UNESP por me acolher desde o cursinho pré-vestibular, e à NVIDIA Corporation (GPU Grant Program) pela Titan xp concedida.

“Durante muito tempo eu atirava sem conseguir abrir direito o arco, até que um dia o mestre me ensinou um exercício de respiração, e tudo ficou fácil. Perguntei por que demorara tanto para me corrigir. Ele respondeu: Se desde o início eu tivesse lhe ensinado os exercícios respiratórios, você acharia que eram desnecessários.”

(A arte cavalheiresca do arqueiro zen, Eugen Herrigel, 1995)

Resumo

Na sociedade atual, a identificação precisa e rápida dos indivíduos é uma necessidade. Devido às crescentes preocupações sobre segurança, a Biometria tem sido proposta para este fim. A região ocular da face, que inclui o olho, as pálpebras, os cílios e as sobrancelhas, é uma das mais recentes modalidades biométricas sendo pesquisadas. Além da alta unicidade desta região da face, sua utilização representa um bom *trade-off* entre a utilização de toda a região da face e a utilização apenas da textura da íris dos olhos, pois possibilita uma gama maior de distâncias do indivíduo sendo identificado ao sensor. Este trabalho apresenta um novo método de autenticação de pessoas baseado em características oculares profundas, que são extraídas da região ocular da face usando uma CNN (*Convolutional Neural Network*). Em nosso método, em vez de usar diretamente os características profundas para a autenticação, usamos a diferença entre as características de referência e teste, gerando um vetor diferença. Então, nosso método adota uma estratégia de pares. Em seguida, um classificador SVM (*Support Vector Machine*) binário é treinado para determinar se um vetor diferença é genuíno ou impostor. O novo método proposto para autenticação de pessoas baseado em características oculares foi avaliado em diferentes bases de dados, contendo toda a face ou apenas a região ocular. Em nossos experimentos, a fusão de características oculares com características faciais obteve melhores resultados do que o uso de características de toda a face quando há variações de iluminação, expressão facial e oclusões parciais da face.

Palavras-chave: *Identificação de Pessoas, Biometria, Reconhecimento da Região Ocular, Aprendizado em Profundidade, Redes Neurais Convolucionais.*

Abstract

In modern society, accurate and quick identification of individuals is a necessity. Due to growing security concerns, Biometrics has been proposed for this purpose. The ocular region of the face, which includes the eye, eyelids, eyelashes and eyebrows, is one of the most recent biometric modalities being investigated. In addition to the high uniqueness of this region of the face, its use represents a good trade-off between the use of the entire face region and using only the texture of the iris of the eyes, since it allows a greater range of distances of the individual being identified to the sensor. This work presents a new method for identity authentication based on ocular deep features, which are extracted from the ocular region of the face by using a very deep CNN (Convolutional Neural Network). In our method, instead of using directly the deep features for the authentication, we use the difference between the probe and reference deep features, creating a difference vector. So, our method adopts a pairwise strategy. Then, a binary SVM (Support Vector Machine) classifier is trained to determine whether a given difference of deep features is genuine or impostor. The proposed new method for identity authentication based on ocular features was evaluated on different databases, containing the entire face or only the ocular region. In our experiments, the fusion of ocular features with facial features obtained better results than the use of features of the whole face when there are variations of illumination, facial expression, and partial occlusions of the face.

Keywords: *Identification of people, Biometrics, Recognition of the ocular region, Deep Learning, Convolutional Neural Networks*

Lista de ilustrações

Figura 1 – Exemplos de características biométricas. a) DNA; b) Orelha; c) Face; d) Termograma Facial; e) Termograma da Mão; f) Veias da Mão; g) Impressão Digital; h) Caminhar; i) Geometria da Mão; j) Íris; k) Impressão da Palma; l) Retina; m) Assinatura; n) Voz.	24
Figura 2 – Imagem frontal do olho humano.	25
Figura 3 – Exemplos de imagens oculares com problemas de captura. a) Reflexão especular; b) Obstrução pelas pálpebras e pelo uso de óculos; c) Posicionamento do olho; e d) Dilatação da pupila.	27
Figura 4 – Face com marcações de alguns pontos chaves para determinar o reconhecimento.	29
Figura 5 – Exemplos de regiões oculares de diferentes indivíduos	30
Figura 6 – Sistema biométrico bimodal mostrando os três níveis de fusão	33
Figura 7 – Processo para gerar código da íris	36
Figura 8 – Pré processamento da região periocular	36
Figura 9 – Quatro abordagens analisadas	37
Figura 10 – Exemplo de arquitetura de CNN	41
Figura 11 – Estrutura da rede CNN implementada no trabalho de Zhao e Kumar (2017a)	42
Figura 12 – Etapa de extração de característica da rede OcularNet.	43
Figura 13 – Exemplo de uso do método MTCNN para detecção da face	45
Figura 14 – Resultado obtido na aplicação do método MTCNN em um dos experimentos deste trabalho, na base de dados faciais AR Face (MARTINEZ; BENAVENTE, 1998).	46
Figura 15 – Diagrama do método proposto para a autenticação biométrica de pessoas por meio de características profundas.	48
Figura 16 – Diferentes imagens do conjunto de dados UBIPr (PADOLE; PROENCA, 2012) capturadas de um mesmo indivíduo.	50

Figura 17 – Curvas ROC obtidas por meio de características profundas extraídas de cinco diferentes redes neurais profundas. Em vermelho as curvas ROC obtidas pelo método proposto com o classificador SVM, em verde as curvas ROC obtidas pela aplicação da distância Cosseno, e em azul as curvas ROC obtidas pela aplicação da distância Euclidiana ao conjunto de imagens oculares do conjunto de dados UBIPr (PADOLE; PROENCA, 2012).	53
Figura 18 – Curvas ROC obtidas pelas cinco CNNs avaliadas neste experimento como meio de obtenção de características oculares profundas. Como se pode observar o VGG-Face obteve o melhor resultado.	55
Figura 19 – Diferentes imagens do conjunto de dados AR Face (MARTINEZ; BENAVENTE, 1998) capturadas de um mesmo indivíduo.	55
Figura 20 – Vetor diferença calculado entre duas amostras, A e B.	57
Figura 21 – Curvas ROC obtidas pelos experimentos realizados na base de dados AR Face (MARTINEZ; BENAVENTE, 1998).	59
Figura 22 – Diferentes imagens do conjunto de dados FRGC (PHILLIPS et al., 2005) capturadas de um mesmo indivíduo, à esquerda imagens capturadas em ambiente controlado, e a direita imagens capturas em ambientes não controlados.. . . .	60
Figura 23 – Curvas ROC obtidas pelos experimentos realizados na base de dados FRGC (PHILLIPS et al., 2005).	61
Figura 24 – Diferentes imagens do conjunto de dados Ubiiris.v2 (PROENCA et al., 2010) capturadas de um mesmo indivíduo em diferentes condições de iluminação, posicionamento e distância do sensor.	63
Figura 25 – Curva ROC obtida pelo experimento realizado na base de dados Ubiiris.v2 (PROENCA et al., 2010).	63
Figura 26 – Diferentes imagens do conjunto de dados FOCS (NIST, 2018) capturadas de um mesmo indivíduo em diferentes condições.	65
Figura 27 – Imagens obtidas após a aplicação do método MTCNN em algumas amostras do conjunto de dados FOCS (NIST, 2018) capturadas de um mesmo indivíduo em diferentes condições.	65

Figura 28 – Curvas ROC obtidas pelo experimento realizado na base de dados FOCS (NIST, 2018).	67
--	----

Lista de tabelas

Tabela 1 – Comparação entre as principais características biométricas. A, M e B representam, respectivamente, os níveis alto, médio ou baixo de atendimento aos requisitos desejáveis das características biométricas. .	25
Tabela 2 – Resultados experimentais com a base de dados UBIPr (PADOLE; PRO-ENCA, 2012).	54
Tabela 3 – Resultados experimentais com a base de dados AR Face (MARTINEZ; BENAVENTE, 1998).	58
Tabela 4 – Resultados experimentais com a base de dados FRGC (PHILLIPS et al., 2005).	62
Tabela 5 – Resultados experimentais com a base de dados FOCS (NIST, 2018). . .	66

Lista de abreviaturas e siglas

2-D	Duas dimensões
NIR	<i>Near-infrared</i>
LBP	<i>Local Binary Patterns</i>
HOG	<i>Histogram of Oriented Gradients</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
BSIF	<i>Binarized Statistical Image Features</i>
SVM	<i>Support Vector Machine</i>
CNN	<i>Convolutional Neural Network</i>
CRBM	<i>Convolutional Restricted Boltzman Machines</i>
MTCNN	<i>Multitask Cascaded Convolutional Networks</i>
ROC	<i>Receiver Operating Characteristics</i>
PM	<i>Prababilistic Matching</i>
FAR	<i>False Accept Rate</i>
FRR	<i>False Reject Rate</i>
EER	<i>Equal Error Rate</i>

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	18
1.2	Objetivos	19
1.3	Organização da Dissertação	20
2	BIOMETRIA	21
2.1	Introdução	21
2.2	Sistemas biométricos	22
2.3	Íris	24
2.4	Face	28
2.5	Região ocular	30
2.6	Fusão de características biométricas	31
3	MÉTODOS PARA RECONHECIMENTO DA REGIÃO OCULAR	35
3.1	Métodos tradicionais	35
3.2	Métodos de aprendizado de máquina em profundidade	37
3.2.1	Redes Neurais Convolucionais (CNN)	39
4	MÉTODO PROPOSTO	44
4.1	Segmentação facial e ocular	44
4.2	Módulo de Extração de Características	45
4.3	Módulo de Reconhecimento da Região Ocular	46
5	EXPERIMENTOS	49
5.1	Experimento 1: UBIPr	50
5.1.1	Resultados e Discussões	52
5.2	Experimento 2: AR Face	54
5.2.1	Resultados e Discussões	57
5.3	Experimento 3: FRGC	59
5.3.1	Resultados e discussões	61

5.4	Experimento 4: Ubiris.v2	62
5.4.1	Resultados e Discussões	64
5.5	Experimento 5: FOCS	64
5.5.1	Resultados e Discussões	66
6	CONCLUSÕES E TRABALHOS FUTUROS	68
6.1	Conclusões	68
6.2	Trabalhos Futuros	69
	REFERÊNCIAS	70

1 Introdução

É possível notar que os avanços tecnológicos dos últimos anos acabaram criando um cenário onde a identificação automática de indivíduos se tornou uma grande necessidade, afinal, a sociedade atual busca cada vez mais agilizar processos por meio da tecnologia. Os métodos tradicionais de identificação, como senhas ou cartões (posse), estão cada vez mais sendo deixados de lado, para dar espaço à abordagem que busca a identificação por meio de características físicas ou comportamentais de indivíduos, chamada de Biometria ([JAIN; ROSS; NANDAKUMAR, 2011](#)).

Uma das principais vantagens da utilização de sistemas biométricos é a maior segurança contra fraudes, devido ao fato de que a identificação por meio de cartões ou senhas pode ser mais facilmente burlada, seja por meio de processos de clonagens, ou utilização de técnicas para descobrimento de senhas, sem contar que posse pode ser perdida e senha pode ser esquecida pelo próprio indivíduo ([PANKANTI; BOLLE; JAIN, 2000](#)).

A partir deste contexto, diversos tipos de sistemas biométricos são utilizados e estudados, abordando métodos por reconhecimento de uma variedade de características físicas ou comportamentais dos indivíduos. Dentre esses métodos, uma abordagem com bastante destaque é a utilização de características da íris, que possui técnicas de implementação consolidadas ([DAUGMAN, 2004](#)).

Mesmo com métodos eficazes para reconhecimento da íris, estudos mostram que para imagens da íris obtidas em condições não ideais, os resultados não são tão bons, e uma alternativa de melhoria é a fusão das informações de íris com informações da região periocular ([WOODARD et al., 2010](#)), que trata-se da região que compreende os elementos ao redor dos olhos, como cílios, pálpebras e sobrancelhas. Seja por problemas ligados a iluminação, obstrução, posicionamento, entre outros, as informações da íris do indivíduo acabam não tendo tanta precisão como deveriam, o que pode ocasionar erros de identificação no sistema biométrico. Dessa forma, imagens da região periocular que abrangem cílios, sobrancelha e pálpebras podem ser utilizadas em conjunto com informações dos olhos para se obter resultados melhores.

1.1 Motivação

Após uma fase inicial de pesquisa e desenvolvimento de sistemas biométricos, na qual apenas dados obtidos em ambientes controlados eram possíveis de serem utilizados, o interesse passou a ser o desenvolvimento de sistemas capazes de operar também em ambientes de aquisição não controlados, o que aumenta significativamente os desafios do reconhecimento biométrico (SANTOS; PROENÇA, 2013). A biometria periocular também vem ganhando espaço nesta área de pesquisa devido às preocupações com o desempenho aquém do desejável dos sistemas biométricos baseados na íris ou na face em condições não controladas (SANTOS; PROENÇA, 2013; ALONSO-FERNANDEZ; BIGUN, 2016; NIGAM; VATSA; SINGH, 2015).

Devido à capacidade de reconhecer padrões complexos de forma holística e inspirada nos modelos biológicos, os métodos de aprendizado de máquina em profundidade têm ganhado popularidade. Segundo Deng e Dong (2014), por extraírem informações de alto nível e autoaprendidas dos dados sendo tratados, as técnicas de aprendizado de máquina em profundidade conseguem obter alta abstração, poder de generalização e, conseqüentemente, robustez em suas aplicações.

Os métodos de aprendizado de máquina em profundidade têm sido aplicados com êxito em tarefas de reconhecimento de padrões em diversos domínios. As CNN (Convolutional Neural Network), por exemplo, têm demonstrado robustez às variações espaciais intraclasses, muito comuns nas aplicações da vida real, e têm sido bem sucedidas no reconhecimento de caracteres manuscritos (LECUN et al., 1998), na detecção de objetos (GIRSHICK et al., 2014), na classificação de imagens em larga escala (SZEGEDY et al., 2015).

Este sucesso também tem sido verificado na área de Biometria (SUN; WANG; TANG, 2014; WANG et al., 2014; MERAOUIMIA et al., 2017; SUN; WU; HOI, 2017). No reconhecimento de faces, por exemplo, métodos baseados em CNN têm superado significativamente os métodos tradicionais baseados em características *handcrafted* (SUN; WANG; TANG, 2014; TAIGMAN et al., 2014). No reconhecimento de íris, por exemplo, a aplicação de características profundas (*deep features*), extraídas da VGG-Net (SIMONYAN; ZISSERMAN, 2014), em um algoritmo de classificação simples, mostrou que as características aprendidas via treinamento para uma tarefa muito diferente de reconhecimento

de objetos, podem ser transferidas para o reconhecimento biométrico das íris, superando os melhores resultados obtidos por técnicas tradicionais (MINAEE; ABDOLRASHIDIY; WANG, 2016).

No entanto, para o reconhecimento da região ocular, são poucos os trabalhos encontrados na literatura baseados em aprendizado em profundidade. No levantamento bibliográfico realizado por Alonso-Fernandez e Bigun (2016) sobre pesquisas a respeito da biometria periocular, publicado em 2016, consta apenas o trabalho de Nie, Kumar e Zhan (2014). Neste trabalho, publicado em 2014, os autores propõem a CRBM (Convolutional Restricted Boltzman Machines), uma versão convolucional das RBM para realizar o reconhecimento periocular. A CRBM é uma rede neural generativa que aprende uma distribuição de probabilidade a partir de um conjunto de dados de entrada gerados por filtros que capturam a orientação das bordas e as conexões espaciais entre partes da imagem. Outro trabalho encontrado sobre este tema, publicado em 2017 por Zhao e Kumar (2017a), propõe uma nova rede neural, denominada SCNN (Semantics-Assisted Convolutional Neural Network) para reconhecer de forma eficiente e precisa imagens perioculares adquiridas automaticamente em ambientes não controlados.

A escassez de trabalhos publicados sobre o reconhecimento da região ocular para a identificação biométrica de indivíduos baseados em métodos de aprendizado de máquina em profundidade, aliada aos ótimos resultados apresentados por tais técnicas no reconhecimento de outras características biométricas mesmo em ambientes pouco controlados (e também em outros domínios de aplicação de reconhecimento de padrões em imagens) motivaram a realização desta dissertação de mestrado.

1.2 Objetivos

Esta dissertação de mestrado visa a proposição de um novo método para o reconhecimento da região ocular para a identificação biométrica de pessoas utilizando aprendizado de máquina em profundidade, em particular as Redes Neurais Convolucionais (CNN - Convolutional Neural Networks).

Com a utilização de imagens de bases de dados de faces, também foi possível efetuar a comparação entre resultados obtidos utilizando somente a região ocular e utilizando toda a região facial para se autenticar indivíduos.

1.3 Organização da Dissertação

Esta dissertação está organizada em seis capítulos. No Capítulo 1 apresenta-se a introdução ao tema da dissertação de mestrado, sua motivação e seus objetivos.

No Capítulo 2 apresenta-se o conceito de Biometria, sistemas biométricos, e uma introdução ao tipo de Biometria da íris e da face com seus desafios que motivam o uso da região ocular para estabelecer o reconhecimento de pessoas, e também a prática da fusão de características biométricas.

No Capítulo 3 apresenta-se um estudo sobre métodos correlatos, encontrados na literatura para promover o reconhecimento da região ocular.

No Capítulo 4 apresenta-se o método proposto nesta dissertação para promover o reconhecimento da região ocular para identificação biométrica de pessoas.

No capítulo 5 apresenta-se os experimentos realizados e os resultados obtidos.

No capítulo 6 apresenta-se as conclusões do trabalho e propostas para trabalhos futuros.

2 Biometria

Neste capítulo são abordados os aspectos e métodos relacionados à biometria humana, analisando seus pontos positivos em relação a abordagens de identificação tradicionais, com enfoque na biometria da região ocular. Partindo em um primeiro momento do reconhecimento da íris e da face, serão analisados os possíveis problemas e dificuldades com essa abordagem, que motivam a busca por abordagens diferentes como a biometria da região ocular e técnicas de fusão de características biométricas.

2.1 Introdução

Os constantes avanços tecnológicos do mundo atual, o grande aumento dos processos computacionais em larga escala, nos mais variados tipos de problemas, evidenciam cada vez mais o tema Biometria, que se refere ao estudo estatístico das características físicas e comportamentais de indivíduos ([JAIN; ROSS; NANDAKUMAR, 2011](#)). Isto porque o seu uso propõe uma alternativa para identificação automática de pessoas, em relação à métodos tradicionais baseados em posse ou conhecimento.

Segundo [Bolle et al. \(2013\)](#) existem basicamente três categorias de métodos para identificação de pessoas:

- **Baseados em posse:** São os métodos em que para se identificar, a pessoa precisa apresentar algum objeto, como por exemplo um cartão;
- **Baseados em conhecimento:** São os métodos em que é solicitada à pessoa alguma informação pré cadastrada para confirmar sua identificação, como por exemplo uma senha de acesso.
- **Baseados em biometria:** São os métodos baseados em características pessoais dos seres humanos, que podem ser mensuradas, e fazer a distinção singular entre os indivíduos, como por exemplo a impressão digital.

O emprego das abordagens mais tradicionais, baseadas em posse ou conhecimento, pode acarretar alguns problemas, tanto para o usuário, quanto para o sistema em si, pois

uma posse pode ser perdida ou clonada, e uma senha pode ser esquecida ou inferida.

Os sistemas de segurança baseados em biometria são uma boa alternativa para contornar alguns problemas enfrentados pelos sistemas mais tradicionais, pois a análise de uma característica biológica ou comportamental única de cada indivíduo é mais difícil de ser fraudada, e não corre o risco de ser perdida ou esquecida.

Apesar do tema ter ganhado bastante atenção no final da década de 1990 na área de acesso primário à informação, os primeiros estudos envolvendo biometria surgiram em meados do século XIX ([MAGUIRE, 2009](#)). Os estudos de Henry Faulds, William Herschel e Francis Galton, contribuíram para estabelecer uma forma de apontar singularidades em impressões digitais, identificando padrões como por exemplo, pontos de minúcia. A publicação de Mitchell Trauring em 1963 é considerado o estudo pioneiro nesta área, onde foi abordado o reconhecimento automático de impressões digitais ([JAIN; NANDAKUMAR; ROSS, 2016](#))

A princípio, os estudos relacionados à sistemas biométricos tiveram seu enfoque em aplicações forenses, e posteriormente foram se expandindo para tratar problemas envolvendo acesso restrito de pessoas. Nos dias de hoje tal tecnologia já faz parte da rotina das pessoas, sendo possível encontrar aplicações para celulares e dispositivos portáteis em geral, visando não somente a segurança, mas também a praticidade e a agilidade.

2.2 Sistemas biométricos

Dependendo de sua aplicação, um sistema biométrico pode basicamente ter finalidade de identificação ou validação. Um sistema que busca a identificação de indivíduos, opera de forma a encontrar a identidade do usuário com base apenas em sua entrada biométrica no sistema, fazendo uma busca em todo o banco de dados. Já uma abordagem de validação refere-se ao problema de confirmar a identidade do usuário com base em uma entrada prévia, em que sabendo qual indivíduo deseja-se analisar, a comparação é feita apenas com o dado biométrico cadastrado para aquele indivíduo.

A abordagem escolhida para compor um sistema biométrico dependerá muito da aplicação que se deseja desenvolver. Atualmente podemos encontrar uma tecnologia biométrica para validação em caixas eletrônicos, onde a pessoa, após inserir seu cartão de

identificação, deve confirmar a identidade por meio da leitura de impressão digital, por exemplo. Neste caso, a impressão digital é utilizada para validar se a pessoa tentando acessar a conta é realmente a dona. Já no sistema de acesso de uma academia, por exemplo, é realizada a leitura da impressão digital do indivíduo, e toda a base de dados é comparada com aquelas características biométricas para se determinar a sua identidade.

Uma vantagem do primeiro modo é que para a identificação do cliente de um banco torna-se mais segura, pois é feita de modo mais minucioso, baseada em posse e em biometria. No segundo modo, uma vantagem é a conveniência ao sócio da academia, que não precisa portar a carteira de sócio para o ingresso às suas instalações.

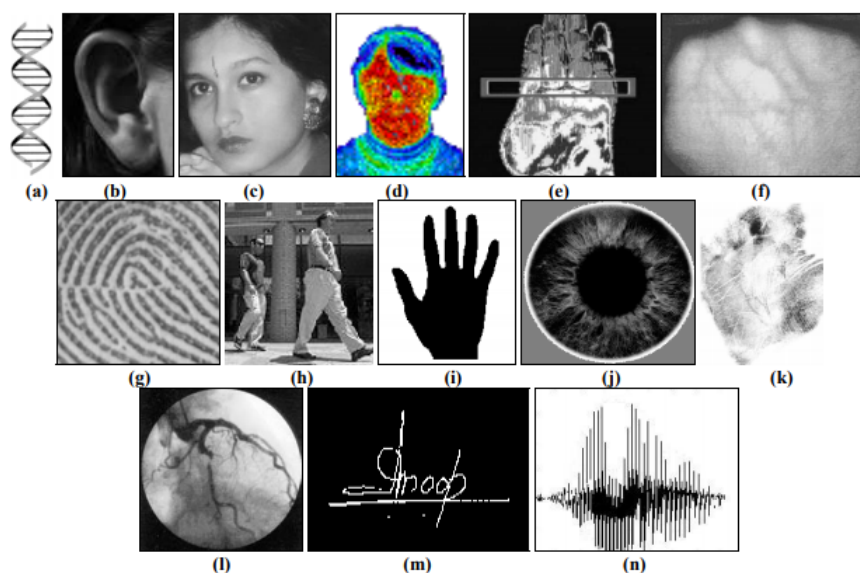
Conforme já foi definido anteriormente, um sistema biométrico realiza a análise de características biológicas ou comportamentais de indivíduos. Entretanto, para garantir que o sistema tenha um bom desempenho, isto é, consiga diferenciar indivíduos com uma margem de erro aceitável, e estabeleça o reconhecimento correto de dados de entrada genuínos, segundo [Jain, Ross e Prabhakar \(2004\)](#), é ideal que as características satisfaçam os seguintes requisitos:

- **Universalidade:** Garantir que cada pessoa tenha a característica;
- **Unicidade:** Ser suficientemente distinguível entre os indivíduos;
- **Permanência:** Ser pouco mutável com o passar do tempo;
- **Coletabilidade:** Ser mensurável quantitativamente;
- **Performance:** Promover de modo eficaz e eficiente o reconhecimento possibilitado pela característica, incluindo recursos necessários, operações, e fatores ambientais que venham a influenciar o desempenho;
- **Aceitabilidade:** Ter aceitação das pessoas que utilizarem aquela característica como forma de identificação;
- **Segurança:** Ser robusta contra fraudes, isto é, a característica não deve ser facilmente violada.

A Figura 1 ilustra algumas características biométricas que apresentam bons resultados levando em consideração os requisitos anteriormente descritos, e que, por isso, são utilizadas para compor diversos tipos de sistemas biométricos.

A Tabela 1 mostra uma comparação entre algumas características biométricas levando em consideração os requisitos apresentados.

Figura 1 – Exemplos de características biométricas. a) DNA; b) Orelha; c) Face; d) Termograma Facial; e) Termograma da Mão; f) Veias da Mão; g) Impressão Digital; h) Caminhar; i) Geometria da Mão; j) Íris; k) Impressão da Palma; l) Retina; m) Assinatura; n) Voz.



Fonte: Jain, Ross e Prabhakar (2004).

2.3 Íris

A íris é a região do olho que corresponde à um fino diafragma circular, contendo uma perfuração central chamada pupila. A função da íris é controlar a quantidade de luz que entra através da pupila por meio do esfíncter e músculos dilatadores, que ajustam o diâmetro da pupila dependendo da incidência de luz no ambiente (MASEK et al., 2003). A Figura 2, mostra uma imagem frontal do olho apontando a localização da íris e da pupila.

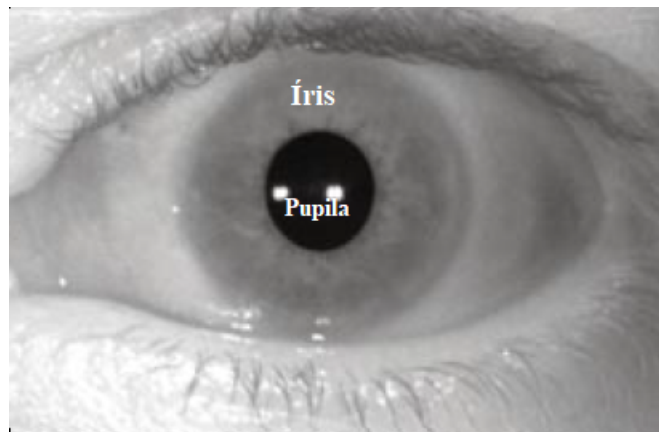
A identificação de pessoas a partir do reconhecimento da íris é um dos métodos mais populares, pelo fato de ser uma característica biométrica que satisfaz a maioria dos requisitos apresentados na subseção 2.2.

Tabela 1 – Comparação entre as principais características biométricas. A, M e B representam, respectivamente, os níveis alto, médio ou baixo de atendimento aos requisitos desejáveis das características biométricas.

Característica Biométrica	Universalidade	Unicidade	Continuidade	Coletabilidade	Performance	Aceitabilidade	Segurança
DNA	A	A	A	B	A	B	A
Face	A	B	M	A	B	A	B
Termograma facial	A	A	B	A	M	A	A
Impressão digital	M	A	A	M	A	M	M
Geomatria da mão	M	M	M	A	M	M	M
Veias da mão	M	M	M	M	M	M	A
Íris	A	A	A	M	A	B	A

Fonte: Adaptado de [Jain, Ross e Prabhakar \(2004\)](#).

Figura 2 – Imagem frontal do olho humano.



Fonte: Adaptado de [Negin et al. \(2000\)](#).

Segundo [Allen et al. \(2010\)](#), dentre as principais vantagens da íris, podemos citar:

- Estrutura física extremamente rica de dados, permitindo que as informações de íris se tornem bons descritores;
- Independência genética, garantindo que a chance de duas pessoas terem o mesmo padrão de íris seja praticamente zero;
- Padrão estável ao longo do tempo;

- A proteção estabelecida pela córnea, garante a visibilidade externa da íris e ao mesmo tempo age como protetor físico.

Imagens adquiridas da íris, geralmente no espectro infra-vermelho, também chamadas de imagens NIR (*near-infrared*), apresentam padrões complexos de textura com numerosos atributos individuais. Estas características da íris a tornam muito atrativa para compor descritores particulares de cada indivíduo (JAIN; KUMAR, 2012). Sendo assim, técnicas de processamento de imagens podem ser empregadas em imagens da íris para se obter um modelo biométrico de cada pessoa.

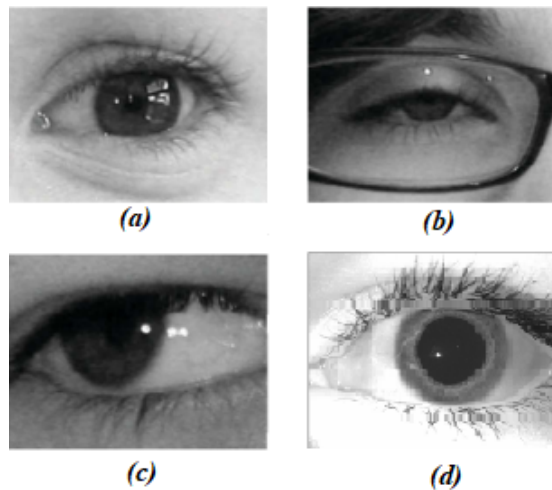
Mesmo que anteriormente protótipos já haviam sido propostos para tratar o reconhecimento da íris, foi na década de 1990 que o tema ganhou força com o método baseado em *iris code* desenvolvido pelo pesquisador da Universidade de Cambridge, John Daugman (DAUGMAN, 1994). Desde então, seu método foi estudado em diversas bases de dados apresentando resultados com taxas de erro muito próximas a zero (MASEK et al., 2003). As imagens NIR são mais utilizadas em reconhecimento de íris pelo fato dos padrões de texturas serem mais escuros, e com isso, mais difíceis de serem observados sob luz visível (JAIN; NANDAKUMAR; ROSS, 2016).

Além de boa iluminação e de um bom equipamento para capturar as imagens oculares, para se obter bons resultados é necessário cooperação do indivíduo, pois o posicionamento da cabeça e a distância entre o sensor e a pessoa afetam diretamente na capacidade de se obter boas imagens para extração de informações da íris (TAN; HE; SUN, 2010). Outro aspecto importante, são as possíveis obstruções da região da íris devido às pálpebras parcialmente fechadas ou ao posicionamento dos cílios sobre os olhos, que acabam impedindo que as informações de texturas sejam capturadas de maneira adequada.

Em se tratando do reconhecimento da íris, também é preciso considerar as alterações no padrão de textura devido às dilatações e às contrações da pupila. Conforme dito, a íris é um diafragma circular que controla a entrada de luz através da pupila, e desse modo seu padrão acaba se alterando devido às contrações ou dilatações da pupila. Segundo Ortiz, Bowyer e Flynn (2014) os efeitos da dilatação da pupila na íris aumentam a taxa de falsas não correspondências, isto é, imagens oculares obtidas de um mesmo indivíduo, mas com alterações no tamanho da pupila, podem gerar uma grande dissimilaridade entre elas, apontando para uma decisão de impostor.

A Figura 3 ilustra algumas imagens oculares com diferentes tipos de problemas associados à captura, e que podem dificultar o reconhecimento da íris.

Figura 3 – Exemplos de imagens oculares com problemas de captura. a) Reflexão especular; b) Obstrução pelas pálpebras e pelo uso de óculos; c) Posicionamento do olho; e d) Dilatação da pupila.



Fonte: Adaptado de [Tan, He e Sun \(2010\)](#).

O *iris code* gerado pelo método de Daugman representa a textura da íris em um padrão binário, entretanto, estudos como o de [Hollingsworth, Bowyer e Flynn \(2009\)](#), e [Rathgeb, Uhl e Wild \(2011\)](#), mostram que determinadas regiões deste código de bits são mais inconsistentes e que podem ocorrer maiores variações nos valores de bits para imagens diferentes de uma mesma pessoa, e dessa forma acabam prejudicando o processo de comparação de similaridade.

Os possíveis problemas que podem ser encontrados ao se trabalhar com biometria da íris foram estudados e algumas alternativas foram propostas para se obter melhores resultados com imagens em condições não tão ideais, como o estudo de [Ortiz, Bowyer e Flynn \(2014\)](#) que analisa as variações no diâmetro da pupila no momento do cadastramento do modelo biométrico do indivíduo, para se obter uma métrica de comparação entre dilatações.

Outras abordagens são utilizadas por meio do emprego de máscaras sobre o modelo binário. O estudo de [Rathgeb, Uhl e Wild \(2011\)](#) analisa as regiões do modelo binário que possuem maiores inconsistências, os chamados bits frágeis ([BOLLE et al., 2004](#)), para criar máscaras específicas para essas regiões, melhorando o reconhecimento da íris.

2.4 Face

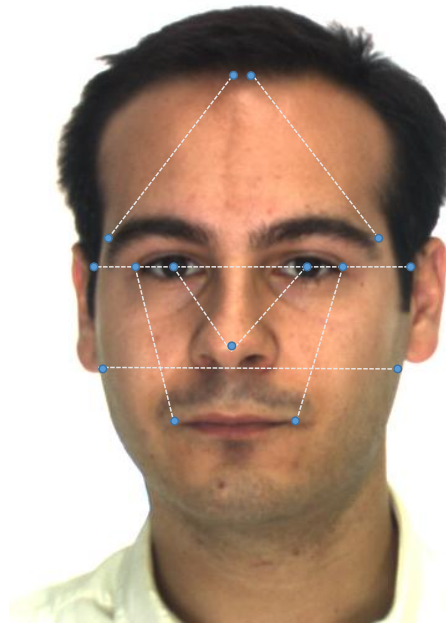
A identificação e verificação de indivíduos por meio da face remete aos anos 60, quando os estudos voltados à visão computacional se tornaram capazes de resolver os problemas que a abordagem traz (BUCIU; GACSADI, 2016). Embora a tecnologia de reconhecimento facial seja de alguma forma inferior ao desempenho em comparação com outras modalidades biométricas (como o reconhecimento da íris, por exemplo), ela é mais aceita devido à sua grande vantagem: ser uma biometria muito confiável a distância e, além disso, a autenticação dos usuários pode acontecer sem sua interação explícita com o sensor ou seu conhecimento (BUCIU; GACSADI, 2016).

O problema da biometria facial envolve, antes de se pensar na autenticação de indivíduos, a detecção da face na imagem ou vídeo, já que os sensores responsáveis por capturar essas imagens podem trazer informações do corpo da pessoa ou do ambiente. As primeiras técnicas para detecção da região facial consistia em criar uma imagem binária representando os contornos da figura, para depois extrair as características relacionadas com a face humana, como localização do topo da cabeça na matriz de contorno da imagem, nariz, boca, olhos e o eixo da face (MORAES, 2010).

Posteriormente um algoritmo ficou muito conhecido para realizar a detecção de faces, utilizando imagem integral e uma combinação de classificadores, apresentado no trabalho de (VIOLA; JONES, 2004). Mais recentemente, outro algoritmo que tem ganhado atenção é método MTCNN (*Multitask Cascaded Convolutional Networks*) (ZHANG et al., 2016), com uma abordagem utilizando redes de aprendizado em profundidade.

Com a detecção da face, os processos mais populares para se promover o reconhecimento de pessoas baseiam-se: (i) na localização e forma dos atributos faciais, determinando pontos-chaves, como olhos, sobrancelhas, nariz, lábios e queixo e suas relações espaciais, ou (ii) na análise global da imagem da face que representa uma face como uma combinação ponderada de um número de faces canônicas (JAIN; ROSS; PRABHAKAR, 2004). Para que um sistema de reconhecimento facial funcione bem na prática, ele deve automaticamente: (i) detectar se um rosto está presente na imagem adquirida; (ii) localizar a face se houver uma; e (iii) reconhecer o rosto de um ponto de vista geral (isto é, de qualquer pose) (JAIN; ROSS; PRABHAKAR, 2004). A Figura 4 mostra a imagem de uma face com a marcação de alguns pontos-chaves para se determinar o reconhecimento.

Figura 4 – Face com marcações de alguns pontos chaves para determinar o reconhecimento.



Fonte: Elaborado pelo autor.

Mesmo com os aspectos positivos da utilização da face como característica biométrica, existem cenários que acabam criando dificuldades no emprego dessa abordagem. Existem pelo menos cinco fatores chaves que afetam significativamente o desempenho de sistemas biométricos faciais ([ABATE et al., 2007](#)):

- **Iluminação:** variações na iluminação causam refletância da pele e também influenciam controles internos da câmera;
- **Pose:** mudanças na pose introduzem deformações projetivas e auto-occlusão;
- **Expressões faciais:** Variações de expressões faciais também introduzem deformações, principalmente nas regiões da boca, das bochechas e dos olhos, dificultando a autenticação ou identificação;
- **Permanência:** A face muda ao longo do tempo, de forma não linear. Em geral, este problema é mais difícil de resolver em relação aos outros, especialmente para variações de idade.

- **Oclusões:** Afeta bastante o desempenho de sistemas biométricos faciais, tornando o momento da captura um fator importante.

2.5 Região ocular

A região ocular refere-se a região do rosto que contém o olho e sua vizinhança, como os cílios, as pálpebras, as sobrancelhas e a pele ao seu redor. Características dessa região têm sido utilizadas para compor sistemas biométricos que buscam aspectos únicos em indivíduos, seja para serem utilizadas sozinhas, ou em conjunto com a íris ou a face ([SANTOS; PROENÇA, 2013](#)), uma vez que as informações da região ocular podem ser capturadas pelo mesmo sensor utilizado para capturar as imagens das íris e das faces. Desta forma, a Biometria da região ocular pode se apresentar como uma alternativa para situações em que haja dificuldades em reconhecer a íris ou a face, seja por obstrução da imagem ou pelo posicionamento inadequado do sensor no momento da captura.

A Figura 5 apresenta imagens das regiões oculares de quatro indivíduos distintos. Pode-se observar facilmente nestas imagens que as regiões oculares apresentam características distintas para indivíduos distintos. Elas também podem ser utilizadas para identificação de gênero e etnias dos indivíduos.

Figura 5 – Exemplos de regiões oculares de diferentes indivíduos



Fonte: Adaptado de [Lyle et al. \(2010\)](#).

A utilização da região ocular representa um bom compromisso entre a utilização de toda a região da face ou da utilização apenas da textura da íris, pois quando toda a face é capturada à distância, a informação da íris é tipicamente de baixa resolução; por outro lado, quando a íris é capturada com uma boa resolução (com a câmera próxima),

a face inteira pode não estar disponível, forçando assim o sistema de reconhecimento a confiar apenas na íris. Desse modo, a biometria ocular pode ser útil em uma ampla gama de distâncias do sensor ([PARK et al., 2011](#)).

Outra vantagem da utilização da região ocular como característica biométrica ocorre no momento da captura das informações, pois a aquisição de imagens dessa região é mais fácil se comparada com a aquisição da íris, que exige um nível alto de colaboração da pessoa sendo identificada. Além disso, segundo [Santos e Proença \(2013\)](#), a região ocular não é tão afetada pelo envelhecimento quanto são outras regiões da face, como a região das bochechas e da boca.

Outro aspecto importante desse tipo de Biometria é apresentado no estudo de [Juefei-Xu e Savvides \(2014\)](#), onde a região ocular se mostrou mais tolerante às mudanças de expressão facial e às oclusões, apresentando maior capacidade para se encontrar similaridades entre imagens parciais da face.

Neste contexto, as características da região ocular permitem aprimorar as técnicas de Biometria e tornam os sistemas biométricos mais robustos, e com maior aplicabilidade para situações onde se deseja realizar a identificação de pessoas, em ambientes pouco controlados.

2.6 Fusão de características biométricas

Outra técnica utilizada para melhorar a acurácia de sistemas biométricos é a fusão de características para se estabelecer o reconhecimento de um indivíduo, isto é, a união de duas ou mais características de um mesmo indivíduo que juntas acabam apresentando melhores resultados para a identificação da pessoa. O trabalho de [Sim et al. \(2014\)](#), por exemplo, apresenta uma abordagem de fusão de imagens da íris em condições não ideais, com imagens da face do mesmo indivíduo, obtendo uma melhoria no processo de reconhecimento.

Um sistema biométrico baseado em duas ou mais características biométricas é chamado de sistema biométrico multimodal ([PANCHAL; SINGH, 2013](#)). Seu uso pode ser uma boa alternativa para aumentar a robustez de um sistema biométrico, uma vez que a combinação de duas ou mais características aumentam o número de informações

capturadas de um mesmo indivíduo, criando um cenário em que as abordagens se completam e diminuem a probabilidade de encontrar erros, pois em determinado momento que uma característica não tenha bom resultado, outras poderão dar mais embasamento para se tomar uma decisão.

Outro aspecto importante do uso de sistemas biométricos multimodais é o aumento na robustez contra fraudes. Isto se dá pelo fato de que a imposição de duas ou mais características biométricas para realizar a identificação dificulta o emprego de técnicas fraudulentas, pois, um impostor que decida burlar o sistema terá que falsificar todas as características de um determinado indivíduo sendo utilizadas. Dessa forma, o sistema apresenta uma maior taxa de precisão e proteção contra falsificação.

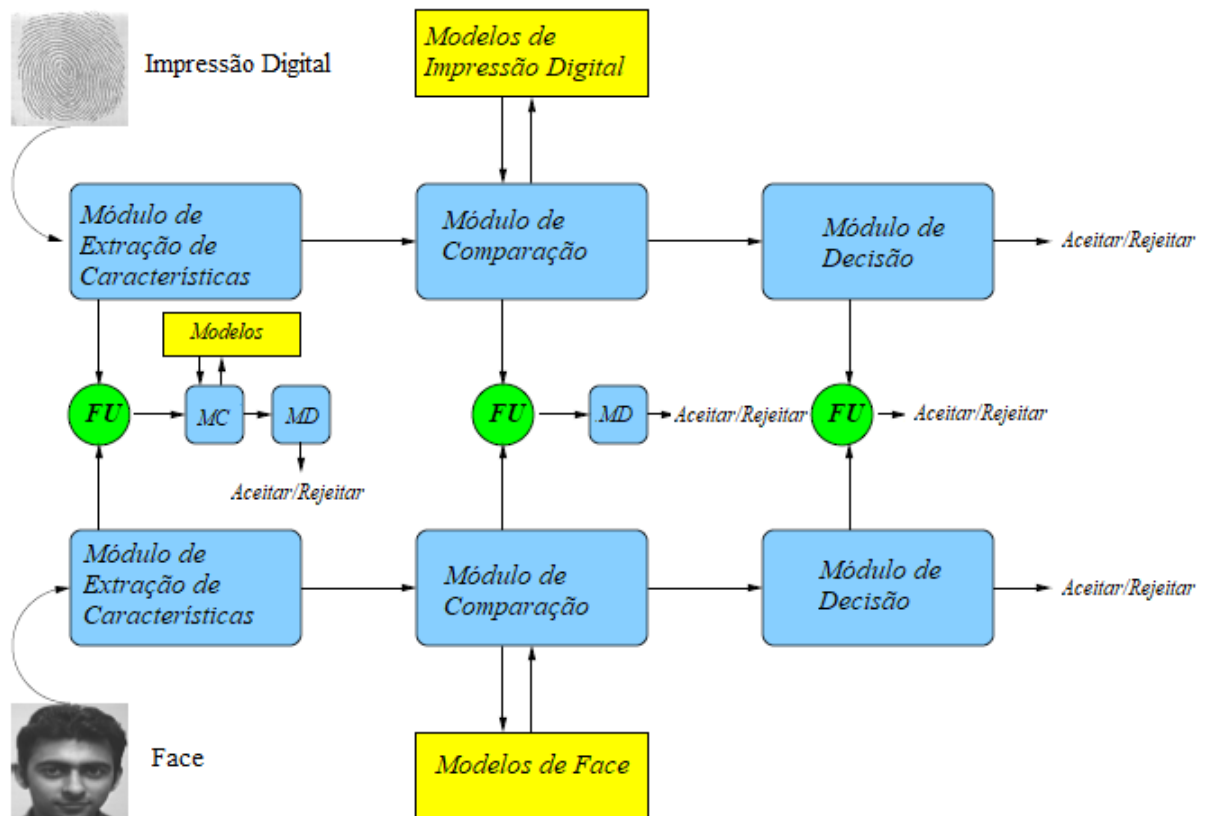
Segundo [Ross e Jain \(2003\)](#), a etapa de fusão de dois tipos de biometrias distintas pode ocorrer em três níveis distintos do processo de reconhecimento de um sistema biométrico:

1. **Nível de características:** As informações obtidas por meio dos dados capturados por diferentes sensores são transformadas em vetores de características, e seus descritores combinados, como por exemplo por meio de uma concatenação. Um novo vetor de características com mais dimensões é obtido para representar o indivíduo, ou também técnicas de redução de dimensionamento podem ser aplicadas;
2. **Nível de pontuação:** Cada sistema oferece uma pontuação de similaridade entre os descritores analisados, e essas pontuações são combinadas para afirmar ou não a veracidade do modelo sendo analisado em questão;
3. **Nível de decisão:** Cada sistema baseado na sua pontuação de similaridade, toma uma das duas decisões - aceitar ou rejeitar. Esquemas como voto da maioria ou ponderamentos específicos podem ser aplicados para se chegar à decisão final

A Figura 6 ilustra o esquema de um sistema biométrico bimodal, com as diferentes possibilidades de fusão entre duas características biométricas, levando em consideração as abordagens em três níveis. As caixas azuis maiores representam os níveis do processo de reconhecimento de um sistema biométrico, sendo que o sentido para a direita corresponde ao modelo tradicional unimodal, e ao se direcionar para o círculo verde a fusão é estabelecida

com a outra característica, que segue o restante do processo de reconhecimento, ilustrado pelas caixas azuis menores, com as informações combinadas.

Figura 6 – Sistema biométrico bimodal mostrando os três níveis de fusão



Fonte: Adaptado de Ross e Jain (2003).

Panchal e Singh (2013) propõem mais um tipo de fusão de característica a nível do sensor de captura, isto é, características biométricas obtidas por meio de sensores diferentes, ou em momentos diferentes, são combinados para originar o dado de entrada para o sistema.

Além do método de fusão em diferentes níveis, é possível realizar a combinação de biometrias em um modelo cascata. A abordagem multimodal em diferentes níveis, funciona com dois ou mais sistemas biométricos em paralelo, que em determinado momento terão suas informações combinadas para se tomar uma decisão. O modelo de classificador em cascata, por sua vez, opera em um esquema de estágios, sendo um tipo de biometria definido como um primeiro estágio a ser cumprido, limitando o processo de busca de um segundo estágio, ou simplesmente excluindo a necessidade de uma segunda verificação.

Um exemplo de biometria com classificador em cascata pode ser encontrado no trabalho de [Zheng et al. \(2012\)](#), onde informações de imagens cumulativas da pressão dos pés sobre solo durante uma caminhada, são utilizadas em um primeiro nível para limitar a base de dados de busca do segundo nível, com imagens sequencias da forma de caminhar do indivíduo.

3 Métodos para Reconhecimento da Região Ocular

Neste capítulo são abordados estudos realizados utilizando técnicas para se promover a biometria da região ocular. O capítulo está dividido em duas seções, a primeira apresenta trabalhos utilizando métodos tradicionais de processamento de imagem para o reconhecimento da região ocular, também chamados de métodos *handcrafted*. Já a segunda seção, apresenta trabalhos que usam métodos baseados em aprendizado de máquina em profundidade.

3.1 Métodos tradicionais

Esta seção apresenta trabalhos que utilizaram abordagens tradicionais, também chamados de métodos *handcrafted*, para promover o reconhecimento da região ocular.

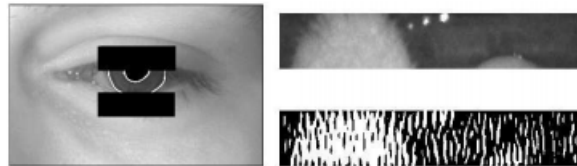
Podemos citar o trabalho de [Park, Ross e Jain \(2009\)](#) como sendo pioneiro no uso da região periocular para reconhecimento de pessoas. O estudo propõe uma abordagem de extração de características de imagens oculares em dois níveis, global e local, capturadas no espectro visível. Os descritores globais foram obtidos por meio de dois métodos, o LBP (*Local Binary Patterns*) que gera histogramas dos padrões binários locais por toda a imagem, considerando a vizinhança de cada pixel ([PIETIKÄINEN, 2005](#)), e o método HOG (*Histogram of Oriented Gradients*) que também gera histogramas baseados em informações de orientação de borda ([DALAL; TRIGGS, 2005](#)). Localmente, os descritores foram gerados a partir do método SIFT (*Scale-Invariant Feature Transform*), que permite a extração de características de pontos de interesse em uma imagem ([LOWE, 2004](#)).

A abordagem proposta pelos autores, com a fusão dos descritores na etapa de pontuação, não teve uma acurácia melhor se comparada ao reconhecimento total da face, porém, o estudo concluiu que a região periocular pode ser usada como característica biométrica, seja como alternativa para cenários em que somente imagens parciais da face são analisadas, ou como um segundo nível em um sistema biométrico de modelo cascata.

O trabalho de [Woodard et al. \(2010\)](#) propõe uma abordagem de união de informações

da íris e da região periocular, sendo que cada região tem suas características extraídas de forma isolada, e por meio de métodos distintos. A partir de imagens NIR de toda a região ocular, a íris é detectada, segmentada, e tem seu modelo binário gerado pelo método de Daugman. A Figura 7 ilustra o processo para gerar o código da íris do indivíduo.

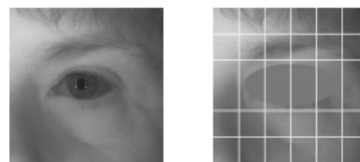
Figura 7 – Processo para gerar código da íris



Fonte: [Woodard et al. \(2010\)](#).

Para a obtenção dos descritores da região periocular, as informações da íris são excluídas da imagem ocular, e toda sua região é dividida em blocos de dimensões iguais, conforme mostra a Figura 8. Para cada um dos blocos, é gerado um descritor LBP local, que no final são concatenados para gerar um descritor global da imagem.

Figura 8 – Pré processamento da região periocular



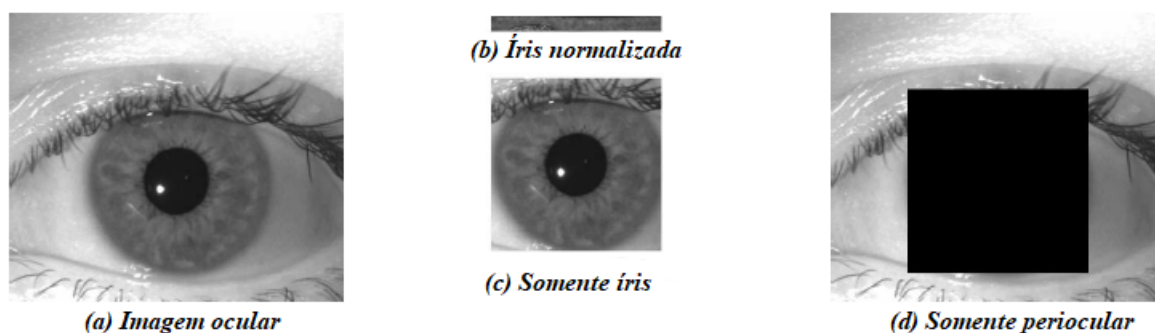
Fonte: adaptado de [Woodard et al. \(2010\)](#).

Por fim, ambos descritores são normalizados e unidos na etapa de pontuação por uma regra da soma simples, com peso 0,4 para íris e 0,6 para a região periocular. Os resultados obtidos, utilizando a base de dados MGBC (*Multi Biometric Grand Challenge*), foram melhores com a técnica de fusão, sendo que para o olho esquerdo a acurácia foi de 92,5% para a íris e 96,5% com a fusão, e para o olho direito a acurácia foi de 88,7% para a íris e 92,4% com a fusão.

As imagens da região ocular também podem ser utilizadas para a determinação do sexo de indivíduos. O recente trabalho de [Bobeldyk e Ross \(2016\)](#), analisa imagens NIR,

em quatro abordagens: (1) utilizando informações de toda a região ocular; (2) informações de íris normalizadas; (3) informações de íris sem normalização; (4) somente a região periocular. A Figura 9 ilustra as quatro abordagens analisadas.

Figura 9 – Quatro abordagens analisadas



Fonte: adaptado de Bobeldyk e Ross (2016).

Para as quatro abordagens, as imagens são submetidas ao descritor BSIF (Binarized Statistical Image Features), que opera aplicando n filtros com respostas binárias à cada pixel, gerando uma sequência de n dígitos binários para cada pixel da imagem (KANNALA; RAHTU, 2012). Posteriormente, a imagem é dividida em blocos de dimensões iguais e para cada bloco são computados histogramas a partir das respostas do descritor BSIF. Por fim, estes histogramas são normalizados e concatenados, para gerar um único vetor de características que descreve o indivíduo.

Os vetores de características gerados são utilizados para treinar uma SVM (*Support Vector Machine*), que fica responsável por classificar se uma amostra apresentada é do sexo masculino ou feminino. Analisando os resultados gerados pelas quatro abordagens, a que obteve menor taxa de erro foi por meio das imagens que apresentam informações de toda a região ocular, ao passo que os piores resultados foram obtidos com as imagens de íris normalizadas.

3.2 Métodos de aprendizado de máquina em profundidade

Nos últimos anos, uma nova área de pesquisa em aprendizado de máquina vem sendo estudada, que opera por meio de uma estrutura de aprendizado em profundidade.

Diversas técnicas foram analisadas que impactaram em uma grande quantidade de trabalhos referentes à processamento de sinais (DENG; YU et al., 2014).

Resultados teóricos indicam que a extração de características de alto nível a partir de um conjunto de dados, essenciais a sistemas que lidam com tarefas complexas nas áreas de Visão Computacional, Processamento de Linguagem Natural, etc., requer arquiteturas em profundidade, isto é, multicamadas (DENG; DONG, 2014). Tais arquiteturas são compostas por múltiplas camadas de estruturas que realizam operações pré-definidas sobre o conjunto de dados inicial. Os resultados obtidos por uma camada inferior alimentam a próxima camada. Deste modo, aos dados originais vão sendo aplicadas uma repetição de operações até se obter características complexas (de alto nível). A busca no espaço de parâmetros de arquiteturas em profundidade pode ser complicada, mas algoritmos têm sido propostos recentemente, os quais lidam com este problema com sucesso, obtendo melhores resultados do que técnicas do estado-da-arte em muitas áreas (DENG; DONG, 2014). Este aprendizado baseado em camadas é conhecido como Aprendizado de Máquina em Profundidade (*Deep Learning*).

Em 2013, Bengio já afirmava que os algoritmos emergentes de aprendizado em profundidade levam os sistemas de aprendizado de máquina à descoberta de múltiplos níveis de representação de características, propiciando alcançar resultados superiores (BENGIO, 2013). O Aprendizado de Máquina em Profundidade tem atraído a atenção da comunidade científica, comercial e industrial. Empresas como Google, Microsoft, Apple, IBM, dentre outras, têm investido em métodos de aprendizado multicamadas. A empresa Google, por exemplo, utiliza técnicas de aprendizado em profundidade em seus *softwares* de identificação de objetos (Google *Goggles*) e busca de imagens (Google *Image Search*).

A arquitetura de uma rede de aprendizado em profundidade pode ser definida por múltiplas camadas de estruturas que realizam operações pré-definidas sobre o conjunto de dados inicial. Os resultados obtidos por uma camada inferior alimentam a próxima camada, e assim por diante os dados vão sendo tratados em cada camada. Deste modo, ao final da rede se obtém características de alto nível, após as consecutivas operações realizadas ao longo das camadas.

3.2.1 Redes Neurais Convolucionais (CNN)

As Redes Neurais Convolucionais (CNN, do inglês *Convolutional Neural Networks*) são estruturas tradicionais de aprendizado em profundidade (LECUN et al., 1998). Uma CNN é constituída por uma ou mais camadas nas quais espécies de filtros (operações Convolucionais e amostragem) são aplicados aos dados de entrada (imagens bidimensionais, por exemplo) (SIMARD; STEINKRAUS; PLATT, 2003; CHELLAPILLA; PURI; SIMARD, 2006). O resultado de uma camada inferior serve de entrada para a camada imediatamente superior. Em contraste com as redes neurais totalmente conectadas, de complexa construção e uso, as CNNs compreendem redes de topologia simplificada, com camadas de convolução e amostragem, como dito, e camadas opcionais no topo, constituídas de nós completamente conectados (CHELLAPILLA; PURI; SIMARD, 2006).

Dada uma imagem bidimensional de entrada, em cada camada, um conjunto de n filtros Convolucionais K_i , com $i = 1, 2, \dots, n$, pode ser aplicado, de modo a obter várias bandas C_i da imagem original (também chamadas de *feature maps*):

$$C_i(p) = \sum_{\forall q \in N(p)} I(q) \cdot K_i(q) \quad (3.1)$$

em que q corresponde a um pixel pertencente à vizinhança de p considerada (na imagem I), isto é, $N(p)$. $I(q)$ corresponde ao valor da imagem na posição q e $K_i(q)$ corresponde ao valor na posição respectiva a q no *kernel* i .

Na realidade, o resultado das operações de convolução pode passar ainda por funções de ativações como a de retificação linear, dada por:

$$C_i(p) = \max\{C_i(p); 0\} \quad (3.2)$$

Operações de *pooling* espacial (amostragem) também são realizadas a partir dos *feature maps* da imagem original a fim de se obter invariância translacional. Em geral, emprega-se operações de *pooling* máximo:

$$M_i(r) = \max_{\forall p \in \beta(r)} \{C_i(p)\} \quad (3.3)$$

em que $\beta(r)$ corresponde à região de *pooling* referente a r , no *feature map* C_i .

Mais operações similares podem ser agregadas. Os resultados (estruturas de dados geradas, isto é, *feature maps*) são passados para camadas superiores da rede. Ao final, tem-se uma representação de alto nível dos dados originais, codificada em um vetor de características numérico (CHELLAPILLA; PURI; SIMARD, 2006; PINTO et al., 2009; BERGSTRA; YAMINS; COX, 2013; MENOTTI et al., 2015). Tal representação pode então alimentar um classificador a fim de identificar o padrão sob análise.

Vale ressaltar, entretanto, que muitas vezes a própria rede CNN pode ser transformada num classificador ao acoplar-se camadas completamente conectadas, em seu topo, com neurônios especiais, unidades *softmax*, por exemplo. Basicamente tais unidades servem para indicar a qual classe pertence um sinal de entrada e, no caso de unidades *softmax*, seguem a equação:

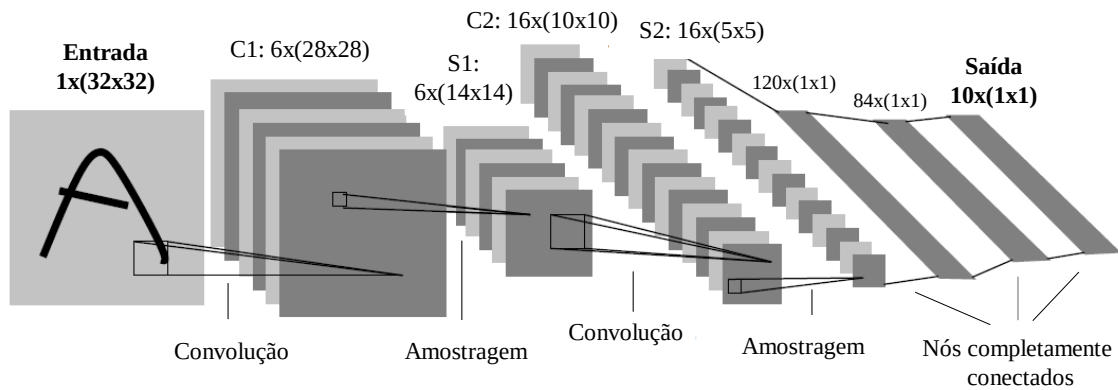
$$S_{max}(u_i) = \frac{e^{u_i}}{\sum_{j=1}^n e^{u_j}} \quad (3.4)$$

na qual u_i corresponde à soma ponderada dos sinais de entrada no neurônio *softmax* i da rede e n corresponde à quantidade de neurônios *softmax* existentes (classes do problema sendo tratado). Devido ao denominador ser uma função de partição, a saída dos neurônios *softmax* estará sempre no intervalo $[0; 1]$ e o neurônio cuja ativação for a máxima indica a classe do sinal de entrada.

A Figura 10 ilustra um exemplo de arquitetura CNN de duas camadas (sem a camada de classificação). Conforme pode-se observar, dada uma imagem inicial, operações de convolução e amostragem são aplicadas seguidas por camadas de nós completamente conectados, obtendo-se deste modo um vetor de característica reduzido de alto nível para a mesma (as dimensões e quantidades de *feature maps* gerados após cada operação da rede são indicadas na parte superior da figura).

Conforme mencionado por LeCun et al. (1998) e Simard, Steinkraus e Platt (2003), as redes convolucionais apresentam grande robustez a distorções, variações na escala e translação dos objetos nas imagens, justamente devido às operações com que trabalham. Os neurônios que respondem às operações de convolução, por exemplo, por estarem espacialmente distribuídos e compartilharem os mesmos *kernels* na geração de cada *feature map*, acabam detectando as características importantes na imagem mesmo que estas estejam deslocadas (a característica pode ficar transladada, mas continua presente no

Figura 10 – Exemplo de arquitetura de CNN



Fonte: [LeCun et al. \(1998\)](#).

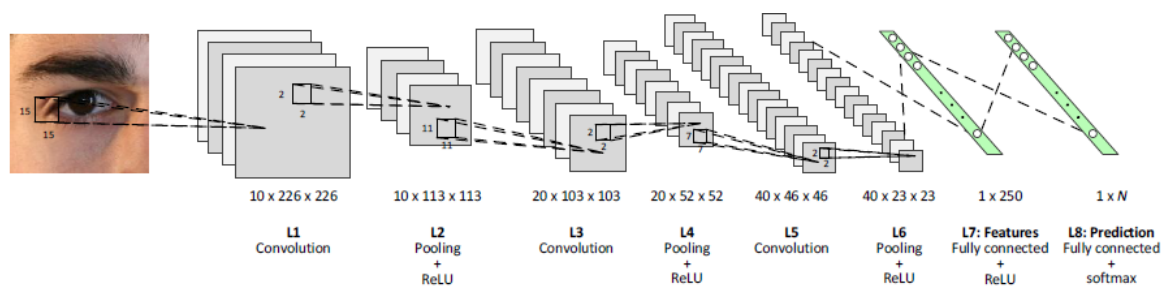
respectivo *feature map*). Os nós que respondem às operações de *pooling* espacial acabam por atenuar diferenças de escala e pequenas distorções nos objetos, preservando apenas os aspectos principais dos mesmos e o posicionamento relativo das características detectadas na operação de convolução.

Como em outras redes neurais, as CNNs também aprendem por *backpropagation* ([LECUN et al., 1998](#)). Os pesos dos *kernels* de convolução, por exemplo, podem ser vistos como os pesos sinápticos entre os neurônios das camadas da rede, os quais são inicializados e atualizados durante a aprendizagem - em [Menotti et al. \(2015\)](#), por exemplo, os autores inicializam tais pesos com valores amostrados de uma distribuição uniforme. Vale ressaltar que uma grande vantagem de tais redes em relação às redes tradicionais completamente conectadas reside no fato de que, devido aos nós compartilharem pesos sinápticos (pesos dos *kernels*), a quantidade de parâmetros livres do sistema acaba ficando menor, viabilizando o treinamento mesmo com menor quantidade de amostras disponíveis.

No contexto da biometria, estudos também têm apresentado sucesso com abordagens de aprendizado de máquina em profundidade ([SUN; WANG; TANG, 2014](#); [WANG et al., 2014](#); [TAIGMAN et al., 2014](#)). Entretanto, para a biometria da região ocular a quantidade de estudos ainda é escassa, mas é possível citar o trabalho de [Zhao e Kumar \(2017a\)](#) que busca estabelecer o reconhecimento de pessoas por meio de imagens da região ocular, utilizando uma abordagem baseada em CNN.

A Figura 11, ilustra a arquitetura da rede utilizada no trabalho de [Zhao e Kumar \(2017b\)](#), com seu final sendo composto por uma etapa de classificação com camadas totalmente conectadas. Os autores ainda utilizam uma variação da rede, adotando informações semânticas sobre o determinado indivíduo sendo analisado, obtendo melhores resultados do que os métodos tradicionais para o reconhecimento de pessoas. O método foi analisado em diferentes bases de dados, sendo que em todas as bases analisadas o método obteve melhores resultados de acurácia comparados a métodos tradicionais, com diminuição do erro de até 13,02%

Figura 11 – Estrutura da rede CNN implementada no trabalho de [Zhao e Kumar \(2017a\)](#)



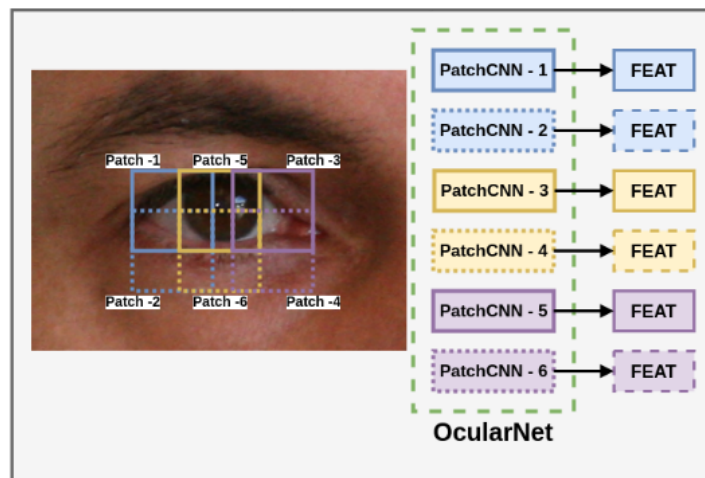
Fonte: [Zhao e Kumar \(2017a\)](#).

Em ([REDDY; RATTANI; DERAKHSHANI, 2018](#)) os autores propõem uma nova arquitetura de CNN chamada OcularNet, baseada na detecção de seis pontos chaves na região ocular e no treinamento de pequenas redes CNNs para cada uma das regiões. A Figura 12 mostra um exemplo da aplicação da etapa de extração de características deste trabalho.

Na etapa de verificação entre amostras, a distância euclidiana é computada para cada um dos seis segmentos obtidos pelas pequenas CNNs, e uma fusão das pontuações é realizada para se obter o resultado final da verificação. Os experimentos mostraram que a arquitetura proposta da OcularNet obteve um melhoria de pelo menos 8% do erro em comparação com a arquitetura popular da Resnet50.

Também em ([AHUJA et al., 2017](#)) uma estratégia utilizando CNNs é utilizada para a biometria da região ocular. Neste caso, uma abordagem híbrida utilizando uma CNN supervisionada e outra não-supervisionada. O método é testado para situações de captura de imagens por smartphones. Os resultados obtidos dos experimentos com a base

Figura 12 – Etapa de extração de característica da rede OcularNet.



Fonte: Adaptado de [Reddy, Rattani e Derakhshani \(2018\)](#).

de dados VISOB mostraram uma performance melhor do que os resultados obtidos pelos líderes do desafio ICIP VISOB ([RATTANI et al., 2016](#)), com uma taxa de verificação de 99,5%.

Analisando este contexto, as CNNs foram objeto de estudo dos experimentos desenvolvidos nesta dissertação de mestrado para realizar a extração de características da região ocular de cada indivíduo, buscando analisar a performance para se obter uma boa acurácia em um método de autenticação biométrica, e utilizando da possibilidade de fusão entre região ocular e face, em bases de dados com informações de toda a face.

4 Método Proposto

O método proposto neste trabalho para a identificação de pessoas por meio do reconhecimento da região ocular utilizando aprendizado em profundidade é constituído por dois módulos, um módulo para a extração de características e em outro módulo para o reconhecimento da região ocular. Entretanto, nosso objetivo também foi analisar o uso da região ocular em base de dados de imagens faciais, em relação à abordagens usando toda a face. Sendo assim, utilizamos um método para detecção e segmentação da face e regiões oculares em determinadas bases de dados em nossos experimentos. As próximas três seções abordam o método utilizado para detecção e segmentação da face e regiões oculares, o módulo para extração de características e o módulo de reconhecimento da região ocular, respectivamente.

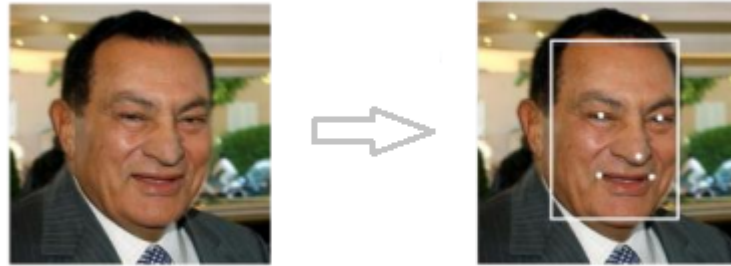
4.1 Segmentação facial e ocular

Além das bases de dados contendo somente a região ocular, utilizamos em nossos experimentos bases que são compostas por imagens obtidas com o indivíduos mais distante do sensor de captura, isto é, imagens que apresentam o rosto e até mesmo o corpo da pessoa. Para este tipo de base de dados é primordial que se faça uma etapa de pré-processamento realizando a segmentação do rosto e posteriormente a região ocular do indivíduo na cena.

Para a etapa de pré-processamento de segmentação da face foi utilizado o método MTCNN (*Multitask Cascaded Convolutional Networks*), que foi desenvolvido recentemente e superou resultados quando comparado à métodos mais tradicionais da literatura ([ZHANG et al., 2016](#)). O método MTCNN além de possibilitar a segmentação do rosto realizando o enquadramento da face, também retorna as informações de posicionamento de olhos, nariz e boca na imagem. A Figura 13 mostra um exemplo da detecção da facial utilizando método MTCNN, e de pontos chaves do rosto (olhos, nariz e boca).

Estas informações obtidas pela aplicação do método MTCNN possibilitam que as regiões oculares da esquerda e direita sejam facilmente detectadas utilizando heurísticas baseadas na antropometria da face humana, isto é, com o posicionamento dos olhos, nariz e enquadramento da face, traçar segmentos para cada uma dos olhos que enquadrem a região

Figura 13 – Exemplo de uso do método MTCNN para detecção da face



Fonte: Adaptado de [Zhang et al. \(2016\)](#).

ocular é fácil. A Figura 14 mostra o resultado obtido na aplicação do método MTCNN em um dos experimentos deste trabalho, na base de dados faciais AR Face ([MARTINEZ; BENAVENTE, 1998](#)).

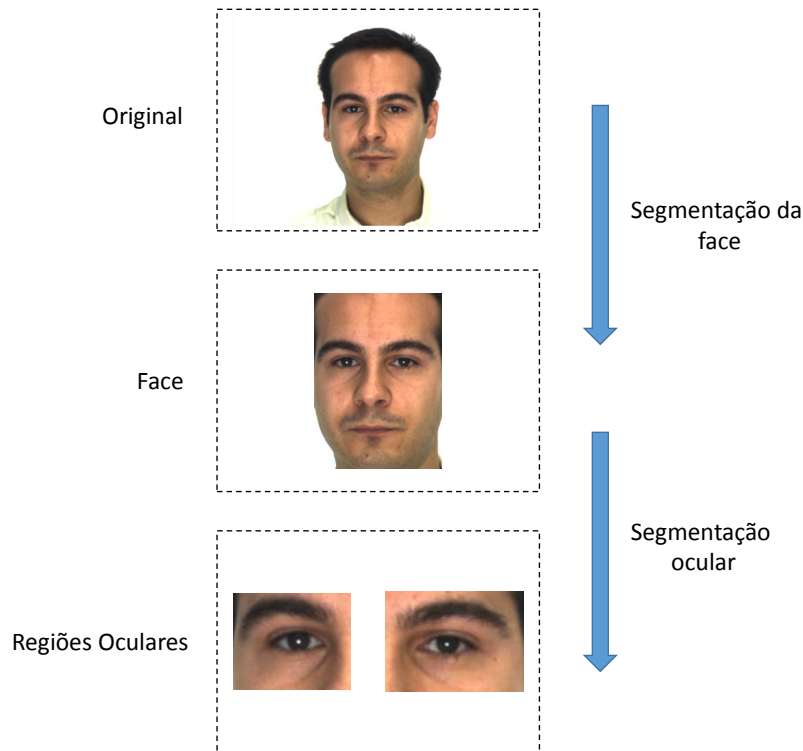
4.2 Módulo de Extração de Características

Este módulo é responsável por receber como entrada as imagens das bases de dados utilizadas nos experimentos, seja bases oculares, ou bases faciais que passaram pela aplicação do método MTCNN, e retornar como saída o vetor de características de cada amostra. Dessa forma, cada imagem que é submetida ao módulo de extração de características tem uma representação numérica em forma de vetor de características, que possui dimensões conforme a arquitetura da rede de profundidade utilizada.

As características são extraídas por meio de CNNs pré treinadas, isto é, arquiteturas de redes de aprendizado em profundidade previamente treinadas com diferentes imagens, em diferentes cenários. As imagens submetidas nas redes passam por sucessivas camadas de convoluções, nas quais a saída de uma camada funciona como entrada para outra. Dessa forma, as características extraídas em cada camada vão se aprofundando, e antes da etapa de classificação, são extraídas da rede para compor o vetor de características de cada amostra, pois em nosso trabalho as CNNs foram utilizadas somente como extrator de características, e não classificadores.

Como abordado na seção anterior, as bases de dados de imagens faciais passaram

Figura 14 – Resultado obtido na aplicação do método MTCNN em um dos experimentos deste trabalho, na base de dados faciais AR Face ([MARTINEZ; BENAVENTE, 1998](#)).



Fonte: Elaborado pelo autor.

pela aplicação do método MTCNN, dessa forma, cada indivíduo teve sua representação em termos de região facial completa, região ocular esquerda e região ocular direita, separadamente. Portanto, as três abordagens passaram pela extração de características separadamente também.

4.3 Módulo de Reconhecimento da Região Ocular

Este módulo é responsável pela identificação biométrica do indivíduo a partir dos vetores de características obtidos no módulo de extração de características descrito na seção 4.2. A identificação foi realizada no modo de autenticação, no qual o objetivo principal é verificar se a característica biométrica obtida do indivíduo no momento da identificação coincide com a característica biométrica armazenada no banco de dados no registro daquele indivíduo. Em muitos trabalhos que realizam autenticação a extração

de características é independente para cada imagem, sem analisar possíveis relações nas primeiras etapas do pipeline de autenticação. Nestes casos, as relações existentes entre os vetores de características não são modeladas a priori para os estágios de treinamento e de classificação. Assumindo a hipótese de que a modelagem das relações existentes entre vetores de características pode ser útil para aumentar a robustez e o desempenho das tarefas de autenticação, neste trabalho é proposta uma abordagem que compara um par de regiões oculares por meio de características relacionais existentes entre os vetores de características. A diferença dos vetores de características, chamado de vetor diferença, é considerada como sendo o vetor de características a ser apresentado ao classificador que decide se a comparação é genuína ou impostora.

Os vetores de características no primeiro módulo são comparados com amostras intra-classe, e amostras inter-classe, gerando os vetores diferenças, que são os descritores das classes genuína e impostora respectivamente. Desta forma, o classificador fica responsável por determinar se uma comparação é impostora ou genuína, tornando o problema que era multi-classe (cada indivíduo da base de dados sendo uma classe), em binário (genuíno ou impostor).

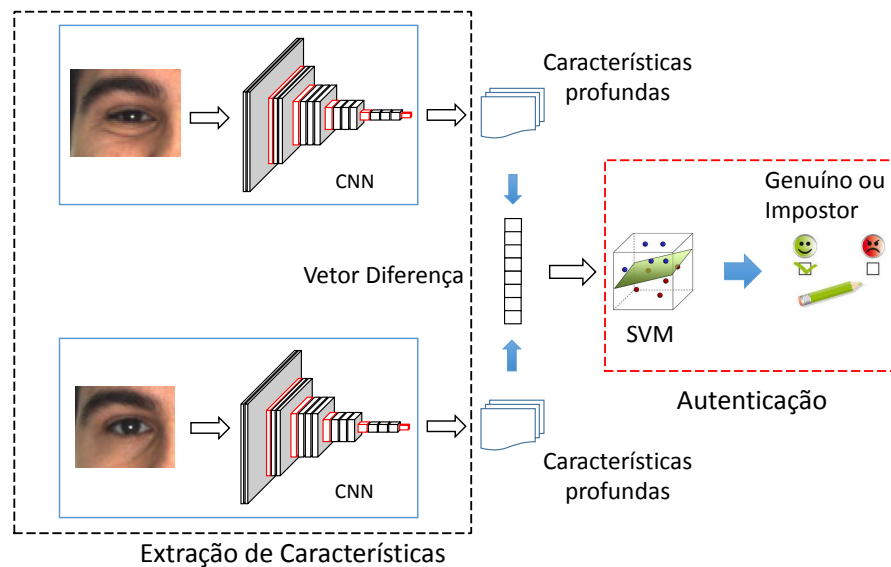
O classificador utilizado é uma máquina de vetores de suporte (SVM, *Support Vector Machine*) (GUENTHER; SCHONLAU et al., 2016). Como este classificador fornece as probabilidades do vetor de características pertencer às classes, o valor de probabilidade da classe genuína é tomado como sendo a pontuação da comparação e permite calcular as curvas ROC (Receiver Operating Characteristics).

Para a etapa de treinamento do classificador uma porção da base de dados é utilizada para gerar duas classes distintas, uma que representa a diferença entre vetores de características de classes semelhantes, comparação genuína, e outra que representa a diferença entre vetores de características de classes distintas, comparação impostora. Um conjunto de vetores de diferenças para cada uma das classes é criado e utilizado para treinar o classificador SVM.

O classificador SVM treinado, tem como função decidir se uma amostra de vetores de diferença à ele submetido refere-se a uma comparação genuína ou impostora. Desta forma, a outra porção da base de dados é utilizada para se testar a eficiência do classificador em separar as duas classes corretamente.

A Figura 15 apresenta um diagrama do método proposto neste trabalho para a autenticação de pessoas por meio das características biométricas da região ocular, sendo que no exemplo ilustrado trata-se de uma comparação entre duas imagens do mesmo indivíduo que portanto espera-se que se obtenha uma verificação genuína.

Figura 15 – Diagrama do método proposto para a autenticação biométrica de pessoas por meio de características profundas.



Fonte: Elaborado pelo autor.

Após a etapa de teste de cada abordagem (face e oculares), as pontuações determinadas pelo classificador para cada amostra foram combinadas em uma fusão no nível de pontuação, devido a possibilidade que a saída em forma de probabilidades do classificador permite. A fusão é realizada por uma soma ponderada, com pesos iguais para todas as características, das pontuações obtidas em cada abordagem.

5 Experimentos

Este Capítulo descreve os experimentos realizados utilizando o método proposto neste trabalho. Cinco experimentos diferentes foram conduzidos, cada um com uma base de dados específica de imagens somente da região ocular ou de imagens faciais. Cada base de dados possibilita analisar abordagens diferentes em um sistema biométrico, com base nas características em que as imagens foram capturadas, como por exemplo diferentes expressões faciais e iluminação, distâncias do indivíduo em relação ao sensor de captura, oclusões parciais e imagens capturadas ao ar livre.

Todos os experimentos foram inteiramente desenvolvidos com a linguagem de programação Python, que já possui bibliotecas especializadas para a implementação de redes de aprendizado em profundidade e processamento de imagens. Sendo assim, a biblioteca Keras com base em TensorFlow foi utilizada para implementar as CNNs. O ambiente experimental foi configurado em um computador com sistema operacional Linux Ubuntu 16.04 LTS, Intel® Core™ i5-3230M CPU @ 2.60GHz 4, 8GB de memória RAM e um disco rígido de 500GB. Os experimentos desenvolvidos foram implementados com a abordagem utilizando processamento em CPU.

Em todos os experimentos a estratégia de diferença entre pares foi utilizada para gerar os descritores das classes genuínas e impostoras, em um Vetor Diferença. Os conjuntos foram divididos em subconjuntos de treinamento e teste, em que o subconjunto de treinamento foi utilizado para treinar um classificador SVM com núcleo RBF (*Radial Basis Function*). O subconjunto de teste foi submetido ao classificador que retorna a probabilidade da comparação em questão ser genuína ou impostora.

As pontuações obtidas na fase de teste são computadas em termos da probabilidade da amostra ser genuína, e projetadas em uma curva ROC, sendo que eixo das abscissas refere-se à taxa de falsa aceitação (FAR), e o eixo das ordenadas à taxa de falsa rejeição (FRR). Portanto, os valores de FAR correspondem às amostras que tratam-se de comparações impostoras mas foram consideradas genuínas pelo classificador, e os valores de FRR as comparações que tratam-se de casos genuínos que foram considerados impostoras. Sendo assim, quanto mais perto a curva está dos eixos, menor é o erro na autenticação do sistema

biométrico.

Outra métrica analisada nos experimentos é o valor de EER (*Equal Error Rate*), que representa o valor em que FAR e FRR são iguais. A análise desse valor é importante para determinar um ponto de equilíbrio do sistema de autenticação biométrica, para não permitir que muitos impostores sejam aceitos, e nem muitos genuínos sejam recusados.

As próximas cinco seções tratam dos experimentos realizados com cada uma das bases de dados, especificando os protocolos experimentais em cada caso, e os resultados obtidos expressos em projeções de curvas ROC e dos valores de EER.

5.1 Experimento 1: UBIPr

O conjunto de dados UBI Reconhecimento Periocular (UBIPr) ([PADOLE; PROENCA, 2012](#)) fornece meios de avaliar os métodos baseados em biometria periocular em níveis mais elevados de heterogeneidade. Os fatores de ruído desta base foram introduzidos por meio da configuração da aquisição: variação das distâncias de aquisição, iluminação irregular e variações de pose. As dimensões das imagens variam de acordo com a distância de aquisição, entre 501×401 (8m) e 1001×801 (4m). A base de dados possui um total de 10.950 imagens capturadas de 261 sujeitos. A Figura 16 mostra um exemplo de diferentes imagens do conjunto UBIPr capturadas de um mesmo indivíduo.

Figura 16 – Diferentes imagens do conjunto de dados UBIPr ([PADOLE; PROENCA, 2012](#)) capturadas de um mesmo indivíduo.



Fonte: Elaborado pelo autor.

Os experimentos realizados com a base de dados UBIPr tiveram como objetivo analisar o uso da região ocular como características biométrica em sistemas de autenticação, por meio do método proposto neste trabalho. Um aspecto do experimento foi a comparação

do uso de diferentes Redes Neurais Convolucionais para a etapa de extração de características das imagens, e também, se o fato da rede ter sido treinada com imagens faciais melhora o desempenho em relação às redes treinadas com imagens aleatórias. Outro objetivo do experimento foi comparar o uso de um classificador para a etapa de autenticação em relação à aplicação de funções de distância para determinar se um par de amostras trata-se de uma comparação genuína ou impostora.

Seguindo a divisão do banco de dados em subconjuntos de treinamento e teste, conforme em (ZHAO; KUMAR, 2018), 3359 das imagens da região ocular esquerda foram utilizadas para gerar os vetores de características de diferença genuína e impostora para treinar o classificador SVM. O conjunto restante das imagens da região ocular esquerda foi usado para gerar os vetores de características de diferença que foram usados na etapa de teste.

Neste primeiro experimento com a base de dados UBIPr, para a extração de características, foram avaliadas cinco CNNs diferentes pré treinadas: Resnet50 (HE et al., 2016), VGG16 (SIMONYAN; ZISSERMAN, 2014), VGG19 (SIMONYAN; ZISSERMAN, 2014), Xception (CHOLLET, 2017) e VGG-Face (PARKHI; VEDALDI; ZISSERMAN, 2015), todos os modelos foram implementados com a biblioteca Keras. A VGG-Face foi treinada usando um banco de dados com milhares de imagens faciais¹, enquanto as outras quatro CNNs foram treinadas usando a base de dados ImageNet², com milhões de diferentes imagens.

Para todas as cinco redes, os vetores de características foram extraídos da camada totalmente conectada antes da etapa de classificação. As redes Resnet50 e Xception resultaram em vetores de características com 1000 dimensões, enquanto as outras três redes resultaram em vetores de características de 4096 dimensões, devido à arquitetura de cada uma das redes.

Para gerar vetores de características de diferença genuína, cada vetor de característica obtido de um indivíduo foi subtraído de todos os outros vetores de característica obtidos do mesmo indivíduo. Para gerar vetores de características de diferença de impostor, cada vetor de característica obtido de um indivíduo foi comparado com todos os outros vetores de características obtidos de todos os outros indivíduos diferentes.

¹ http://www.robots.ox.ac.uk/vgg/data/vgg_face/

² <http://www.image-net.org>

Finalmente os descritores obtidos pelo conjunto de imagens de treinamento foram apresentados ao classificador SVM, e posteriormente foi realizada a etapa de teste com o conjunto restante. E também, além da utilização do vetor diferença e o treinamento de um classificador SVM, outra abordagem foi analisada nos experimentos, onde os *scores* de genuínos e impostores foram determinados por funções de distância entre as amostras, isto é, em vez de gerar o vetor diferença entre as amostras e treinar o classificador, apenas as funções de distâncias euclidiana e cosseno foram usadas, e os valores também usados para computar as curvas ROC. A próxima subseção mostra os resultados obtidos nos experimentos com a base de dados UBIPr.

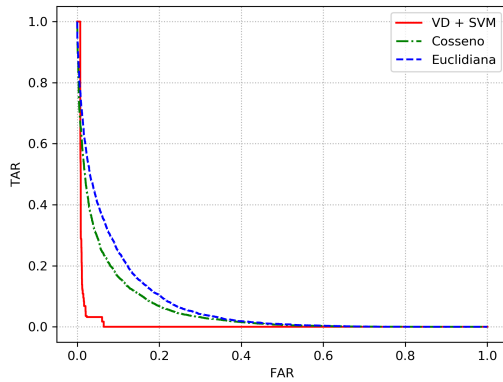
5.1.1 Resultados e Discussões

A Figura 17 mostra os resultados obtidos por cada uma das redes CNNs analisadas. A curva vermelha (—) refere-se à abordagem utilizada pelo método proposto com a aplicação do vetor diferença e uso do classificador SVM, a curva azul (— —) refere-se à abordagem utilizando a função de distância euclidiana e a curva verde (— · —) a função de distância cosseno.

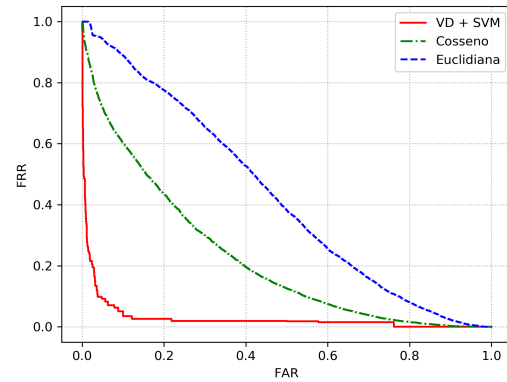
A Tabela 2 mostra os valores de EER obtidos para cada uma das cinco CNNs testadas, nas três abordagens diferentes, onde é possível analisar que o melhor resultado foi obtido pelo método proposto utilizando a rede VGG-Face para a etapa de extração de características, com o EER de 3,28%. As outras redes que não foram treinadas com imagens faciais obtiveram valores de erro muito maiores, e a rede Resnet50 obteve o valor mais próximo do melhor resultado com EER de 7,18%. Os experimentos utilizando as funções de distâncias obtiveram piores resultados do que o método proposto utilizando o classificador SVM em quase todos os casos, apenas com a utilização da rede VGG16 a função cosseno obteve o valor de 20,70% de erro contra 30,23% do método proposto.

Trabalhos recentes em biometria da região ocular no espectro visível utilizando o banco de dados da UBIPr mostram resultados próximos aos apresentados neste trabalho. Na pesquisa (RATTANI; DERAKHSHANI, 2017) os autores listam quatro artigos que utilizaram este banco de dados, mas aplicaram diferentes técnicas, com resultados variando de 6,4% a 20% de EER. Um trabalho recente (HERNANDEZ-DIAZ; ALONSO-FERNANDEZ; BIGUN, 2018), que explora a fusão de redes profundas com técnicas

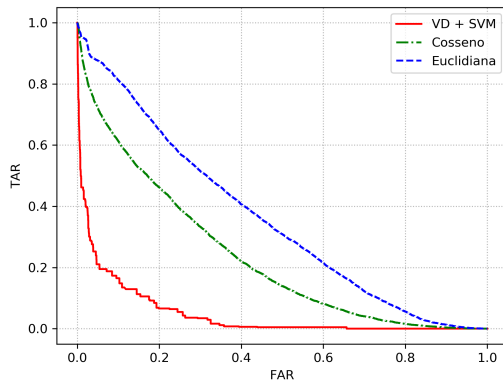
Figura 17 – Curvas ROC obtidas por meio de características profundas extraídas de cinco diferentes redes neurais profundas. Em vermelho as curvas ROC obtidas pelo método proposto com o classificador SVM, em verde as curvas ROC obtidas pela aplicação da distância Cosseno, e em azul as curvas ROC obtidas pela aplicação da distância Euclidiana ao conjunto de imagens oculares do conjunto de dados UBIPr ([PADOLE; PROENCA, 2012](#)).



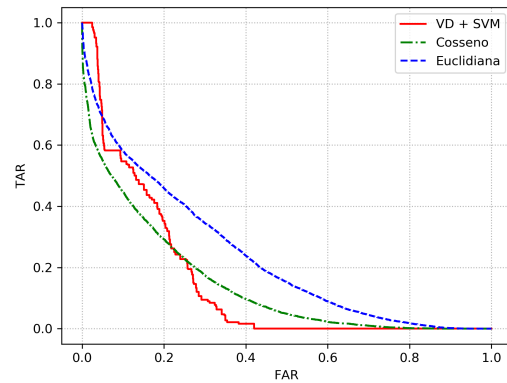
(a) VGG-Face



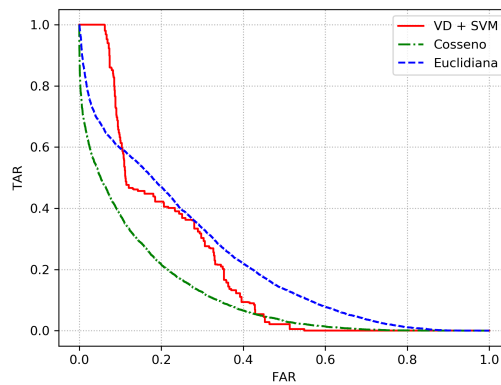
(b) Resnet50



(c) Xception



(d) VGG19



(e) VGG16

Fonte: Elaborado pelo autor.

Tabela 2 – Resultados experimentais com a base de dados UBIPr (PADOLE; PROENCA, 2012).

Rede	EER		
	Vetor diferença + SVM	Distância cosseno	Distância Euclidiana
Resnet50	7.11%	29.98%	45.19%
VGG16	30.23%	20.70%	31.53%
VGG19	23.97%	24.06%	32.35%
Xception	12.90%	31.52%	40.36%
VGG-Face	3.18%	12.74%	15.25%

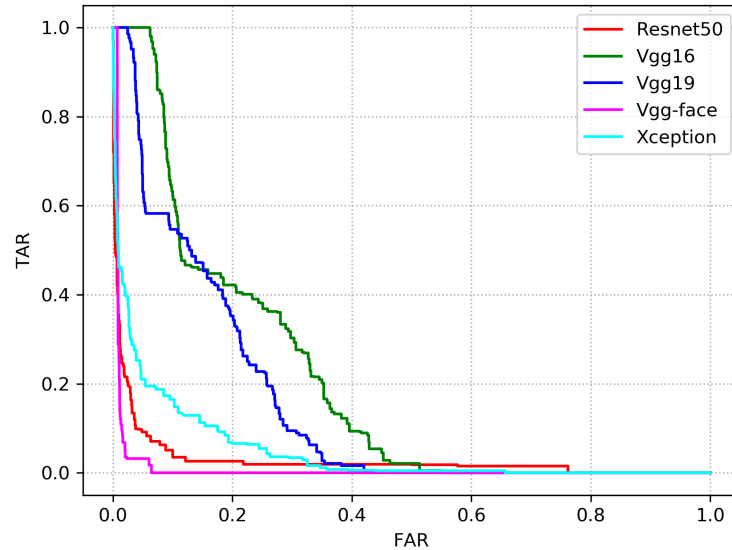
tradicionais, obteve 5,1% de EER, com a melhor abordagem analisada na base de dados UBIPr. Dois outros trabalhos muito recentes (ZHAO; KUMAR, 2017b; ZHAO; KUMAR, 2018), que adotaram o mesmo protocolo experimental adotado em nosso trabalho para o banco de dados UBIPr, obtiveram valores EER de 4.67% (ZHAO; KUMAR, 2017b) em uma estratégia que incorpora informação semântica explícita para automaticamente recuperar características perioculares abrangentes e 2.26% (ZHAO; KUMAR, 2018) em uma estratégia que propõe uma nova arquitetura de aprendizagem profunda que incorpora um modelo de atenção para enfatizar regiões importantes nas imagens perioculares.

A Figura 18 mostra o resultado obtido pela melhor abordagem em cada uma das cinco redes de aprendizado em profundidade analisadas. Pode-se verificar que a VGG-Face obteve o melhor resultado em termos da curva ROC.

5.2 Experimento 2: AR Face

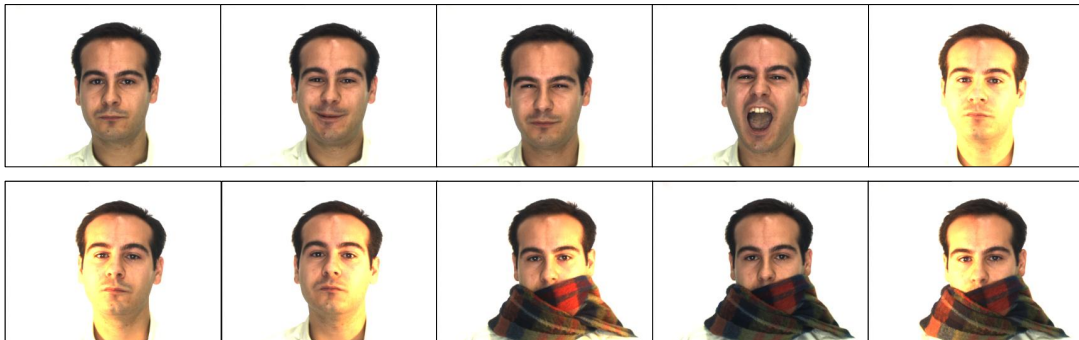
No segundo experimento foi utilizada a base de dados AR Face. O conjunto AR Face (MARTINEZ; BENAVENTE, 1998) consiste em mais de 3.000 imagens coloridas de 116 pessoas (63 homens e 53 mulheres), sem restrições relacionadas a roupas, óculos, maquiagem, penteado, etc. As imagens tem 768×576 pixels e 24 bits de profundidade. Esta base de dados é desafiadora para sistemas biométricos por conter variações em iluminação e expressões faciais, e oclusões parciais da face por óculos escuros e cachecol. Como nossos experimentos procuram analisar o desempenho do uso da região ocular em comparação com a face inteira para autenticação de identidade, imagens com pessoas usando óculos escuros não foram usadas em nossos experimentos. A Figura 19 mostra um exemplo de imagens diferentes de um mesmo indivíduo do conjunto AR Face, evidenciando a diversificação que há entre as amostras de cada indivíduo.

Figura 18 – Curvas ROC obtidas pelas cinco CNNs avaliadas neste experimento como meio de obtenção de características oculares profundas. Como se pode observar o VGG-Face obteve o melhor resultado.



Fonte: Elaborado pelo autor.

Figura 19 – Diferentes imagens do conjunto de dados AR Face ([MARTINEZ; BENAVENTE, 1998](#)) capturadas de um mesmo indivíduo.



Fonte: Elaborado pelo autor.

A escolha da base de dados AR Face para o experimento se deu principalmente pela possibilidade de analisar o uso da região ocular em relação a face toda em situações em que há oclusões parciais do rosto por cachecol. A partir do banco de dados AR Face, 100 indivíduos foram selecionados aleatoriamente (50 homens e 50 mulheres). Para verificar a capacidade de autenticação da região ocular, foram selecionados dois conjuntos de imagens. No primeiro, as primeiras 6 imagens faciais de cada indivíduo foram selecionadas para o

estágio de treinamento, e 4 imagens faciais das imagens remanescentes dos mesmos 100 indivíduos, sem a presença de óculos escuros ou cachecol, foram selecionadas aleatoriamente para a etapa de teste. Sendo assim, no primeiro conjunto da base AR Face, em ambos subconjuntos de teste e treinamento não haviam imagens de indivíduos com oclusões.

Para o segundo conjunto, utilizamos as mesmas imagens do subconjunto de treinamento utilizadas no conjunto 1, mas para o subconjunto de teste as 4 imagens faciais foram selecionadas somente na situação em que os indivíduos usavam cachecol, ou seja, subconjunto de treinamento com imagens sem oclusões, e subconjunto de teste somente de imagens com oclusões, que acaba criando um cenário que dificulta a autenticação biométrica.

O primeiro passo para aplicação do método proposto na base de dados foi a utilização do método MTCNN para detecção e segmentação da face do indivíduo na imagem, e também a segmentação das regiões oculares esquerda e direita, conforme abordado no Capítulo 4. Com as imagens segmentadas, primeiramente foi analisada a abordagem usando toda a face para os conjuntos 1 e 2, onde o teste se deu com imagens sem oclusões e imagens com oclusões respectivamente. Posteriormente, foram utilizadas as imagens oculares esquerda e direita do conjunto 2, separadamente. E finalmente, foi analisada a fusão entre face, ocular esquerdo e ocular direito, da abordagem utilizando o conjunto 2. Sendo assim, cinco diferentes experimentos foram conduzidos utilizando a base de dados AR Face.

Após a etapa de aplicação do método MTCNN, e a segmentação da face e regiões oculares, as imagens foram submetidas à rede CNN VGG-Face, conforme utilizado no experimento com a base de dados UBIPr pois esta apresentou menores taxas de erro. Neste experimento as características profundas foram extraídas antes mesmo da camada totalmente conectada da rede, responsável pela classificação. Sendo assim, o vetor característica extraído tem dimensões de $7 \times 7 \times 512$, ou seja, 512 núcleos (kernels) de 49 características que possuem a representação numérica de cada amostra submetida a rede.

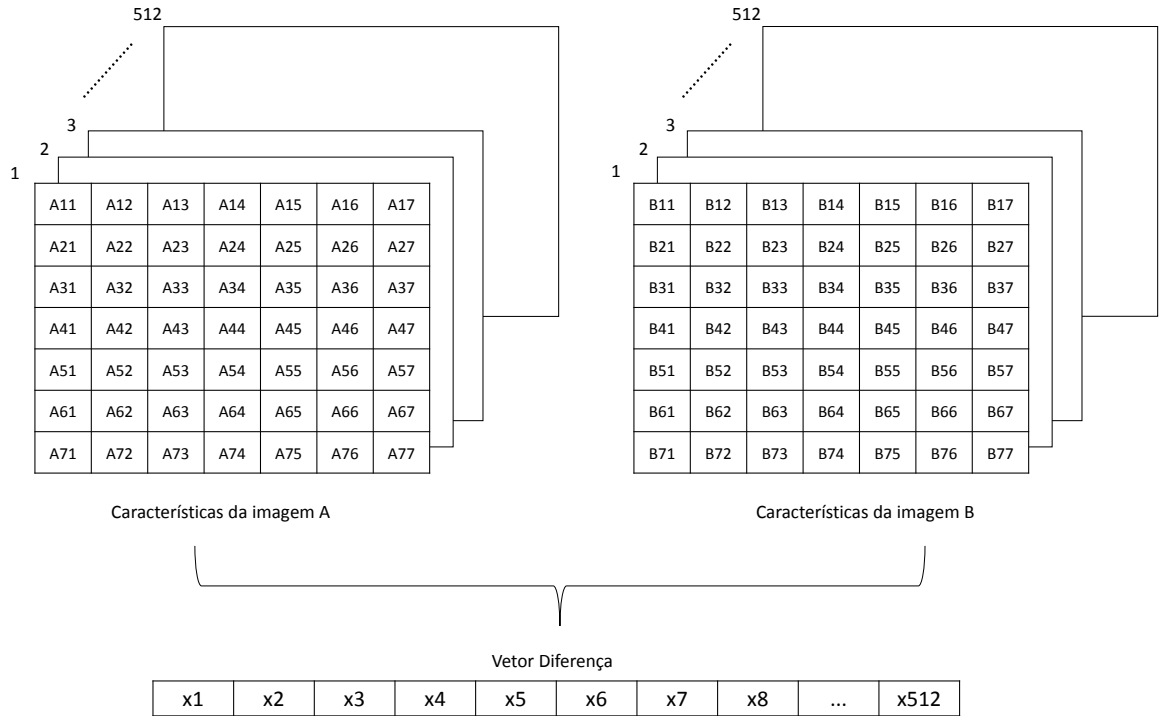
Os vetores de diferença entre as imagens de amostra são obtidos calculando as diferenças entre cada núcleo. Assim, para cada um dos 512 núcleos, um valor de diferença entre as 49 características de duas amostras é obtido, resultando em um vetor de diferença de 512 características para cada comparação. A Figura 20 mostra um exemplo de obtenção

do vetor diferença entre duas amostras, A e B , onde o k -ésimo elemento, x_k , do vetor diferença é obtido pela equação 5.1:

$$x_k = \sum_{i=1}^7 \sum_{j=1}^7 (A_{i,j,k} - B_{i,j,k}), k = 1, 2, \dots, 512. \quad (5.1)$$

Desta forma, vetores de características extraídos do conjunto de imagens do mesmo indivíduo são comparados entre si, fornecendo padrões de comparação genuínos. Por outro lado, os vetores de características extraídos do conjunto de imagens de diferentes indivíduos são comparados entre si, fornecendo padrões de comparação de impostores. Um classificador SVM é treinado separadamente para cada abordagem, faciais e regiões oculares, e posteriormente é realizada a etapa de teste.

Figura 20 – Vetor diferença calculado entre duas amostras, A e B.



Fonte: Elaborado pelo autor.

5.2.1 Resultados e Discussões

Conforme esperado, a biometria da face utilizando imagens com oclusões por cachecol na etapa de teste, obteve um pior resultado do que a biometria da face utilizando

Tabela 3 – Resultados experimentais com a base de dados AR Face (MARTINEZ; BENAVENTE, 1998).

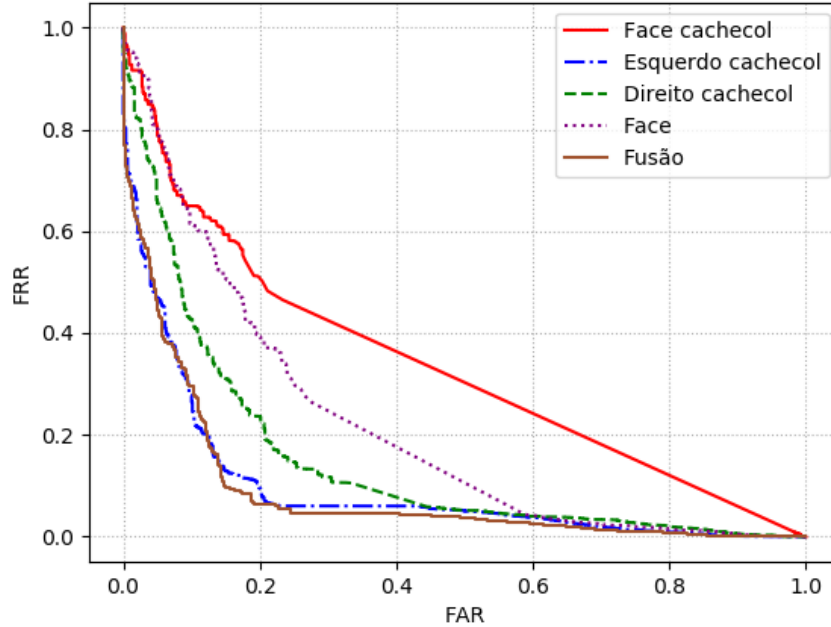
Característica Biométrica	Treinamento	Teste	Equal Error Rate
Face	sem oclusão	sem oclusão	27,06%
Face	sem oclusão	com oclusão	37,73%
Ocular esquerdo	sem oclusão	com oclusão	14,37%
Ocular direito	sem oclusão	com oclusão	20,74%
Fusão	sem oclusão	com oclusão	14,00%

imagens sem oclusões, com um aumento de 10,67% no erro. Entretanto, nas abordagens utilizando as regiões oculares ambas apresentaram uma taxa de EER menor, com a região esquerda sendo 14,37%, e região direita 20,74%. Por fim, a fusão entre as pontuações das abordagens utilizando face e regiões oculares obteve o melhor resultado do experimento, com EER igual a 14,00%. A Tabela 3 mostra os valores de EER obtidos em cada uma das abordagens, e na Figura 21 as curvas ROC geradas, em que a cor vermelha (—) representa a abordagem utilizando imagens faciais com oclusões, azul (— · —) utilizando a região ocular esquerda extraída de imagens faciais com oclusões, verde (— — —) utilizando a região ocular direita de imagens também com oclusões, roxo (· · · · ·) utilizando a região da face com oclusões e marrom (— · — · —) utilizando a fusão entre face e as duas regiões oculares e também sendo a curva mais próxima dos eixos.

Em outros trabalhos usando a base de dados AR Face, a perda de precisão usando imagens com oclusões também foi verificada ao usar imagens de face inteira. Em (WANG; LU; YANG, 2015), foram observadas quedas de até 37,9% na taxa de reconhecimento, quando se utilizam imagens com oclusões por cachecol, em comparação com imagens sem oclusões. Em (XU et al., 2014), uma queda de 7,45% na taxa de reconhecimento foi observada ao usar imagens com oclusões.

Com base nos experimentos, é possível verificar o problema que oclusões parciais do rosto do indivíduo causam em um sistema de autenticação biométrica facial, e que no experimento em questão, as regiões oculares acabaram tendo uma melhor performance do que a face toda, exatamente por não ter sido afetada pela oclusão específica analisada. E a fusão das três características obteve o melhor resultado, mas com pouca diferença para a abordagem utilizando somente a região ocular esquerda.

Figura 21 – Curvas ROC obtidas pelos experimentos realizados na base de dados AR Face ([MARTINEZ; BENAVENTE, 1998](#)).



Fonte: Elaborado pelo autor.

5.3 Experimento 3: FRGC

O conjunto de dados FRGC (Face Recognition Grand Challenge) ([PHILLIPS et al., 2005](#)) foi criado na Universidade de Notre Dame e consiste em milhares de imagens coloridas de alta resolução (1200×1400 pixels), capturadas em ambientes controlados e não controlados. O subconjunto controlado foi capturado em um estúdio sob iluminação uniforme, em que os indivíduos ficavam parados enquanto olhavam diretamente para a câmera e ensaiavam diferentes expressões. Quanto ao conjunto não controlado, as imagens foram obtidas em diferentes cenários, desconsiderando o fundo e a iluminação. A Figura 22 mostra um exemplo de diferentes imagens capturadas de um mesmo indivíduo, à esquerda imagens capturadas em ambiente controlado, e à direita imagens capturas em ambientes não controlados.

O conjunto de dados FRGC consiste em seis diferentes experimentos, com comparações variadas entre indivíduos, em imagens 2D e 3D. Em nosso experimento abordamos o experimento 4 da base de dados FRGC, que consiste em testar a comparação entre uma única imagem de um indivíduo em ambiente controlado, com uma imagem em ambiente

Figura 22 – Diferentes imagens do conjunto de dados FRGC (PHILLIPS et al., 2005) capturadas de um mesmo indivíduo, à esquerda imagens capturadas em ambiente controlado, e a direita imagens capturas em ambientes não controlados..



Fonte: Elaborado pelo autor.

não controlado. Todos os 370 indivíduos do subconjunto de imagens 2D "fall 2013" foram selecionados aleatoriamente para realização do experimento. Doze imagens de cada indivíduo foram selecionadas aleatoriamente para gerar os vetores diferenças que treinaram o classificador SVM.

As imagens selecionadas da base de dados FRGC também foram submetidas ao método MTCNN, e tiveram face e regiões oculares segmentadas. Assim, os experimentos foram conduzidos nas três abordagens diferentes: toda a face, região ocular esquerda e região ocular direita.

Os vetores diferença foram obtidos conforme abordado no experimento com a base de dados AR Face na seção 5.2, com a extração das características antes mesmo da camada totalmente conectada da rede VGG-Face, e com os descritores genuínos gerados por comparações intra-classe e os vetores de diferença impostores gerados pelas comparações inter-classe.

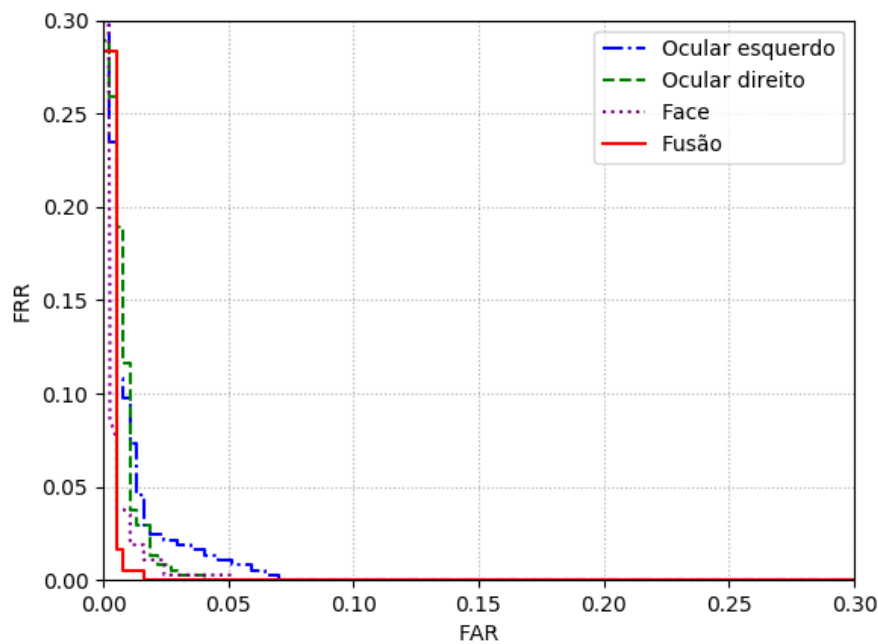
Na etapa de teste, seguindo o protocolo do experimento 4 da base de dados FRGC, uma amostra em ambiente controlado e uma amostra em ambiente não controlado foram selecionadas para gerar os vetores diferenças. E da mesma forma que na etapa de treinamento, comparações intra-classe geraram descritores genuínos, e comparações

inter-classe descritores impostores.

5.3.1 Resultados e discussões

Os descritores gerados na etapa de teste foram submetidos ao classificador SVM, que retornou as probabilidades das comparações serem genuínas ou não, e assim, foram geradas as curvas ROC para cada uma das características biométricas. A Figura 23 apresenta os resultados dos experimentos em termos da curva ROC, onde a cor azul (— · —) representa os resultados obtidos com a região ocular esquerda, verde (— · —) com a região ocular direita, roxo (.....) com toda a face, e vermelho (——) a fusão das três características. O gráfico é apresentado com uma escala onde os eixos tem o limite de 0.30, para facilitar a visualização das diferenças entre cada curva.

Figura 23 – Curvas ROC obtidas pelos experimentos realizados na base de dados FRGC (PHILLIPS et al., 2005).



Fonte: Elaborado pelo autor.

Na Tabela 4 é possível observar os valores de EER obtidos para cada uma das características. Dentre as três características analisadas, usando-se a face toda obteve-se o melhor resultado com uma taxa de EER de 1,62%. Entretanto, a fusão das três características apresentou um melhor resultado do que cada uma delas sozinhas, com um

EER de 0,81%.

Tabela 4 – Resultados experimentais com a base de dados FRGC ([PHILLIPS et al., 2005](#)).

Característica Biométrica	Equal error rate
Face	1,62%
Ocular esquerdo	2,25%
Ocular direito	1,89%
Fusão	0,81%

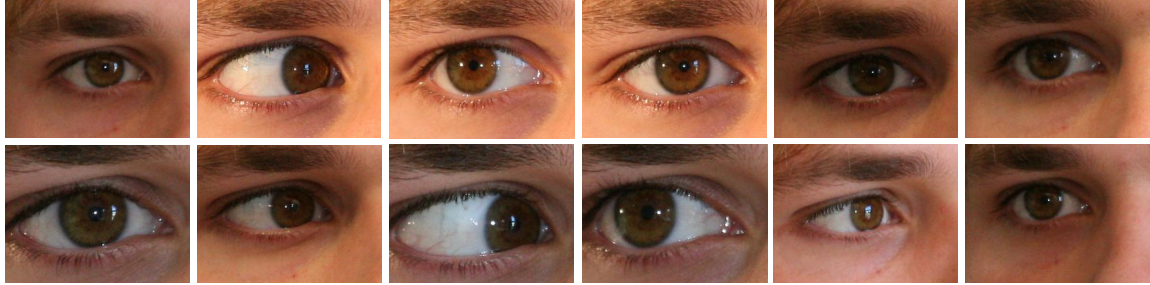
Neste experimento as abordagens utilizando somente as regiões oculares não obtiveram melhores resultados do que a face como um todo, mas a fusão de seus resultados com a face, ajudaram a diminuir o erro que a característica sozinha apresentou. Outros trabalhos também analisaram o experimento 4 da base de dados FRGC, onde as comparações são realizadas entre imagens em ambiente controlado e não controlado. Em ([CHINMAY et al., 2017](#)) os autores propõem um método para extração de características baseado em *Gabor Wavelets*, e um classificador SVM para fazer a identificação facial, atingindo no melhor caso um resultado de 13% de EER. Em ([DING et al., 2015](#)) é proposta uma nova estratégia como descritor facial chamado *Dual-Cross Patterns*, buscando trazer robustez a cenários que hajam variações de expressões faciais e iluminação, atingindo no melhor caso um resultado de 93,39% de taxa de verificação.

5.4 Experimento 4: Ubiris.v2

O conjunto de dados UBIRIS.v2 ([PROENCA et al., 2010](#)) é utilizado principalmente para estudos voltados ao reconhecimento e segmentação da íris, pela fato de conter imagens capturadas com o sensor mais próximo do indivíduo, mas também contém imagens em que o sensor está mais distante, permitindo analisar a região ocular. As imagens foram capturadas no espectro visível de sujeitos em movimento, em ambientes não controlados, em distâncias diferentes e em condições de iluminação desafiadoras, simulando problemas de aquisição realistas com ruídos associados. As 11.102 imagens adquiridas representam um total de 261 sujeitos, de diferentes idades e etnias. A Figura 24 mostra o exemplo de diferentes imagens capturadas de um mesmo indivíduo na base de dados Ubiris.v2.

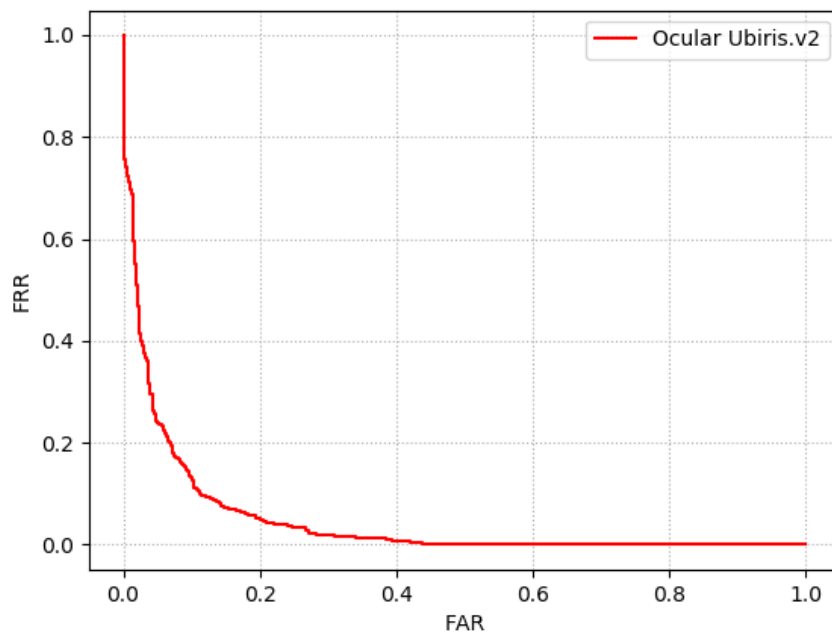
Assim como em ([ZHAO; KUMAR, 2017b](#)), em nosso experimento foram selecionados 171 indivíduos da base de dados Ubiris.v2 aleatoriamente, com 20 imagens de cada, 14 das 20 imagens de cada indivíduo foram selecionadas para gerar os descritores genuínos e

Figura 24 – Diferentes imagens do conjunto de dados Ubiris.v2 (PROENCA et al., 2010) capturadas de um mesmo indivíduo em diferentes condições de iluminação, posicionamento e distância do sensor.



Fonte: Elaborado pelo autor.

Figura 25 – Curva ROC obtida pelo experimento realizado na base de dados Ubiris.v2 (PROENCA et al., 2010).



Fonte: Elaborado pelo autor.

impostores da etapa de treinamento do classificador SVM. As imagens restantes formaram o agrupamento responsável por gerar os descritores da etapa de teste.

As amostras tiveram a extração de características por meio da rede VGG-Face, que foram obtidas antes da camada totalmente conectada da rede, que é responsável pela classificação das amostras, assim como nos experimentos das seção 5.2 e seção 5.3.

5.4.1 Resultados e Discussões

A Figura 25 mostra o resultado do experimento projetado na curva ROC. O valor de EER obtido foi de 10,85%. Este experimento buscou principalmente analisar o desempenho do método proposto neste trabalho em uma base de dados que possui bastante variação na quantidade de informação capturada da região ocular, isto é, com imagens capturadas a diferentes distâncias do sensor, e também com variações de iluminação e posicionamento.

Na pesquisa (RATTANI; DERAKHSHANI, 2017) os autores listam artigos que utilizaram este banco de dados para biometria da região ocular, mas aplicaram diferentes técnicas, com resultados de 24,4% e 31,87% de EER. Em (ZHAO; KUMAR, 2017b) com uma estratégia que incorpora informação semântica explícita para automaticamente recuperar características perioculares abrangente, obteve um EER de 10,98%.

5.5 Experimento 5: FOCS

A base de dados FOCS (*Face and Ocular Challenge Series*) (NIST, 2018) foi utilizada neste experimento. Esta base possui imagens de 437 indivíduos, que foram agrupadas em três partições: *Good*, *Bad* and *Ugly*, formadas por imagens frontais dos indivíduos, capturadas sob condições não controladas de iluminação, tanto em ambientes internos quanto externos. As três partições foram definidas analisando-se os resultados obtidos no FRVT (Face Recognition Vendor Test - NIST) 2006. O conjunto *Good* consiste em imagens de faces para as quais obteve-se taxas de reconhecimento acima da média, o conjunto *Bad* consiste em imagens de faces para as quais obteve-se taxas de reconhecimento médio e o conjunto *Ugly* consiste em imagens de faces para as quais obteve-se taxas de reconhecimento abaixo da média. Nos nossos experimentos foram utilizadas todas as imagens das três partições (*Good*, *Bad* and *Ugly*) em conjunto, ou seja, os pares de imagens de comparação foram selecionados aleatoriamente sem qualquer distinção quanto à expressão facial do indivíduo ou qualidade da imagem. A Figura 26 mostra um exemplo de diferentes imagens da base de dados FOCS capturadas de um mesmo indivíduo.

As imagens do conjunto de dados FOCS foram submetidas ao método de detecção de face MTCNN, e assim, as regiões da face e oculares segmentadas. As imagens contendo apenas as regiões faciais e as imagens contendo apenas as regiões oculares foram submetidas

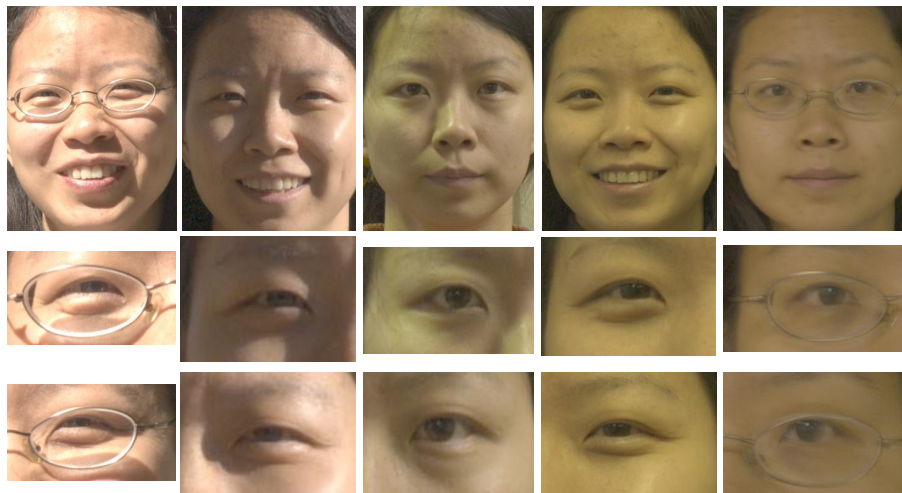
Figura 26 – Diferentes imagens do conjunto de dados FOCS (NIST, 2018) capturadas de um mesmo indivíduo em diferentes condições.



Fonte: Elaborado pelo autor.

a VGG-Face para a obtenção das características profundas e a subsequente obtenção dos vetores de características. A Figura 27 mostra as imagens obtidas após a aplicação do método MTCNN, com a face e regiões oculares segmentadas, nas amostras da Figura 26 respectivamente, em que a fileira de cima representa toda a face, a fileira do meio a região ocular esquerda e a fileira de baixo a região ocular direita.

Figura 27 – Imagens obtidas após a aplicação do método MTCNN em algumas amostras do conjunto de dados FOCS (NIST, 2018) capturadas de um mesmo indivíduo em diferentes condições.



Fonte: Elaborado pelo autor.

Os conjuntos de vetores diferença das classes genuíno e impostor foram obtidos, respectivamente, por meio da subtração de pares de vetores de características profundas oriundos de diferentes imagens do mesmo indivíduo, e por meio da subtração de pares de vetores de características profundas oriundos de diferentes indivíduos, conforme conduzido no experimento 2 deste trabalho na seção 5.2. Os conjuntos dos vetores das classes genuíno

e impostor foram divididos em subconjuntos de treinamento (50% dos vetores) e teste (50% dos vetores).

O treinamento supervisionado do classificador SVM foi promovido por meio da submissão dos vetores (amostras) das classes genuíno e impostor do conjunto de treinamento, indicando-se os rótulos das classes de origem das amostras.

Após o treinamento, o classificador foi utilizado na etapa de teste para classificar os vetores do conjunto de teste. A fase de teste se deu basicamente pela apresentação de cada uma das amostras do conjunto de teste aos classificadores, que ficaram especializados em fornecer as probabilidades (pontuações) de cada amostra pertencer à classe genuíno. Além das três características biométricas analisadas, também foi conduzida a fusão entre as três.

5.5.1 Resultados e Discussões

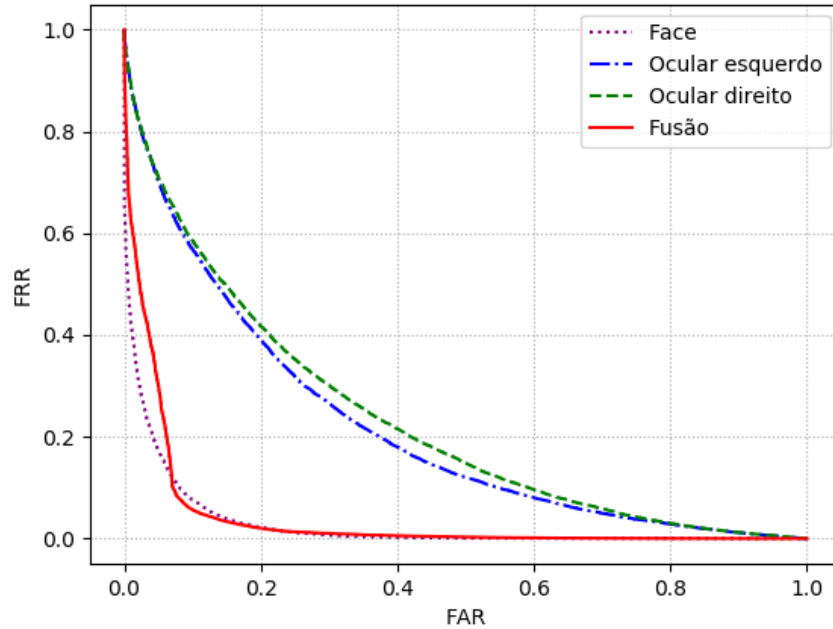
A Figura 28 mostra a curva ROC obtida com a aplicação no método proposto nas imagens da base de dados FOCS, para cada característica biométrica e cada abordagem (individual ou fundida). A cor azul (— · —) representa os resultados obtidos com a região ocular esquerda, verde (— — —) com a região ocular direita, roxo (— · · · —) com toda a face, e vermelho (——) a fusão das três características. A Tabela 5 mostra os valores de EER obtidos para cada uma das abordagens do experimento.

Tabela 5 – Resultados experimentais com a base de dados FOCS (NIST, 2018).

Característica Biométrica	Equal Error Rate
Face	8,90%
Ocular esquerdo	30,00%
Ocular direito	28,20%
Fusão	7,95%

Entre as abordagens utilizando somente a região ocular a que obteve melhor resultado foi da região ocular esquerda com $EER = 28,20\%$, e no geral, a abordagem utilizando a fusão entre toda a face com as regiões oculares com $EER = 7,95\%$, superando o resultado de $EER = 8,90\%$ obtido pela abordagem utilizando somente a face. Os resultados obtidos com os experimentos mostram um bom desempenho do método proposto, levando em consideração que toda a base de dados FOCS foi utilizada, sem separação entre os

Figura 28 – Curvas ROC obtidas pelo experimento realizado na base de dados FOCS (NIST, 2018).



Fonte: Elaborado pelo autor.

conjuntos *Good*, *Bad* e *Ugly*, que segundo (NIST, 2018) obtiveram acurácias de 98%, 80% e 15%, respectivamente.

6 Conclusões e Trabalhos Futuros

6.1 Conclusões

Levando-se em consideração os avanços tecnológicos e a busca por métodos de identificação rápidos e eficazes, a biometria se torna uma alternativa muito atrativa no cenário atual. A possibilidade de implementação de sistemas biométricos para cenários em que se deseja uma abordagem menos cooperativa do usuário, ou que os dados obtidos possam estar em condições não ideais, impulsiona a busca por métodos de reconhecimento menos sensíveis a ruídos originados no momento da captura de dados.

Os métodos tradicionais de biometria, levando em consideração características biométricas consolidadas, como reconhecimento da íris ou face, podem não apresentar resultados tão bons para estas informações capturadas em ambientes com pouco controle. Diante deste cenário, para contornar esses problemas, o uso de características biométricas alternativas e a fusão de características biométricas surgem como boas opções para compor sistemas multimodais, visando uma melhor forma de reconhecer indivíduos.

Em se tratando da biometria da região ocular, estudos tem concluído que a região periocular em união das informações oculares pode aumentar o poder de reconhecimento de um sistema biométrico. A utilização dessas características é uma boa alternativa para aumentar as informações referentes aos indivíduos em uma base de dados.

As estratégias de aprendizado de máquina em profundidade vem sendo estudadas nos últimos anos para o desenvolvimento de sistemas biométricos. A biometria da região ocular se beneficia nesse aspecto pela possibilidade de a partir de um mesmo sensor se obter informações da íris simultaneamente, facilitando com que tais características tenham seus padrões analisados pela mesma entrada de uma única rede de profundidade.

Os experimentos realizados mostraram que o método proposto para estabelecer o reconhecimento da região ocular para a identificação biométrica de pessoas pode ser uma alternativa interessante no processo de autenticação de indivíduos, tanto pelo uso individual da região ocular, como também da sua utilização em conjunto com a face, criando um sistema mais robusto para situações em que as imagens capturadas dos indivíduos possuam

ruídos, obstruções parciais ou diferentes poses e expressões.

O uso das características extraídas de redes de profundidades se mostraram muito eficientes para compor os descritores de cada indivíduo por meio da diferença entre amostras genuínas e impostoras, possibilitando o treinamento de uma SVM como classificador do sistema. Nos experimentos realizados a rede de aprendizado em profundidade VGG-Face obteve o melhor resultado, com a menor taxa de erro.

Sendo assim, é possível concluir que o atual cenário tecnológico, em conjunto com diferentes métodos de extração de características e estratégias utilizando aprendizado de máquina em profundidade, tornam relevante o estudo de abordagens que busquem alternativas para os métodos biométricos tradicionais, visando a utilização de diferentes características de um mesmo indivíduo para promover a identificação, tornando os sistemas biométricos mais robustos.

6.2 Trabalhos Futuros

Como possíveis trabalhos futuros, pode-se apontar:

- Realizar uma análise mais detalhada das características extraídas das redes de aprendizado em profundidade, podendo aplicar técnicas que determinam as melhores características para descrever o problema, resultando também em redução de dimensionalidade dos descritores;
- Analisar outras opções para a etapa de classificação das amostras, em alternativa ao classificador SVM, verificando diferentes maneiras de tomar a decisão quanto à autenticidade dos indivíduos;
- Trabalhar em abordagens que consigam utilizar melhor as camadas totalmente conectadas das redes de profundidade, seja para classificar as amostras, ou para possibilitar o retreinamento da rede antes da etapa de extração de características;
- Implementar etapas de pré-processamento nas imagens antes de submetê-las às redes para extração de características, podendo remover possíveis ruídos, ou identificar pontos-chaves para a identificação.

Referências

- ABATE, A. F. et al. 2d and 3d face recognition: A survey. *Pattern recognition letters*, Elsevier, v. 28, n. 14, p. 1885–1906, 2007.
- AHUJA, K. et al. Convolutional neural networks for ocular smartphone-based biometrics. *Pattern Recognition Letters*, Elsevier, v. 91, p. 17–26, 2017.
- ALLEN, H. L. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, Nature Research, v. 467, n. 7317, p. 832–838, 2010.
- ALONSO-FERNANDEZ, F.; BIGUN, J. Periocular biometrics: databases, algorithms and directions. In: *2016 4th International Conference on Biometrics and Forensics (IWBF)*. [S.l.: s.n.], 2016. p. 1–6.
- BENGIO, Y. Deep learning of representations: Looking forward. In: SPRINGER. *International Conference on Statistical Language and Speech Processing*. [S.l.], 2013. p. 1–37.
- BERGSTRA, J.; YAMINS, D.; COX, D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: *Anais da International Conference on Machine Learning*. Estados Unidos: JMLR, 2013.
- BOBELDYK, D.; ROSS, A. Iris or periocular? exploring sex prediction from near infrared ocular images. In: IEEE. *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*. [S.l.], 2016. p. 1–7.
- BOLLE, R. M. et al. *Guide to biometrics*. [S.l.]: Springer Science & Business Media, 2013.
- BOLLE, R. M. et al. Iris individuality: A partial iris model. In: IEEE. *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. [S.l.], 2004. v. 2, p. 927–930.
- BUCIU, I.; GACSADI, A. Biometrics systems and technologies: a survey. *International Journal of Computers Communications & Control*, v. 11, n. 3, p. 315–330, 2016.
- CHELLAPILLA, K.; PURI, S.; SIMARD, P. High performance convolutional neural networks for document processing. In: *Anais da International Workshop on Frontiers in Handwriting Recognition*. França: CCSD, 2006.
- CHINMAY, B. et al. *Face Identification*. 2017. IJESC 2017, Research article volume 7 No 5, disponivel em: <<http://ijesc.org/upload/5dd719eb3ad5ea1c0b8087dc22ce5eab.Face%20Identification.pdf>> acesso em: 31/05/2019.
- CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, p. 1610–02357, 2017.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.], 2005. v. 1, p. 886–893.

- DAUGMAN, J. How iris recognition works. *IEEE Transactions on circuits and systems for video technology*, IEEE, v. 14, n. 1, p. 21–30, 2004.
- DAUGMAN, J. G. *Biometric personal identification system based on iris analysis*. [S.l.]: Google Patents, 1994. US Patent 5,291,560.
- DENG, L.; DONG, Y. Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, v. 7, n. 3–4, p. 197–387, 2014.
- DENG, L.; YU, D. et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 7, n. 3–4, p. 197–387, 2014.
- DING, C. et al. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 38, n. 3, p. 518–531, 2015.
- GIRSHICK, R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2014. (CVPR '14), p. 580–587. ISBN 978-1-4799-5118-5. Disponível em: <http://dx.doi.org/10.1109/CVPR.2014.81>.
- GUENTHER, N.; SCHONLAU, M. et al. Support vector machines. *Stata J*, v. 16, n. 4, p. 917–937, 2016.
- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- HERNANDEZ-DIAZ, K.; ALONSO-FERNANDEZ, F.; BIGUN, J. Periocular recognition using cnn features off-the-shelf. *arXiv preprint arXiv:1809.06157*, 2018.
- HOLLINGSWORTH, K. P.; BOWYER, K. W.; FLYNN, P. J. The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 31, n. 6, p. 964–973, 2009.
- JAIN, A. K.; KUMAR, A. Biometric recognition: an overview. In: *Second generation biometrics: The ethical, legal and social context*. [S.l.]: Springer, 2012. p. 49–79.
- JAIN, A. K.; NANDAKUMAR, K.; ROSS, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, Elsevier, v. 79, p. 80–105, 2016.
- JAIN, A. K.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, IEEE, v. 14, n. 1, p. 4–20, 2004.
- JAIN, A. K.; ROSS, A. A.; NANDAKUMAR, K. Introduction. In: *Introduction to Biometrics*. [S.l.]: Springer, 2011. p. 1–49.
- JUEFEI-XU, F.; SAVVIDES, M. Subspace-based discrete transform encoded local binary patterns representations for robust periocular matching on nist's face recognition grand challenge. *IEEE transactions on image processing*, IEEE, v. 23, n. 8, p. 3490–3505, 2014.

- KANNALA, J.; RAHTU, E. Bsif: Binarized statistical image features. In: IEEE. *Pattern Recognition (ICPR), 2012 21st International Conference on*. [S.l.], 2012. p. 1363–1366.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Anais da IEEE*, v. 86, n. 11, p. 2278–2324, 1998.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, n. 2, p. 91–110, 2004.
- LYLE, J. R. et al. Soft biometric classification using periocular region features. In: IEEE. *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. [S.l.], 2010. p. 1–7.
- MAGUIRE, M. The birth of biometric security. *Anthropology Today*, Wiley Online Library, v. 25, n. 2, p. 9–14, 2009.
- MARTINEZ, A.; BENAVENTE, R. The ar face database. *CVC Technical Report #24*, 1998.
- MASEK, L. et al. Recognition of human iris patterns for biometric identification. Master's thesis, University of Western Australia, 2003.
- MENOTTI, D. et al. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, v. 10, n. 4, p. 864–879, 2015.
- MERAOUIMIA, A. et al. Improving biometric identification performance using pcanet deep learning and multispectral palmprint. In: *Biometric Security and Privacy*. [S.l.]: Springer, 2017. p. 51–69.
- MINAEE, S.; ABDOLRASHIDIY, A.; WANG, Y. An experimental study of deep convolutional features for iris recognition. In: *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. [S.l.: s.n.], 2016. p. 1–6.
- MORAES, J. *Controle de Acesso Baseado em Biometria Facial*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2010.
- NEGIN, M. et al. An iris biometric system for public and personal use. *Computer*, IEEE, v. 33, n. 2, p. 70–75, 2000.
- NIE, L.; KUMAR, A.; ZHAN, S. Periocular recognition using unsupervised convolutional rbm feature learning. In: *2014 22nd International Conference on Pattern Recognition*. [S.l.: s.n.], 2014. p. 399–404. ISSN 1051-4651.
- NIGAM, I.; VATSA, M.; SINGH, R. Ocular biometrics: A survey of modalities and fusion approaches. *Information Fusion*, Elsevier, v. 26, p. 1–35, 2015.
- NIST. *NIST Face and Ocular Challenge Series database*. 2018. Disponível em: <<http://www.nist.gov/itl/iad/ig/focs.cfm>>. Acesso em: 29 june 2018.
- ORTIZ, E.; BOWYER, K. W.; FLYNN, P. J. An optimal strategy for dilation based iris image enrollment. In: IEEE. *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. [S.l.], 2014. p. 1–6.

- PADOLE, C. N.; PROENCA, H. Periocular recognition: Analysis of performance degradation factors. In: *2012 5th IAPR International Conference on Biometrics (ICB)*. [S.l.: s.n.], 2012. p. 439–445. ISSN 2376-4201.
- PANCHAL, T.; SINGH, A. Multimodal biometric system. *International Journal of Advanced Research in Computer Science and Software Engineering*, v. 3, n. 5, 2013.
- PANKANTI, S.; BOLLE, R. M.; JAIN, A. Biometrics: The future of identification [guest editors' introduction]. *Computer, IEEE*, v. 33, n. 2, p. 46–49, 2000.
- PARK, U. et al. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, v. 6, n. 1, p. 96–106, March 2011. ISSN 1556-6013.
- PARK, U.; ROSS, A.; JAIN, A. K. Periocular biometrics in the visible spectrum: A feasibility study. In: IEEE. *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. [S.l.], 2009. p. 1–6.
- PARKHI, O. M.; VEDALDI, A.; ZISSERMAN, A. Deep face recognition. In: *British Machine Vision Conference*. [S.l.: s.n.], 2015.
- PHILLIPS, P. J. et al. Overview of the face recognition grand challenge. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. [S.l.: s.n.], 2005. v. 1, p. 947–954 vol. 1. ISSN 1063-6919.
- PIETIKÄINEN, M. Image analysis with local binary patterns. *Image Analysis*, Springer, p. 6–9, 2005.
- PINTO, N. et al. A high-throughput screening approach to discovering good forms of biologically-inspired visual representation. *PLoS Computational Biology*, v. 5, n. 11, p. e1000579, 2009.
- PROENCA, H. et al. The ubiris.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 8, p. 1529–1535, Aug 2010. ISSN 0162-8828.
- RATHGEB, C.; UHL, A.; WILD, P. On combining selective best bits of iris-codes. In: SPRINGER. *BioID*. [S.l.], 2011. p. 227–237.
- RATTANI, A.; DERAKHSHANI, R. Ocular biometrics in the visible spectrum: A survey. *Image and Vision Computing*, Elsevier, v. 59, p. 1–16, 2017.
- RATTANI, A. et al. Icip 2016 competition on mobile ocular biometric recognition. In: IEEE. *2016 IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2016. p. 320–324.
- REDDY, N.; RATTANI, A.; DERAKHSHANI, R. Ocularnet: Deep patch-based ocular biometric recognition. In: IEEE. *2018 IEEE International Symposium on Technologies for Homeland Security (HST)*. [S.l.], 2018. p. 1–6.
- ROSS, A.; JAIN, A. Information fusion in biometrics. *Pattern recognition letters*, Elsevier, v. 24, n. 13, p. 2115–2125, 2003.

- SANTOS, G.; PROENÇA, H. Periocular biometrics: An emerging technology for unconstrained scenarios. In: IEEE. *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2013 IEEE Workshop on*. [S.l.], 2013. p. 14–21.
- SIM, H. M. et al. Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images. *Expert Systems with Applications*, Elsevier, v. 41, n. 11, p. 5390–5404, 2014.
- SIMARD, P.; STEINKRAUS, D.; PLATT, J. Best practices for convolutional neural networks applied to visual document analysis. In: *Anais da International Conference on Document Analysis and Recognition*. Estados Unidos: IEEE, 2003. p. 958–963.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- SUN, X.; WU, P.; HOI, S. C. Face detection using deep learning: An improved faster rcnn approach. *arXiv preprint arXiv:1701.08289*, 2017.
- SUN, Y.; WANG, X.; TANG, X. Deep learning face representation from predicting 10,000 classes. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2014. (CVPR '14), p. 1891–1898. ISBN 978-1-4799-5118-5. Disponível em: <http://dx.doi.org/10.1109/CVPR.2014.244>.
- SZEGEDY, C. et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9. ISSN 1063-6919.
- TAIGMAN, Y. et al. Deepface: Closing the gap to human-level performance in face verification. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2014. p. 1701–1708. ISSN 1063-6919.
- TAN, T.; HE, Z.; SUN, Z. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and vision computing*, Elsevier, v. 28, n. 2, p. 223–230, 2010.
- VIOLA, P.; JONES, M. J. Robust real-time face detection. *International journal of computer vision*, Springer, v. 57, n. 2, p. 137–154, 2004.
- WANG, D.; LU, H.; YANG, M.-H. Kernel collaborative face recognition. *Pattern Recognition*, Elsevier, v. 48, n. 10, p. 3025–3037, 2015.
- WANG, W. et al. Face recognition based on deep learning. In: SPRINGER. *International Conference on Human Centered Computing*. [S.l.], 2014. p. 812–820.
- WOODARD, D. L. et al. On the fusion of periocular and iris biometrics in non-ideal imagery. In: IEEE. *Pattern Recognition (ICPR), 2010 20th International Conference on*. [S.l.], 2010. p. 201–204.
- XU, Y. et al. Data uncertainty in face recognition. *IEEE transactions on cybernetics*, IEEE, v. 44, n. 10, p. 1950–1961, 2014.

- ZHANG, K. et al. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, IEEE, v. 23, n. 10, p. 1499–1503, 2016.
- ZHAO, Z.; KUMAR, A. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Transactions on Information Forensics and Security*, v. 12, n. 5, p. 1017–1030, May 2017. ISSN 1556-6013.
- ZHAO, Z.; KUMAR, A. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 12, n. 5, p. 1017–1030, 2017.
- ZHAO, Z.; KUMAR, A. Improving periocular recognition by explicit attention to critical regions in deep neural network. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 13, n. 12, p. 2937–2952, 2018.
- ZHENG, S. et al. A cascade fusion scheme for gait and cumulative foot pressure image recognition. *Pattern Recognition*, Elsevier, v. 45, n. 10, p. 3603–3610, 2012.