

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Artur Azeredo Santos Servian

**Automação Inteligente de Processos para evitar
fraudes analógicas, digitais e os *DeepFakes* das
LLMs Generativas**

Uberlândia, Brasil

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Artur Azeredo Santos Servian

**Automação Inteligente de Processos para evitar fraudes
analógicas, digitais e os *DeepFakes* das LLMs Generativas**

Trabalho de conclusão de curso apresentado
à Faculdade de Engenharia Mecânica da Uni-
versidade Federal de Uberlândia, como parte
dos requisitos exigidos para a obtenção do
título de Bacharel em Engenharia Mecatrô-
nica.

Orientador: Prof. Dr. Luciano Vieira Lima

Universidade Federal de Uberlândia – UFU

Faculdade de Engenharia Mecânica

Bacharelado em Engenharia Mecatrônica

Uberlândia, Brasil

2025

Artur Azeredo Santos Servian

Automação Inteligente de Processos para evitar fraudes analógicas, digitais e os *DeepFakes* das LLMs Generativas

Trabalho de conclusão de curso apresentado à Faculdade de Engenharia Mecânica da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção do título de Bacharel em Engenharia Mecatrônica.

Trabalho aprovado. Uberlândia, Brasil, 05 de junho de 2025:

Prof. Dr. Luciano Vieira Lima
Orientador

Prof. Dr. José dos Reis Vieira de Moura Júnior
Membro da Banca

Me. Thales Oliveira Lima
Membro da Banca

Uberlândia, Brasil
2025

Agradecimentos

Este Projeto de Fim de Curso encerra uma longa e árdua etapa da Graduação em Engenharia Mecatrônica na Universidade Federal de Uberlândia. Um sonho idealizado ainda na infância, que se mostrou desafiador em todos os seus diferentes momentos e aspectos. Pude acompanhar a evolução pessoal e profissional, enquanto o progresso acadêmico se desenvolvia. A Universidade, um verdadeiro universo de conhecimento e saberes, e os mais valiosos dentre estes foram as experiências vividas (mesmo as mais negativas), ter conhecido pessoas incríveis e ter firmado laços imprescindíveis.

Agradeço à minha amada mãe, por ter me apoiado desde os estudos pré-vestibulares até os momentos em que desistir se tornou uma opção. Agradeço aos meus avós por terem dado suporte incondicional à conclusão desta graduação.

Ao meu amigo-irmão Leonardo (*in memorian*), por ter mantido vivo este sonho mútuo.

A Jônatas, Gabriel e Esther, com quem dividi o teto, o pão, as dores e a rotina, tornando esta vivência mais leve e alegre.

A Lucas e Jalles, por mostrarem que, mesmo em uma graduação em Ciências Exatas, a expressão da arte alimenta a alma e move a humanidade.

Aos amigos Paulo Victor e Juliana, por trazerem o acolhimento de nossa cidade de origem para estas paisagens desconhecidas.

Ao professor Juliano Miranda, que primeiro me fez brilhar os olhos pela Robótica, e ao professor Carlos Roberto Lopes, que me permitiu manter esta paixão acesa durante grande parte da graduação.

Este não é o fim. Este não é o começo.

Resumo

A crescente digitalização de serviços e a sofisticação de fraudes digitais, incluindo ataques por apresentação em sistemas de autenticação biométrica facial, demandam soluções de segurança cada vez mais robustas. Este Projeto de Fim de Curso teve como objetivo principal o desenvolvimento e a avaliação de um protótipo de sistema de detecção de vivacidade (*liveness detection*) multifacetado, visando aumentar a segurança contra tais ataques. A metodologia empregou uma combinação de técnicas de visão computacional, utilizando principalmente as bibliotecas Python, OpenCV e MediaPipe para a extração de pontos fiduciais faciais e análise subsequente. O protótipo integrou módulos para validação da pose cefálica como pré-requisito, detecção de piscadas com base no *Eye Aspect Ratio* (EAR), um método de indução implícita do olhar através da interação com alvos visuais clicáveis, análise da reação pupilar à emissão de luz (reflexo fotomotor) e verificação da cor da esclera. Todas as funcionalidades foram orquestradas por um algoritmo piloto que também ofereceu interfaces distintas para o usuário e para o desenvolvedor, culminando na geração de um escore de vivacidade. Os testes experimentais com usuários legítimos demonstraram uma Taxa de Aceitação Verdadeira (TAR) de 90% em testes com webcam e 100% utilizando câmera de smartphone via DroidCam. Em cenários de ataque, o sistema obteve 100% de Taxa de Rejeição Verdadeira (TRR) contra imagens estáticas apresentadas em tela e 90% de TRR contra vídeos *deepfake* gerados pela plataforma Kling AI. Conclui-se que o protótipo desenvolvido apresenta uma base promissora para a detecção de vivacidade, com alta eficácia contra ataques básicos e boa resiliência a *deepfakes*, embora tenham sido identificadas oportunidades de aprimoramento, especialmente na robustez de sub-testes específicos e na necessidade de validação com um espaço amostral mais amplo. As principais contribuições residem na integração de múltiplos indicadores de vivacidade em um sistema funcional e na análise de seu desempenho inicial.

Palavras-chave: Antifraude, *Deepfakes*, Segurança, Detecção de Vivacidade, Biometria Facial.

Lista de ilustrações

Figura 1 – Etapas de visualização de tela durante <i>liveness</i> desenvolvido pela Amazon.	18
Figura 2 – Visão Geral do método proposto em (JIN; YE; CHEN, 2021)	28
Figura 3 – <i>Landmarks</i> faciais utilizados pelo Mediapipe	34
Figura 4 – Fluxograma geral do processo de verificação de vivacidade.	42
Figura 5 – Fluxo simplificado da captura e preparação inicial de frames de vídeo.	46
Figura 6 – Exemplos de cenários na etapa de detecção facial: (a) detecção bem-sucedida de face única e (b) detecção de múltiplas faces.	47
Figura 7 – Exemplos de <i>feedback</i> em tempo real ao usuário: (a) detecção de angulação vertical do rosto e (b) detecção de angulação horizontal do rosto.	49
Figura 8 – Exemplo da malha de 468 pontos fiduciais faciais (Face Mesh) detectada pela biblioteca MediaPipe sobre uma face humana.	51
Figura 9 – Ilustração dos eixos de rotação da cabeça: Guinada (ψ , Yaw), Inclinação (θ , Pitch) e Rotação Axial (ϕ , Roll).	52
Figura 10 – Exemplo de localização da íris utilizando landmarks faciais para delimitar a ROI ocular e algoritmos de processamento de imagem para refinar a detecção do centro pupilar.	56
Figura 11 – Localização dos seis landmarks faciais (P_1 a P_6) utilizados para o cálculo do Eye Aspect Ratio (EAR) para um olho.	57
Figura 12 – Detecção de piscada dos olhos direito e esquerdo.	57
Figura 13 – Variação típica do valor do Eye Aspect Ratio (EAR) ao longo do tempo durante uma sequência de piscadas. A queda abaixo do limiar indica o fechamento do olho.	58
Figura 14 – Ilustração esquemática da inferência da direção do olhar (esquerda e direita) baseada na posição relativa da pupila dentro da Região de Interesse (ROI) do olho.	59
Figura 15 – Sequência de apresentação de alvos visuais (círculos) em diferentes posições da tela para indução do movimento ocular e interação por toque/clique.	62
Figura 16 – Análise comparativa da aparência e reflexibilidade da esclera ocular quando iluminada por fontes de luz de diferentes cores (verde, azul e vermelha), visando identificar padrões que auxiliem na distinção entre um olho real e uma representação artificial.	64

Figura 17 – Sequência ilustrativa da verificação do reflexo fotomotor: (a) diâmetro pupilar basal, (b) aplicação de estímulo luminoso pela tela, (c) constrição pupilar observada.	66
Figura 18 – Exemplo da "Visão do Desenvolvedor" e "Visão do Usuário".	68
Figura 19 – Processo de agregação de evidências dos diferentes módulos de análise para a formação do escore básico de vivacidade e a subsequente tomada de decisão.	69
Figura 20 – Ilustração da configuração do teste de ataque por apresentação utilizando uma imagem estática exibida na tela de um smartphone e apresentada à webcam do sistema.	76
Figura 21 – Exemplos de frames dos vídeos deepfake gerados pela plataforma Kling AI e utilizados nos testes de ataque por apresentação.	77

Lista de tabelas

Tabela 1 – Principais tecnologias e bibliotecas utilizadas no desenvolvimento do protótipo.	45
Tabela 2 – Exemplo de faixas angulares para classificação da orientação da cabeça.	55
Tabela 3 – Exemplo de classificação da direção do olhar baseada na posição normalizada da pupila (X_{norm} , Y_{norm}) dentro da ROI ocular.	59
Tabela 4 – Desafios baseados em ações oculares e critérios de validação.	62
Tabela 5 – Exemplo ilustrativo de contribuição de diferentes verificações para o escore de vivacidade.	70
Tabela 6 – Resultados consolidados dos testes de prova de vida com usuários legítimos.	75
Tabela 7 – Resultados consolidados dos testes de ataques por apresentação (spoofing).	78

Sumário

1	INTRODUÇÃO	11
1.1	Objetivos	11
1.1.1	Objetivo Geral	12
1.1.2	Objetivos Específicos	12
1.2	Organização do Trabalho	13
2	REVISÃO BIBLIOGRÁFICA	14
2.1	Demanda de Antifraude	14
2.2	O Mercado de Antifraude	16
2.2.1	FaceTec	16
2.2.2	iProov	17
2.2.3	Amazon Rekognition Face Liveness	18
2.3	Deteção de Manipulação de Imagens e Vídeos	19
2.3.1	Técnicas Tradicionais e Baseadas em Artefatos	19
2.3.1.1	Análise de Ruído e Inconsistências Locais	20
2.3.1.2	Análise de Compressão JPEG	20
2.3.1.3	Outras Características Forenses	21
2.3.2	Tecnologia por Trás dos Deepfakes	22
2.3.2.1	Autoencoders	23
2.3.2.2	Redes Adversariais Generativas (GANs)	24
2.3.3	Abordagens Baseadas em Deep Learning para Deteção de Manipulação Facial e Geral	25
2.3.3.1	Redes Neurais Convolucionais (CNNs) para Extração de Características e Deteção	25
2.3.3.2	Deteção Baseada em Sinais Biológicos e Fisiológicos	26
2.3.3.3	Mecanismos de Atenção para Localização de Regiões Manipuladas	29
2.4	Análise de Documentos e Identidade (Documentoscopia e Biometria Facial)	31
2.4.1	Verificação de Autenticidade de Documentos	31
2.4.2	Biometria Facial para Verificação de Identidade	32
2.4.2.1	Reconstrução 3D da Face a Partir de Imagens 2D	32
2.4.2.2	Uso de Landmarks Faciais e Aprendizado Profundo para Análise Facial	33
2.4.2.3	Deteção de Vivacidade (<i>Liveness Detection</i>)	34
2.5	Desafios e Limitações Atuais na Deteção de Fraudes Inteligentes	35
2.5.1	Evolução Constante das Técnicas de Manipulação	35
2.5.2	Generalização de Modelos para Tipos de Fraude Não Vistos	36

2.5.3	Robustez Contra Ataques Adversariais	37
2.5.4	Interpretabilidade dos Modelos de Deep Learning	37
2.5.5	Questões Éticas, Legais e de Privacidade	38
3	METODOLOGIA DE DESENVOLVIMENTO DA SOLUÇÃO DE PROVA DE VIDA (<i>LIVENESS DETECTION</i>)	40
3.1	Introdução à Arquitetura do Sistema	40
3.1.1	Visão Geral do Protótipo e Seus Módulos Componentes	41
3.1.2	Fluxograma Geral do Processo de Liveness Detection	42
3.1.3	Principais Tecnologias e Bibliotecas Empregadas	43
3.2	Aquisição e Pré-processamento de Imagens Faciais	45
3.2.1	Captura e Seleção de Frames de Vídeo	45
3.2.2	Deteção e Validação da Presença de Face Única	46
3.2.3	Considerações sobre Qualidade da Imagem para Análise Robusta	47
3.3	Análise Facial Estática e Dinâmica	49
3.3.1	Deteção Facial de Precisão para Mapeamento	49
3.3.2	Extração e Mapeamento de Pontos Fiduciais da Face	50
3.4	Estimativa e Monitoramento da Posição da Cabeça	52
3.4.1	Determinação dos Ângulos de Euler (Yaw, Pitch, Roll)	52
3.4.2	Classificação da Orientação da Cabeça para Interação com o Usuário	54
3.5	Análise Ocular Detalhada para Prova de Vida	55
3.5.1	Deteção e Rastreamento da Íris	55
3.5.2	Deteção de Piscada de Olhos	56
3.5.3	Análise de Direção do Olhar (Gaze Estimation)	58
3.6	Implementação de Interações e Verificações de Prova de Vida	60
3.6.1	Validação Contínua da Posição da Cabeça como Pré-requisito	60
3.6.2	Interações Oculares para Confirmação de Vivacidade	61
3.6.2.1	Verificação de Piscadas Naturais	61
3.6.2.2	Direcionamento Implícito do Olhar através de Interação com Alvos	61
3.7	Mecanismos Adicionais de Antispoofing	62
3.7.1	Análise da Cor da Esclera	63
3.7.2	Análise da Reação Pupilar à Luz (Reflexo Fotomotor)	64
3.7.3	Outras Abordagens e Heurísticas de Deteção de Fraude	65
3.8	Integração dos Módulos e Tomada de Decisão	67
3.8.1	Orquestração das Diferentes Etapas de Verificação pelo Algoritmo Piloto	67
3.8.2	Lógica de Decisão Final para Aprovação/Rejeição da Vivacidade	68
4	RESULTADOS	71
4.1	Configuração Experimental e Métricas de Avaliação	71
4.1.1	Ambiente e Materiais de Teste	71

4.1.2	Métricas de Avaliação de Desempenho	72
4.2	Resultados dos Testes de Uso Legítimo	73
4.2.1	Desempenho com Webcam Dedicada/Integrada	73
4.2.2	Desempenho com Câmera de Smartphone (via DroidCam)	74
4.2.3	Sumário dos Resultados de Uso Legítimo	75
4.3	Resultados dos Testes de Ataques por Apresentação (Spoofing) . .	75
4.3.1	Resistência a Ataques com Imagem Estática	75
4.3.2	Resistência a Ataques com Vídeos Deepfake	76
4.3.3	Sumário dos Resultados de Testes de Ataque	78
4.4	Análise e Discussão Consolidada dos Resultados	78
4.4.1	Interpretação Geral do Desempenho	78
4.4.2	Pontos Fortes do Protótipo	79
4.4.3	Limitações Identificadas e Análise de Falhas	79
4.4.4	Considerações Finais sobre os Resultados	80
5	CONCLUSÃO	81
5.1	Principais Contribuições do Trabalho	81
5.2	Trabalhos Futuros e Recomendações	82
5.2.1	Desenvolvimento de uma Interface de Usuário Comercial e Amigável	82
5.2.2	Validação Experimental Extensiva e Robusta com Amostragem Significativa	83
5.2.3	Refinamento e Aprofundamento dos Algoritmos de Análise Ocular Antifraude	84
5.2.4	Outras Direções e Recomendações	85
	REFERÊNCIAS	86

1 Introdução

A acelerada transformação digital presenciada nas últimas décadas impulsionou uma miríade de serviços e interações para o ambiente online, desde transações financeiras e acesso a serviços governamentais até comunicação e entretenimento. Essa migração, embora traga consigo inúmeras conveniências e eficiências, também expõe usuários e organizações a uma crescente gama de ameaças cibernéticas e fraudes digitais. Nesse cenário, a autenticação segura da identidade dos usuários tornou-se um pilar fundamental para a confiabilidade e integridade dos sistemas digitais (FEBRABAN; DELOITTE, 2024).

A biometria, que utiliza características fisiológicas ou comportamentais únicas para identificar indivíduos, emergiu como uma das abordagens mais promissoras e amplamente adotadas para autenticação. Dentre as diversas modalidades biométricas, o reconhecimento facial se destaca pela sua conveniência e aceitação. Contudo, a simplicidade aparente da captura facial também a torna vulnerável a uma variedade de ataques de apresentação (do inglês, *Presentation Attacks* - PAs), onde artefatos como fotografias, vídeos, máscaras tridimensionais ou mesmo *deepfakes* são utilizados na tentativa de personificar um usuário legítimo e ludibriar o sistema biométrico (GALBALLY; MARCEL; FIERREZ, 2014).

Para mitigar essas vulnerabilidades, as técnicas de Detecção de Vivacidade (do inglês, *Liveness Detection*) ou Prova de Vida são indispensáveis. Tais técnicas buscam verificar se os traços biométricos apresentados ao sensor pertencem a um ser humano vivo e presente no momento da captura, e não a uma representação artificial ou pré-gravada. A crescente sofisticação dos ataques, especialmente com a popularização de ferramentas de geração de *deepfakes*, impõe um desafio contínuo à pesquisa e ao desenvolvimento de mecanismos de detecção de vivacidade cada vez mais robustos e inteligentes (GALBALLY; MARCEL; FIERREZ, 2014).

Este Projeto de Fim de Curso se insere neste contexto desafiador, propondo o desenvolvimento e a avaliação de um protótipo de sistema de detecção de vivacidade multifacetado. O sistema combina diversas análises de características faciais e oculares, bem como interações específicas com o usuário, visando oferecer uma camada adicional de segurança contra ataques de apresentação em processos de autenticação facial.

1.1 Objetivos

O desenvolvimento deste trabalho foi norteado por um objetivo geral e complementado por objetivos específicos que detalham as etapas e metas alcançadas.

1.1.1 Objetivo Geral

O objetivo geral deste Projeto de Fim de Curso é projetar, implementar e avaliar um protótipo de sistema de detecção de vivacidade multifacetado, capaz de analisar diferentes pistas comportamentais e fisiológicas para distinguir entre um usuário genuíno e tentativas de fraude por apresentação em sistemas de autenticação baseados em reconhecimento facial.

1.1.2 Objetivos Específicos

Para alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- Realizar uma revisão bibliográfica abrangente sobre o estado da arte em técnicas de detecção de vivacidade, tipos de ataques de apresentação (incluindo *deepfakes*), e as principais ferramentas e tecnologias empregadas na área.
- Desenvolver e implementar múltiplos módulos de verificação de vivacidade, incluindo:
 - Análise e validação da pose da cabeça como pré-requisito para interações.
 - Detecção de piscadas espontâneas e sob comando, utilizando o *Eye Aspect Ratio* (EAR).
 - Um sistema de interação com alvos visuais na tela para indução e verificação implícita do direcionamento do olhar.
 - Análise da reação pupilar à emissão de luz (reflexo fotomotor).
 - Verificação da cor da esclera como um indicador antifraude.
- Utilizar as bibliotecas OpenCV e MediaPipe em Python como ferramentas principais para o processamento de imagens e extração de características faciais e oculares.
- Integrar os módulos desenvolvidos em um algoritmo piloto coeso, com um sistema de pontuação para agregar as evidências de vivacidade e auxiliar na tomada de decisão.
- Implementar interfaces distintas para "visão do desenvolvedor" (com informações detalhadas para depuração) e "visão do usuário" (com interações simplificadas).
- Avaliar experimentalmente a eficácia e a robustez do protótipo através de testes com usuários legítimos (utilizando webcam e câmera de smartphone via DroidCam) e simulações de ataques de apresentação (utilizando imagens estáticas e vídeos *deepfake*).
- Analisar criticamente os resultados obtidos, identificar os pontos fortes e as limitações da solução proposta, e propor direções para trabalhos futuros.

1.2 Organização do Trabalho

Este Trabalho de Fim de Curso está estruturado em cinco capítulos, de forma a apresentar o desenvolvimento da pesquisa de maneira lógica e coesa:

- **Capítulo 1 (Introdução):** Este capítulo introdutório contextualiza o problema da fraude em sistemas biométricos, justifica a necessidade de soluções de detecção de vivacidade, e apresenta os objetivos geral e específicos do trabalho, bem como a estrutura geral do documento.
- **Capítulo 2 (Revisão Bibliográfica):** Este capítulo (baseado no material previamente desenvolvido pelo autor) discorre sobre a demanda por sistemas antifraude no contexto da transformação digital, o mercado global dessas soluções, as ferramentas e o estado da arte em detecção de manipulações em mídias digitais, e as tecnologias emergentes e recorrentes, com destaque para os desafios impostos pelos *deepfakes* e a importância da reconstrução 3D da face.
- **Capítulo 3 (Metodologia de Desenvolvimento da Solução de Prova de Vida):** Detalha a arquitetura do sistema prototipado, as tecnologias e bibliotecas utilizadas (OpenCV, MediaPipe), os métodos de aquisição e pré-processamento de imagem, a extração de pontos fiduciais faciais, as técnicas de estimativa de pose da cabeça, a análise ocular multifacetada (detecção de íris, cálculo do EAR para piscadas, estimativa de direção do olhar), a implementação das interações de vivacidade (validação da centralização da cabeça, solicitação de piscadas, interação com alvos visuais), os mecanismos adicionais de *antispoofing* (análise da cor da esclera, reação pupilar à luz), e a integração dos módulos no algoritmo piloto com seu sistema de pontuação e as diferentes visões de interface.
- **Capítulo 4 (Resultados):** Apresenta a configuração dos testes experimentais, as métricas de avaliação de desempenho empregadas (TAR, FRR, TRR, TFA), e os resultados quantitativos e qualitativos obtidos nos testes com usuários legítimos (utilizando webcam e DroidCam) e nas simulações de ataques por apresentação (com imagens estáticas e vídeos *deepfake*). Inclui uma análise e discussão consolidada desses resultados.
- **Capítulo 5 (Conclusões):** Sumariza as principais contribuições do trabalho, re-toma os objetivos e verifica em que medida foram alcançados, discute as limitações identificadas na solução proposta e no estudo realizado, e apresenta recomendações e direções para trabalhos futuros, visando o aprimoramento e a expansão do protótipo.

2 Revisão Bibliográfica

Este capítulo conta com uma introdução geral ao tema, apresentando o contexto e o problema que será tratado neste trabalho, bem como trabalhos relacionados e bibliografia de apoio.

A Seção 2.1 discorre sobre a transformação digital nos meios de pagamento, correlacionando com seus riscos, e a necessidade de sistemas que assegurem a integridade dos dados e usuários.

A Seção 2.2 aborda o mercado global de sistemas de Antifraude, levantando suas necessidades e motivações e trazendo um apanhado geral de soluções já existentes. Além disso, traz alguns exemplos de aplicações e empresas que, no papel de cliente, buscam empregar antifraude em suas soluções.

A Seção 2.3 apresenta as ferramentas de detecção de manipulações gerais e o estado da arte para algoritmos de detecção de manipulações em arquivos de imagem e vídeo.

A Seção 2.4 explica como aplicar e utilizar tais ferramentas para garantir a integridade e segurança em associação do uso recorrente de tecnologias digitais para transações que expõem o usuário.

A Seção 2.5 faz um apanhado geral de tecnologias emergentes e recorrentes, e como buscar garantir e prever novos tipos de vetores de ataques, provenientes das constantes atualizações de *deepfakes* por acessibilidade e popularização das ferramentas.

2.1 Demanda de Antifraude

Com a popularização dos *smartphones* e o advento da *internet*, bancos tradicionais e chamados bancos digitais (sem existência de agência física) passaram a adotar aplicativos para facilitar a gestão de contas, transferências e pagamentos. Essa mudança de comportamento tem sido vista entre os brasileiros: cada vez mais eles passaram a utilizar os meios digitais para operações bancárias, pagamentos e compras *on-line*. De acordo com Rodrigo Mulinari, diretor da Federação Brasileira de Bancos (FEBRABAN), cerca de 70% das operações bancárias foram feitas via celular, no período de 2019 a 2023, e, neste mesmo período, houve um aumento de 251% deste uso (FEBRABAN; DELOITTE, 2024).

Um fator que contribuiu para o aumento destes números foi a implementação do Pix, no final do ano de 2020, um método de pagamentos instantâneo que pode ser utilizado em todo o país (FEBRABAN; DELOITTE, 2024).

No entanto, o crescimento de transações financeiras *online* tem atraído a atenção de golpistas e indivíduos mal-intencionados, que empregam diversas estratégias para ludibriar vítimas e apropriar-se de seus recursos financeiros ou dados sensíveis. Segundo a pesquisa ([Instituto DataSenado, 2024](#)), golpes digitais atingem cerca de 24% da população brasileira.

Visando evitar que terceiros utilizem contas bancárias de seus usuários, os bancos adotam ferramentas de Antifraude, que detectam e previnem golpes digitais, assegurando que os dados de seus clientes sejam tratados com segurança e confiabilidade e garantindo que a transação seja legítima do dono da conta.

A segurança em transações online e verificação de identidade requer o uso de diversas ferramentas tecnológicas robustas e eficazes. Dentre elas, destacam-se o Liveness, a Documentoscopia e o FaceMatch, cada qual com funcionalidades específicas e complementares. Liveness é um recurso avançado de biometria facial, e não se limita à simples detecção de um rosto. Ele realiza uma análise multifatorial, considerando aspectos como textura da pele, movimentos faciais sutis, reflexos e até mesmo a detecção de máscaras ou imagens impressas. A verificação vai além da mera presença de um rosto, analisando a dinâmica e características únicas de um rosto real e vivo, rejeitando, assim, tentativas de fraude com fotos, vídeos ou máscaras. Documentoscopia: Essa técnica vai além da simples leitura óptica de caracteres (OCR). A Documentoscopia realiza uma análise forense digital aprofundada do documento, verificando a autenticidade e integridade do documento oficial. Isso inclui a análise de diversos elementos de segurança, como: marcas d'água, microimpressões, elementos holográficos, fios de segurança, padrões de impressão, tipos de papel e tintas utilizadas, além da validação dos dados contra o banco de dados do órgão emissor. Qualquer discrepância ou anomalia é detectada, garantindo a validação da identidade do portador. FaceMatch: Este sistema de comparação biométrica facial utiliza algoritmos avançados para comparar a imagem facial do usuário com um banco de dados, buscando por correspondências. A comparação não se limita à semelhança superficial, mas analisa pontos faciais específicos e características geométricas, garantindo alta precisão e minimizando o risco de falsos positivos. Além disso, a base de dados utilizada é constantemente atualizada e protegida contra acessos não autorizados, garantindo a confiabilidade do processo de verificação.

A combinação dessas tecnologias oferece um sistema de segurança multicamadas, assegurando a verificação de identidade robusta e a prevenção de fraudes em ambientes digitais.

2.2 O Mercado de Antifraude

Como apresentado anteriormente, soluções de Antifraude são essenciais se tratando de transações financeiras e segurança bancária. Tais soluções podem ser desenvolvidas em setor próprio da empresa financeira, ou pode ser terceirizada por outras empresas. Dentre as grandes referências de Antifraude, podemos destacar Facetec, iProov, Amazon Rekognition. Abaixo, iremos discorrer melhor sobre suas soluções.

2.2.1 FaceTec

A FaceTec é uma empresa estadunidense com sede na Califórnia, com foco em implementações tecnológicas de segurança e antifraude. A empresa estabeleceu-se como líder global no segmento de verificação biométrica 3D, oferecendo soluções inovadoras para o mercado de antifraude. Especializada em detecção de "liveness"(prova de vida) e correspondência tridimensional, a empresa desenvolveu uma plataforma que combina segurança avançada com experiência do usuário otimizada. Seu sistema patenteado ZoOm-in FaceScan® cria um FaceMap® 3D a partir de um vídeo-selfie com duração de 2 a 3 segundos, revolucionando processos de verificação de identidade remota (FaceTec, Inc., 2025).

A tecnologia da FaceTec foi concebida especificamente para combater as crescentes ameaças de fraude digital, posicionando-se como uma barreira eficaz contra roubo de identidade, *phishing* e outros vetores de ataque contemporâneos. Sua abordagem baseada em 3D diferencia-se radicalmente das soluções biométricas tradicionais em 2D, oferecendo níveis de segurança significativamente superiores (FaceTec, Inc., 2025).

O diferencial tecnológico da FaceTec reside em seu sistema de criação de FaceMaps 3D, que captura não apenas características faciais superficiais, mas a estrutura geométrica única do rosto do usuário. Este processo envolve: coleta de mais de 100 quadros de vídeo, analisando a distorção de perspectiva conforme o usuário move o rosto; conversão dos dados 2D em um modelo 3D preciso, utilizando algoritmos de IA e visão computacional; armazenamento seguro nos servidores do cliente, não constituindo um *honeypot* biométrico, pois não contêm os dados de liveness necessários para autenticação (FaceTec, Inc., 2025).

A empresa estabeleceu um sistema de detecção de *liveness* com novos padrões na indústria, sendo certificado pelos laboratórios NIST/NVLAP (National Institute of Standards and Technology/National Voluntary Laboratory Accreditation Program)(National Institute of Standards and Technology, 2023). Através de uma abordagem multifatorial, analisa reflexos oculares e textura de pele, profundidade e movimento 3D, características fisiológicas humanas impossíveis de replicar com mídias estáticas. Além disso, a empresa mantém um programa de recompensas por vulnerabilidades (*Spoofy Bounty*) desde 2019,

que incentiva a comunidade de segurança a divulgar e reportar fraudes, além de reforçar a confiança em sua capacidade de resistir a tentativas de *spoofing*.

Dentre os vetores de ataques conhecidos, a FaceTec se garante em neutralizar efetivamente os seguintes métodos de fraude digital: fotos e vídeos 2D, estáticas ou em movimento; máscaras e réplicas, de papel, cera ou bonecos realistas; *deepfakes* e avatares; injeção de vídeo (FaceTec, Inc., 2025).

Efetivamente no mercado global e brasileiro, podemos trazer alguns exemplos, como DocCheck, solução da Exato Digital que utiliza o mapeamento 3D da FaceTec para validação facial e combate a fraudes em operações digitais; e ClearSale, que oferece a tecnologia FaceTec como uma das opções de liveness em sua plataforma de biometria facial para *e-commerce*, bancos e outros setores (Exato Digital, 2025).

2.2.2 iProov

A iProov também está entre as líderes globais em verificação biométrica de identidade, especializada em soluções antifraude baseadas em tecnologia de *liveness* e autenticação facial. Fundada em 2011 no Reino Unido, a empresa desenvolveu um sistema patenteado que combina biometria facial, iluminação controlada e IA, para assegurar que um indivíduo é "a pessoa certa, uma pessoa real, autenticando-se no momento" (iProov Limited, 2025a). Sua atuação abrange setores críticos como serviços financeiros, saúde e viagens.

A empresa utiliza um sistema de verificação facial tridimensional que diferencia seres humanos reais de *deepfakes*, máscaras, fotos estáticas e vídeos pré-gravados. Seu algoritmo de "*Flashmark*" emprega iluminação controlada para detectar reflexos fisiológicos únicos, como textura da pele e movimento ocular. (iProov Limited, 2025a)

De acordo com dados da própria empresa, novos vetores de ataque tem tido expressivo crescimento, como ataques de câmeras virtuais nativas, *face swap* e síntese de identidades sintéticas via IA generativa. Sendo assim, através de um Sistema de Gerenciamento de Ameaças Ativas (iSOC), a iProov monitora e atualiza continuamente suas defesas. Neste sentido, a empresa elaborou um relatório recente, que alerta para o mercado de "*Crime-as-a-Service*", onde ferramentas de deepfake são comercializadas a criminosos de baixa qualificação, e ataques de injeção de vídeo, que burlam sistemas tradicionais de verificação. (iProov Limited, 2025b)

Programas de identidade digital governamental, como o UK Digital Identity & Attributes Trust Framework, utilizam a iProov para verificação remota de documentos (RG, passaporte) com biometria facial, combatendo roubo de identidade e falsificação, e bancos como o UBS (União de Bancos Suíços) adotaram a iProov para *onboarding* digital seguro, reduzindo fraudes em abertura de contas e transações de alto valor.

Dentre os destaques e vantagens competitivas da iProov, podemos destacar seu FAR (*False Acceptance Rate*) de 1 em milhão; neutralidade étnica, evitando vieses e certificações atendendo a padrões internacionais, como: eIDAS, ISO/IEC 30107-3, e FIDO Alliance. (iProov Limited, 2025a)

2.2.3 Amazon Rekognition Face Liveness

A Amazon Rekognition Face Liveness é uma solução de detecção de vivacidade facial que utiliza Inteligência Artificial e Visão Computacional para combater fraudes em processos de verificação de identidade digital. Desenvolvida pela Amazon Web Services (AWS), a tecnologia é projetada para diferenciar usuários reais de tentativas de *spoofing*.

Bem como as soluções das empresas citadas anteriormente, a solução da Amazon é capaz de detectar vetores de ataques como vídeos pré-gravados, *deepfakes* e máscaras 3D.

Através de um vídeo-selfie com duração de 2 a 3 segundos, algoritmos de Machine Learning avaliam reflexos faciais, padrões de iluminação e estrutura 3D do rosto. Alguns diferenciais que a solução de *liveness* da Amazon incluem: um score de confiança, onde valores mais altos indicam maior probabilidade de o usuário ser real, podendo haver um ajuste de *threshold* conforme o risco do caso de uso. Outro diferencial é a jornada do usuário, onde luzes coloridas são emitidas na tela, com intuito de observar a reação dos olhos e pele à essas cores. A conformidade com normativas relevantes, como as diretrizes WCAG 2.1 para acessibilidade e o regulamento eIDAS, que atesta um alto nível de segurança, também constitui um diferencial. (Amazon Web Services, 2023b)

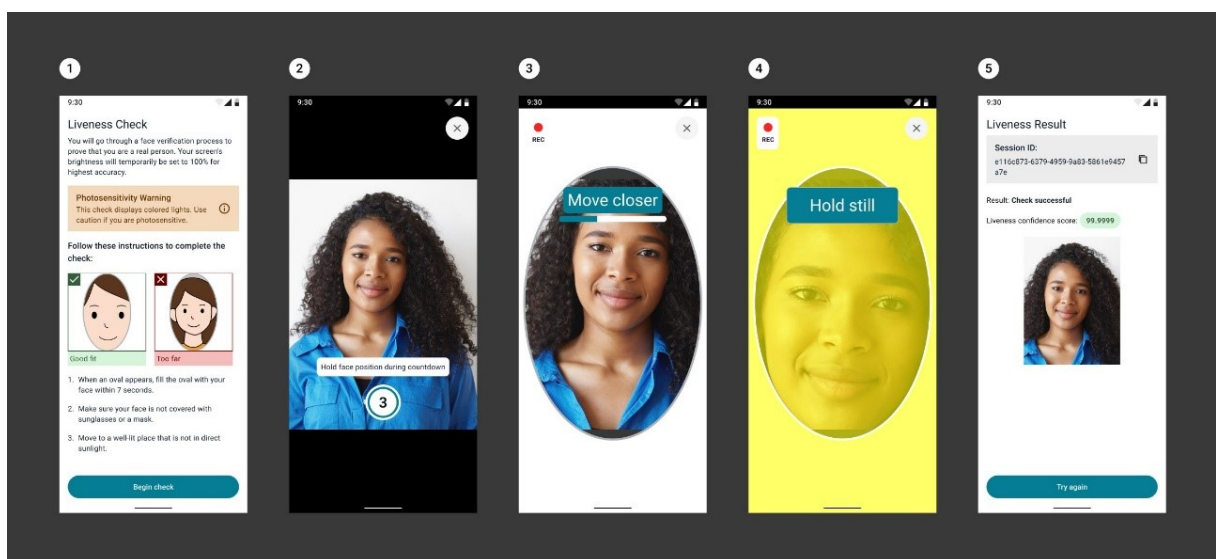


Figura 1 – Etapas de visualização de tela durante *liveness* desenvolvido pela Amazon.

Para assegurar a neutralidade e mitigar vieses discriminatórios, os modelos de ML são treinados com conjuntos de dados (datasets) amplamente diversificados. Essa

diversidade abrange variações significativas em tonalidade de pele, faixa etária e identidade de gênero, contribuindo para uma maior equidade na avaliação. ([Amazon Web Services, 2023c](#))

No mercado de antifraude, a tecnologia encontra diversas aplicações. Em processos de integração digital (*onboarding*), instituições financeiras utilizam a solução para validar a identidade de novos clientes durante a abertura de contas, realizando a comparação biométrica entre a vídeo-selfie e fotografias de documentos oficiais, como carteiras de identidade ou passaportes, frequentemente por meio de interfaces de programação de aplicações (APIs) como a CompareFaces. Adicionalmente, em plataformas de mobilidade urbana, como a Uber, e comércio eletrônico, a tecnologia é implementada para a reautenticação de utilizadores em transações consideradas de alto risco, como transferências bancárias. Setores com restrições etárias, como os de jogos online e tabaco, combinam a detecção de vivacidade com sistemas de estimativa de idade para coibir o acesso de menores. Finalmente, redes sociais e outros serviços digitais empregam esta solução como um substituto avançado para mecanismos de CAPTCHA, exigindo provas de vivacidade para dificultar a proliferação de contas automatizadas (bots) e identidades sintéticas. ([Amazon Web Services, 2023a](#))

2.3 Detecção de Manipulação de Imagens e Vídeos

A crescente sofisticação das ferramentas de edição e geração de imagens e vídeos digitais impulsionou significativamente a necessidade de métodos robustos para a detecção de manipulações. Tais manipulações, que variam desde edições simples até a criação de deepfakes complexos, representam um desafio considerável para a verificação da autenticidade de conteúdo digital, com implicações diretas na disseminação de desinformação e em fraudes. Esta seção discute as abordagens para a detecção dessas manipulações, iniciando pelas técnicas tradicionais baseadas na identificação de artefatos e inconsistências intrínsecas ao processo de alteração digital.

2.3.1 Técnicas Tradicionais e Baseadas em Artefatos

As primeiras abordagens para a detecção de manipulação de imagens frequentemente se concentravam na identificação de anomalias ou "impressões digitais" deixadas pelos processos de edição. Essas técnicas exploram inconsistências que surgem quando um conteúdo digital é alterado, assumindo que o processo de manipulação, por mais sofisticado que seja, introduzirá perturbações detectáveis em relação às características estatísticas de uma imagem autêntica.

2.3.1.1 Análise de Ruído e Inconsistências Locais

Um dos pilares da análise forense de imagens digitais é o estudo dos padrões de ruído. Cada dispositivo de captura, seja uma câmera digital ou um scanner, introduz um padrão de ruído idiossincrático, parcialmente devido à Não-Uniformidade da Resposta do Fotorreceptor (*Photo-Response Non-Uniformity* - PRNU). Este padrão, quase como uma impressão digital do sensor, pode ser utilizado para verificar a origem de uma imagem ou para detectar regiões que foram inseridas ou alteradas. A lógica subjacente é que uma manipulação, como a junção de partes de imagens distintas (*splicing*) ou a cópia de uma região para outra dentro da mesma imagem (*copy-move*), resultará em uma descontinuidade ou inconsistência nos padrões de ruído. O trabalho de Zhou et al. (ZHOU et al., 2018) ilustra essa abordagem ao propor uma arquitetura de rede neural de duas vias, onde um dos fluxos é especificamente projetado para analisar as características de ruído da imagem. Este "fluxo de ruído" emprega filtros de Modelos Ricos em Esteganálise (*Steganalysis Rich Models* - SRM) para extrair descritores de ruído local, com o objetivo de identificar discrepâncias entre as regiões autênticas e as potencialmente manipuladas (ZHOU et al., 2018). A premissa é que, mesmo que um falsificador tente aplicar técnicas de pós-processamento para suavizar as transições, as estatísticas de ruído da região adulterada raramente corresponderão perfeitamente às da imagem hospedeira (ZHOU et al., 2018).

Embora abordagens mais recentes, como a de Wu et al. (WU; CHEN; ZHOU, 2023), utilizem técnicas avançadas como aprendizado contrastivo e agrupamento não supervisionado, elas ainda se fundamentam no princípio de que características forenses intrínsecas podem discriminar entre conteúdo original e forjado. Métodos tradicionais frequentemente utilizavam algoritmos de agrupamento (*clustering*) aplicados a características de ruído extraídas manualmente, como o nível de ruído geral da imagem ou padrões específicos do ruído da câmera, para classificar diferentes blocos ou pixels da imagem como pertencentes a regiões genuínas ou manipuladas (WU; CHEN; ZHOU, 2023). No entanto, a eficácia dessas técnicas que dependem de características artesanais (*hand-crafted features*) é frequentemente limitada diante da crescente complexidade das manipulações e da aplicação de operações de pós-processamento que visam ocultar os vestígios da fraude (WU; CHEN; ZHOU, 2023).

2.3.1.2 Análise de Compressão JPEG

O formato JPEG é, indiscutivelmente, um dos mais prevalentes para o armazenamento e transmissão de imagens digitais. Seu algoritmo de compressão, baseado na Transformada Discreta de Cosseno (DCT) e em tabelas de quantização, introduz artefatos característicos, como o efeito de blocagem em regiões de baixa frequência e padrões específicos nos coeficientes DCT. Esses artefatos, embora muitas vezes sutis ao olho hu-

mano, podem ser explorados para fins forenses. Uma manipulação comum ocorre quando uma imagem já comprimida em JPEG é editada e, subsequentemente, salva novamente no formato JPEG. Este processo, conhecido como dupla compressão JPEG, pode introduzir anomalias e inconsistências nos padrões de compressão que não estariam presentes em uma imagem comprimida uma única vez (ZHOU et al., 2018). A detecção dessas assinaturas de dupla compressão tem sido uma linha de investigação forense estabelecida. Wu et al. (WU; CHEN; ZHOU, 2023) mencionam a utilização do ruído de quantização JPEG como uma das características exploradas por métodos de agrupamento para a detecção de falsificações. De forma complementar, o trabalho CAT-Net, referenciado por Wu et al. (WU; CHEN; ZHOU, 2023), foca na localização de regiões forjadas através da classificação dos coeficientes DCT, que são elementos centrais do processo de compressão JPEG. A análise da consistência desses coeficientes e das tabelas de quantização entre diferentes regiões de uma imagem pode, portanto, revelar áreas que foram submetidas a ciclos de compressão distintos, indicando uma possível manipulação. É importante notar que a robustez destas técnicas pode ser comprometida por operações de pós-processamento que visam atenuar ou remover os vestígios da compressão.

2.3.1.3 Outras Características Forenses

Além das assinaturas de ruído e dos artefatos de compressão, um espectro mais amplo de características forenses pode ser empregado para identificar manipulações visuais. Inconsistências na iluminação, como fontes de luz incompatíveis entre diferentes objetos ou pessoas em uma cena, direções de sombras que desafiam a geometria da cena, ou reflexos que não correspondem ao ambiente circundante, são frequentemente investigadas como indícios de adulteração. Similarmente, anomalias na perspectiva, onde objetos inseridos não se alinham corretamente com o ponto de fuga da imagem original, podem denunciar uma montagem.

A análise de textura, conforme detalhado em (MEDEIROS, 2001), embora aplicada a imagens médicas, oferece um arcabouço conceitual relevante. Texturas são padrões espaciais que podem ser caracterizados estatisticamente. Uma manipulação pode introduzir uma região com uma textura que destoa do seu entorno ou pode interromper abruptamente um padrão textural preexistente. Medidas como as características de Haralick, que são derivadas de matrizes de coocorrência de níveis de cinza, quantificam propriedades texturais como homogeneidade, contraste, correlação e entropia (MEDEIROS, 2001). Por exemplo, o Segundo Momento Angular (SMA), também conhecido como energia, é uma medida da homogeneidade local dos níveis de cinza. Sua fórmula é dada por:

$$SMA = \sum_i \sum_j P(i, j)^2 \quad (2.1)$$

onde $P(i, j)$ é a entrada da matriz de coocorrência normalizada para os níveis de cinza i e j . Uma imagem ou região homogênea tende a ter poucas transições de níveis de cinza,

resultando em uma matriz de coocorrência com poucas entradas de alta magnitude e, consequentemente, um valor de SMA elevado (MEDEIROS, 2001). O Contraste, por outro lado, mede a quantidade de variação local e é calculado como:

$$Contraste = \sum_i \sum_j (i - j)^2 P(i, j) \quad (2.2)$$

Valores elevados de contraste indicam uma maior dispersão dos níveis de cinza (MEDEIROS, 2001). No contexto forense, a inserção de um objeto de outra imagem pode resultar em uma assinatura textural distinta que pode ser capturada por essas métricas.

Adicionalmente, artefatos visuais nas bordas das regiões manipuladas são frequentemente explorados. Bordas que se apresentam excessivamente nítidas ou, inversamente, borradas de forma não natural em relação ao restante da imagem podem indicar uma colagem. Diferenças abruptas de contraste entre uma região suspeita e seu fundo imediato também são sinais de alerta (ZHOU et al., 2018). Outra característica investigada é a consistência dos padrões do Camera Filter Array (CFA). A maioria dos sensores de câmeras digitais utiliza um CFA (como o padrão Bayer) para capturar informações de cor, e o processo de interpolação para reconstruir uma imagem colorida completa deixa traços específicos. Uma região importada de outra câmera ou gerada sinteticamente pode não exibir os mesmos artefatos de CFA ou pode interromper os padrões existentes na imagem hospedeira (ZHOU et al., 2018). Zhou et al. (ZHOU et al., 2018) destacam que o fluxo RGB de sua rede neural é treinado para identificar justamente esses tipos de artefatos visuais, como as mencionadas fortes diferenças de contraste e bordas não naturais.

Embora estas técnicas tradicionais tenham estabelecido as fundações da detecção de manipulação de imagens, a sua eficácia individual é frequentemente limitada pela crescente complexidade e realismo das manipulações, bem como pelo emprego de contra-medidas forenses pelos perpetradores. Esta limitação pavimentou o caminho para o desenvolvimento de abordagens baseadas em aprendizado profundo, que demonstram maior capacidade de adaptação e generalização, e serão o foco da próxima discussão.

2.3.2 Tecnologia por Trás dos Deepfakes

A proliferação e o crescente realismo dos deepfakes são diretamente atribuíveis aos avanços em inteligência artificial (IA), mais especificamente no campo do aprendizado profundo (*deep learning*). Essa tecnologia permite a automação de tarefas cognitivas complexas, como a identificação e substituição de faces em vídeos, com um esforço e custo computacional relativamente baixos quando comparados às técnicas tradicionais de efeitos visuais cinematográficos (LANGGUTH et al., 2021). O cerne da tecnologia deepfake reside no uso de arquiteturas de redes neurais, principalmente Autoencoders e Redes Adversariais Generativas (GANs).

2.3.2.1 Autoencoders

Os Autoencoders são um tipo de rede neural artificial utilizada para aprender representações (codificações) eficientes de dados de forma não supervisionada. Um autoencoder consiste tipicamente em duas partes principais: um codificador (encoder) e um decodificador (decoder). O codificador mapeia os dados de entrada para uma representação latente de dimensionalidade geralmente inferior, capturando as características mais salientes da entrada. O decodificador, por sua vez, tenta reconstruir a entrada original a partir dessa representação latente (JIN; YE; CHEN, 2021). O treinamento de um autoencoder visa minimizar o erro de reconstrução entre a entrada original e a saída reconstruída.

No contexto da criação de *deepfakes*, particularmente para a troca de faces (*face swapping*), uma abordagem comum envolve o treinamento de dois autoencoders separados. Um autoencoder é treinado com um grande volume de imagens da pessoa de origem (cujo rosto será mapeado) e o outro com imagens da pessoa alvo (cujo rosto substituirá o da origem). Crucialmente, ambos os autoencoders compartilham o mesmo codificador, mas possuem decodificadores distintos, um para cada identidade (LANGGUTH et al., 2021). O codificador aprende a extrair características faciais genéricas e independentes da identidade, como expressões faciais e orientação da cabeça. Cada decodificador, então, aprende a reconstruir a face da sua respectiva identidade a partir dessas características latentes compartilhadas.

O processo de troca de faces ocorre ao alimentar o codificador com uma imagem da pessoa de origem e, em seguida, utilizar o vetor latente resultante como entrada para o decodificador treinado para a pessoa alvo. Isso resulta em uma imagem onde a face da pessoa alvo é renderizada com as expressões e a pose da pessoa de origem (LANGGUTH et al., 2021). Matematicamente, se E é o codificador compartilhado, D_A é o decodificador para a pessoa A (alvo), e D_B é o decodificador para a pessoa B (origem), e F_B é uma imagem da pessoa B, a face trocada F'_A é gerada como:

$$F'_A = D_A(E(F_B; \theta_E); \phi_A) \quad (2.3)$$

onde θ_E e ϕ_A são os parâmetros do codificador e do decodificador da pessoa A, respectivamente. Durante o treinamento, o objetivo é minimizar funções de perda como L_A e L_B , para cada pessoa, por exemplo, para a pessoa A, minimiza-se L_A :

$$\min L_A = \frac{1}{N} \sum_{i=1}^N ||F_{A,i} - D_A(E(F_{A,i}; \theta_E); \phi_A)||^2 \quad (2.4)$$

Onde N é o número de imagens de treinamento para a pessoa A, $F_{A,i}$ é a i -ésima imagem da pessoa A, e $|| \cdot ||^2$ denota o erro quadrático (JIN; YE; CHEN, 2021). Uma formulação similar se aplica para a pessoa B. A eficácia dessa abordagem depende da capacidade do codificador de aprender uma representação latente que dissocie as características da identidade das características da expressão e pose.

2.3.2.2 Redes Adversariais Generativas (GANs)

As Redes Adversariais Generativas (GANs), introduzidas por Goodfellow et al., representam um avanço significativo na capacidade de gerar dados sintéticos de alta fidelidade, incluindo imagens e vídeos faciais. Uma GAN consiste em duas redes neurais que competem entre si em um jogo de soma zero: um Gerador (G) e um Discriminador (D) (LANGGUTH et al., 2021).

O Gerador (G) tenta criar amostras de dados (por exemplo, imagens faciais) que sejam indistinguíveis de dados reais. Ele tipicamente recebe um vetor de ruído aleatório $\mathbf{z}$ como entrada e o transforma em uma amostra sintética $G(\mathbf{z})$. O Discriminador (D), por outro lado, é treinado para distinguir entre amostras reais (do conjunto de dados de treinamento) e amostras falsas (geradas por G). $D(\mathbf{x})$ produz a probabilidade de $\mathbf{x}$ ser uma amostra real.

O treinamento de uma GAN envolve um processo adversarial:

1. O Discriminador é treinado para maximizar sua capacidade de classificar corretamente amostras reais e falsas.
2. O Gerador é treinado para maximizar a probabilidade de o Discriminador classificar suas saídas como reais, ou seja, para "enganar" o Discriminador.

Esse processo de treinamento iterativo leva o Gerador a produzir amostras cada vez mais realistas e o Discriminador a se tornar cada vez mais proficiente em detectar falsificações. A função de valor $V(D, G)$ para uma GAN pode ser expressa como:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2.5)$$

onde $p_{data}(x)$ é a distribuição dos dados reais e $p_z(z)$ é a distribuição do ruído de entrada para o gerador.

As GANs foram fundamentais para a segunda geração de softwares *deepfake*, pois permitiram a geração de faces mais realistas e, crucialmente, reduziram a dependência de grandes conjuntos de dados de imagens de uma pessoa específica. O Discriminador, ao ser treinado para reconhecer variações da mesma face (por exemplo, de diferentes ângulos), mesmo com poucas imagens originais, auxilia o Gerador a criar sínteses mais robustas e convincentes (LANGGUTH et al., 2021). Variações de arquiteturas GAN, como StyleGAN, têm demonstrado capacidades impressionantes na geração de faces fotorrealistas, controlando aspectos como estilo e características faciais em diferentes níveis de detalhe.

A combinação de autoencoders com GANs também é comum, onde a qualidade da reconstrução do autoencoder pode ser aprimorada por um discriminador que força as saídas a parecerem mais realistas. A evolução contínua dessas arquiteturas de aprendizado

profundo representa um desafio constante para as técnicas de detecção, pois a qualidade e o realismo das manipulações tendem a aumentar com o tempo.

2.3.3 Abordagens Baseadas em Deep Learning para Detecção de Manipulação Facial e Geral

Com a limitação das técnicas tradicionais em lidar com a crescente sofisticação das manipulações digitais, as abordagens baseadas em aprendizado profundo (*deep learning*) emergiram como o estado da arte na detecção de falsificações em imagens e vídeos. Esses métodos, em particular as Redes Neurais Convolucionais (CNNs), demonstram uma capacidade superior de aprender representações de características complexas diretamente dos dados, permitindo a identificação de padrões sutis de manipulação que frequentemente passam despercebidos pelas técnicas baseadas em artefatos predefinidos.

2.3.3.1 Redes Neurais Convolucionais (CNNs) para Extração de Características e Detecção

As Redes Neurais Convolucionais (CNNs) tornaram-se a espinha dorsal de muitas soluções modernas para a detecção de manipulação de imagens e vídeos. A arquitetura hierárquica das CNNs, composta por camadas convolucionais, de agrupamento (*pooling*) e totalmente conectadas, permite que a rede aprenda automaticamente características relevantes em diferentes níveis de abstração (SALLES; FILHO; JUNIOR, 2021; LANG-GUTH et al., 2021). Nas camadas iniciais, a rede tende a aprender características de baixo nível, como bordas e texturas, enquanto camadas mais profundas capturam padrões mais complexos e semânticos. Essa capacidade de aprendizado de características é particularmente vantajosa na detecção de falsificações, pois os artefatos de manipulação podem ser diversos e não facilmente definidos por regras manuais.

No contexto da detecção de manipulação facial, Stehouwer et al. (STEHOUWER et al., 2019) investigam o uso de CNNs, como a XceptionNet, como espinha dorsal (*backbone*) para classificar imagens faciais como reais ou manipuladas. Eles propõem a inserção de uma camada baseada em atenção para melhorar o desempenho da classificação e localizar as regiões manipuladas. A CNN é treinada para realizar uma classificação binária, e a hipótese é que a rede aprende a identificar as anomalias de alta frequência ou "impressões digitais" modificadas pelo processo de manipulação (STEHOUWER et al., 2019). De maneira similar, Zhou et al. (ZHOU et al., 2018) utilizam uma arquitetura baseada na Faster R-CNN, uma conhecida rede de detecção de objetos, adaptada para a tarefa de localizar regiões manipuladas. Um dos fluxos dessa rede (o fluxo RGB) utiliza uma ResNet-101 para aprender características diretamente dos pixels da imagem, visando capturar artefatos visuais como inconsistências nas bordas e diferenças de contraste (ZHOU et al., 2018). A capacidade das CNNs de aprender representações ricas a partir dos dados

brutos é um diferencial crucial em relação às técnicas que dependem de descritores de características engenheirados manualmente.

Appalaraju e Chaoji ([APPALARAJU; CHAOJI, 2017](#)) também empregam CNNs, especificamente em uma arquitetura de rede siamesa multi-escala, para aprender similaridade de imagens. Embora o objetivo primário seja a busca por imagens similares, a capacidade da CNN de gerar *embeddings* (representações vetoriais compactas) robustos é fundamental. Esses *embeddings* capturam as características distintivas de uma imagem, e em um contexto forense, *embeddings* de regiões manipuladas podem diferir significativamente de regiões autênticas, permitindo a sua detecção através da análise dessas representações. Eles destacam que CNNs, como a VGG16, pré-treinadas em grandes conjuntos de dados como o ImageNet, servem como excelentes extratores de características que podem ser ajustados (*fine-tuned*) para tarefas específicas ([APPALARAJU; CHAOJI, 2017](#)).

A tarefa de detecção de falsificações em vídeos deepfake também se beneficia enormemente das CNNs. Jin et al. ([JIN; YE; CHEN, 2021](#)) utilizam uma CNN para classificar mapas de fotopletismografia (PPG) gerados a partir de vídeos, com o objetivo de distinguir entre vídeos reais e falsos. A rede, composta por seis camadas convolucionais seguidas por camadas totalmente conectadas, aprende a identificar padrões nos sinais biológicos extraídos que são alterados ou destruídos pelo processo de geração de *deepfakes* ([JIN; YE; CHEN, 2021](#)). A arquitetura de uma CNN típica para processamento de imagens envolve neurônios arrançados tridimensionalmente (largura, altura e profundidade), o que se adequa naturalmente ao tratamento de imagens onde a profundidade pode representar os canais de cor (por exemplo, RGB) ou mapas de características aprendidas ([SALLES; FILHO; JUNIOR, 2021](#)).

Um desafio comum no treinamento de CNNs para detecção de manipulação é a necessidade de grandes volumes de dados de treinamento que cubram uma vasta gama de tipos de manipulação e conteúdos. A falta de datasets públicos diversificados e de grande escala foi, por um tempo, um gargalo, mas esforços recentes, como o FaceForensics++ ([LANGGUTH et al., 2021](#)), têm ajudado a mitigar esse problema, fornecendo *benchmarks* para o desenvolvimento e avaliação de novos modelos de detecção baseados em CNNs.

2.3.3.2 Detecção Baseada em Sinais Biológicos e Fisiológicos

Uma vertente promissora na detecção de vídeos faciais manipulados, especialmente os *deepfakes*, explora a análise de sinais biológicos e fisiológicos que, embora muitas vezes imperceptíveis ao olho humano, estão intrinsecamente presentes em gravações de indivíduos reais. A hipótese central é que os algoritmos de geração de *deepfake*, particularmente aqueles baseados em Redes Adversariais Generativas (GANs), embora capazes de produzir resultados visualmente convincentes, podem falhar em replicar com precisão ou podem inadvertidamente destruir esses sinais sutis inerentes à fisiologia humana ([JIN;](#)

YE; CHEN, 2021). Essa discrepância entre o conteúdo real e o sintético oferece uma via para a diferenciação forense.

Um dos sinais biológicos mais investigados para essa finalidade é o sinal fotople-tismográfico (PPG), que pode ser estimado remotamente (rPPG) através da análise de vídeo. O princípio do rPPG baseia-se no fato de que o fluxo sanguíneo pulsátil sob a pele causa microvariações na absorção e reflexão da luz. Essas variações, embora mínimas, podem ser capturadas por uma câmera e processadas para estimar a frequência cardíaca e outros parâmetros cardiovasculares (JIN; YE; CHEN, 2021). Jin et al. (JIN; YE; CHEN, 2021) propõem um método forense que extrai e analisa a diferença nos sinais biológicos entre vídeos reais e falsos. Eles argumentam que, embora os GANs possam enganar a percepção visual humana, eles tendem a destruir os sinais biológicos originais, como os de frequência cardíaca, presentes no vídeo real.

Para capturar essas variações, Jin et al. (JIN; YE; CHEN, 2021) geram mapas, denominados *ppg_{map}*, a partir de sequências de quadros de vídeo. O processo envolve a detecção e alinhamento facial, segmentação da pele para remover influências de fundo e de regiões não cutâneas (como olhos e cabelo), e a divisão da região facial em múltiplos blocos. Os sinais são então extraídos de cada bloco em diferentes dimensões ou espaços de cor. Especificamente, eles exploram:

- Dimensão do Espaço de Cor RGB: As variações nos canais Vermelho (R), Verde (G) e Azul (B) são analisadas, pois a hemoglobina no sangue absorve a luz de maneira diferente nesses comprimentos de onda, e as pulsações alteram sutilmente a coloração da pele.
- Dimensão do Espaço de Cor YUV: Este espaço separa a luminância (Y) dos componentes de cromaticidade (U e V). A análise do componente Y permite focar nas alterações de brilho da pele induzidas pelo fluxo sanguíneo.
- Dimensão de Crominância: Para reduzir os efeitos de ruído ambiental e artefatos de movimento da cabeça, um método baseado em cromaticidade é empregado. O sinal de cromaticidade C pode ser calculado como uma combinação linear de dois sinais X_f e Y_f , que são versões filtradas (passa-banda) de sinais derivados dos canais RGB normalizados (R_n, G_n, B_n).

As formulações são dadas por:

$$C = X_f - \alpha Y_f \quad (2.6)$$

onde o fator de ponderação α é a razão dos desvios padrão dos sinais filtrados:

$$\alpha = \frac{\sigma(X_f)}{\sigma(Y_f)} \quad (2.7)$$

Os sinais X e Y (antes da filtragem) são definidos como:

$$X = 3R_n - 2G_n \quad (2.8)$$

$$Y = 1.5R_n + G_n - 1.5B_n \quad (2.9)$$

Esses ppg_{maps} multidimensionais são então utilizados para treinar uma Rede Neural Convolutacional (CNN) para classificar os vídeos como reais ou falsos (JIN; YE; CHEN, 2021).

Outros sinais fisiológicos também têm sido explorados. O padrão de piscadas de olhos, por exemplo, é um comportamento humano natural que pode não ser consistentemente replicado em vídeos *deepfake*. Li et al. (citado em (JIN; YE; CHEN, 2021) como referência [20] do artigo deles, e também mencionado como um possível artefato em discussões sobre *deepfakes* por Langguth et al. (LANGGUTH et al., 2021)) propuseram métodos para detectar piscadas anormais em vídeos falsos. Similarmente, inconsistências nos movimentos da cabeça e nas microexpressões faciais podem servir como indicadores de manipulação. Agarwal et al. e Yang et al. (citados em (JIN; YE; CHEN, 2021) como referências [19] e [21], respectivamente) investigaram o rastreamento de movimentos faciais e da cabeça e a inconsistência de poses 3D da cabeça para distinguir vídeos reais de falsos.

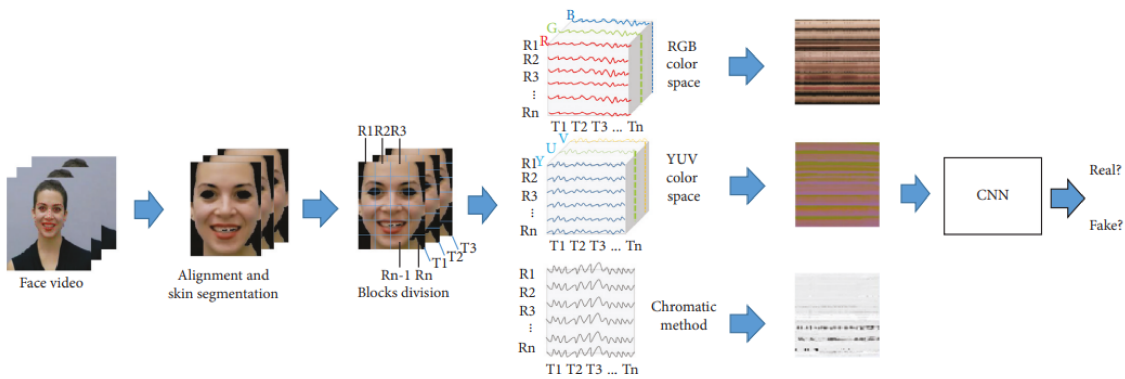


Figura 2 – Visão Geral do método proposto em (JIN; YE; CHEN, 2021)

A principal vantagem da utilização de sinais biológicos e fisiológicos reside na sua natureza fundamentalmente ligada à vida e ao comportamento humano, algo que os algoritmos sintéticos ainda lutam para emular com perfeição em todas as suas nuances temporais e espaciais. No entanto, a extração robusta desses sinais em ambientes não controlados ("*in the wild*") continua sendo um desafio, devido a fatores como variações de iluminação, movimentos do sujeito, oclusões parciais e qualidade da compressão do vídeo (JIN; YE; CHEN, 2021). Apesar desses desafios, a combinação da análise de sinais biológicos com o poder de aprendizado das redes neurais profundas oferece uma direção promissora e complementar às abordagens baseadas puramente em artefatos visuais.

2.3.3.3 Mecanismos de Atenção para Localização de Regiões Manipuladas

Além da classificação binária de uma imagem ou vídeo como autêntico ou manipulado, a capacidade de localizar as regiões específicas que foram alteradas é de grande importância forense. Os mecanismos de atenção, originalmente desenvolvidos em áreas como tradução automática e reconhecimento de objetos, foram adaptados com sucesso para a tarefa de detecção de manipulação, permitindo que as redes neurais não apenas identifiquem a presença de uma falsificação, mas também "concentrem-se" nas áreas da imagem que fornecem as evidências mais fortes dessa manipulação.

Stehouwer et al. (STEHOUSER et al., 2019) propõem uma abordagem que utiliza um mecanismo de atenção para processar e refinar os mapas de características intermediárias de uma rede neural convolucional (CNN) classificadora. A intuição fundamental é que, ao manipular uma imagem facial, seja por troca de identidade, alteração de expressão ou atributos, ou mesmo síntese completa, as modificações raramente são uniformes em toda a imagem e tendem a deixar traços localizados. Um mapa de atenção, nesse contexto, é uma representação espacial que atribui pesos ou probabilidades diferentes a distintas regiões da imagem (ou de seus mapas de características), indicando a relevância de cada região para a tarefa de detecção. A premissa é que um mapa de atenção bem treinado pode destacar as regiões efetivamente manipuladas, permitindo que a rede foque nessas áreas para tomar sua decisão e, como subproduto, fornecendo uma localização visual da manipulação (STEHOUSER et al., 2019).

Os autores argumentam que a eficácia de um mapa de atenção na detecção de manipulações se deve à sua capacidade de discernir anomalias em informações de alta frequência nas *patches* locais da imagem. Acredita-se que as "impressões digitais" da câmera ou os padrões naturais da imagem são perturbados durante a manipulação algorítmica. Assim, ao inserir uma camada de atenção em estágios da rede onde o campo receptivo corresponde a *patches* de tamanho apropriado, os mapas de características que alimentam essa camada podem codificar essas impressões digitais de alta frequência, permitindo a discriminação entre regiões reais e manipuladas em um nível local (STEHOUSER et al., 2019).

Stehouwer et al. (STEHOUSER et al., 2019) exploram duas principais implementações para a camada baseada em atenção, que processa um mapa de características $F \in \mathbb{R}^{B \times H \times W \times C}$ (onde B é o tamanho do lote, H , W , C são altura, largura e número de canais, respectivamente) para gerar um mapa de atenção $M_{att} \in \mathbb{R}^{B \times H \times W}$:

- **Projeção PCA Auto-Atenção (*Self-attention PCA Projection*):** Esta abordagem não introduz parâmetros treináveis adicionais. O mapa de características F é primeiro remodelado para uma matriz $X \in \mathbb{R}^{(B \cdot H \cdot W) \times C}$. Em seguida, a Análise de Componentes Principais (PCA) é utilizada para encontrar o autovetor $v \in \mathbb{R}^{C \times 1}$ correspondente

ao maior autovalor da matriz de covariância de X (após centralização pela média μ). Este autovetor v representa a direção de maior variância nos dados, que, no contexto da classificação entre real e falso, idealmente corresponderia à dimensão mais discriminativa. O mapa de atenção é então obtido projetando os dados centralizados $X - \mu$ sobre v e aplicando uma função sigmoide:

$$M_{att} = \text{Sigmoid}((X - \mu)v) \quad (2.10)$$

A função sigmoide mapeia os valores resultantes para o intervalo $[0, 1]$, onde valores próximos a 1 indicam regiões suspeitas de manipulação e valores próximos a 0 indicam regiões autênticas (STEHOUEW et al., 2019).

- Regressão Direta (Direct Regression): Alternativamente, o mapa de atenção pode ser gerado através de uma ou mais camadas convolucionais auxiliares. Uma implementação simples consiste em adicionar um canal extra na camada convolucional que precede a camada de atenção, resultando em um mapa de características $F' \in \mathbb{R}^{B \times H \times W \times (C+1)}$. Este pode ser dividido em $F_1 \in \mathbb{R}^{B \times H \times W \times C}$ (as características originais) e $M_{conv} \in \mathbb{R}^{B \times H \times W \times 1}$ (o mapa de atenção bruto). Após aplicar uma função sigmoide a M_{conv} , obtém-se M_{att} . Este método, embora simples, introduz parâmetros treináveis adicionais (os pesos da(s) camada(s) convolucional(is) que geram M_{conv}) (STEHOUEW et al., 2019).

Uma vez gerado o mapa de atenção M_{att} , ele é multiplicado (*channel-wise*) pelo mapa de características original F (ou F_1 no caso da regressão). Essa operação pondera as características, amplificando regiões consideradas importantes (manipuladas) e atenuando as irrelevantes (autênticas). O resultado, denotado por $F \odot M_{att}$ (onde $\odot$ indica multiplicação canal a canal), é passado para as camadas subsequentes da rede para classificação final (STEHOUEW et al., 2019).

O aprendizado de M_{att} pode ser guiado por:

$$L_{map} = |M_{att} - M_{gt}| \quad (2.11)$$

Para imagens reais, M_{gt} é um mapa de zeros; para faces totalmente sintéticas, é um mapa de uns. Para manipulações parciais, M_{gt} é obtida pela diferença entre a imagem manipulada e a original. Em cenários onde M_{gt} não está disponível para todas as amostras (aprendizado fracamente supervisionado), uma perda alternativa é proposta:

$$L_{map} = \begin{cases} |M_{att}^{(i)} - 0|, & \text{se real} \\ |\max(M_{att}^{(i)}) - 0.75|, & \text{se falso} \end{cases} \quad (2.12)$$

onde $M_{att}^{(i)}$ é o mapa de atenção para a i -ésima amostra. Esta formulação encoraja o mapa a ser zero para imagens reais e a ter pelo menos uma região com alta ativação (próxima a

0.75) para imagens falsas (STEHOUEW et al., 2019). Também é possível treinar a rede sem supervisão explícita para o mapa ($L_{map} = 0$), permitindo que a camada de atenção aprenda regiões informativas apenas através da retropropagação do erro da tarefa de classificação principal. A combinação da perda de classificação $L_{classifier}$ com a perda do mapa L_{map} forma a perda total $L = \lambda_c L_{classifier} + \lambda_m L_{map}$, onde λ_c e λ_m são pesos de ponderação (STEHOUEW et al., 2019).

A utilização de mecanismos de atenção, portanto, não só oferece um meio de interpretar quais regiões da imagem influenciam a decisão da rede, mas também pode ativamente melhorar o desempenho da detecção ao permitir que o modelo refine suas características e se concentre nos aspectos mais discriminativos da entrada.

2.4 Análise de Documentos e Identidade (Documentoscopia e Biometria Facial)

A verificação da identidade de um indivíduo e a autenticidade dos documentos por ele apresentados são pilares fundamentais na prevenção de fraudes, especialmente em processos de cadastro (*onboarding*) digital, transações financeiras e controle de acesso. A combinação de técnicas de documentoscopia e biometria facial, impulsionadas por avanços em inteligência artificial, oferece um arsenal robusto para mitigar os riscos associados à falsificação de identidade e ao uso de documentos adulterados.

2.4.1 Verificação de Autenticidade de Documentos

A documentoscopia é a disciplina forense dedicada ao exame de documentos com o objetivo de verificar sua autenticidade e identificar possíveis adulterações, falsificações ou fraudes. No contexto digital, a análise de documentos de identificação (como carteiras de identidade, passaportes, carteiras de habilitação, entre outros) submetidos eletronicamente tornou-se uma etapa crítica. Conforme descrito anteriormente na Seção 2.1, a documentoscopia digital vai além da simples leitura óptica de caracteres (OCR). Ela envolve uma análise forense aprofundada que examina diversos elementos de segurança intrínsecos aos documentos oficiais, como marcas d'água digitais ou visíveis sob certas condições de iluminação, microimpressões, elementos holográficos, padrões de impressão específicos, e a consistência dos dados extraídos com bancos de dados oficiais dos órgãos emissores.

Soluções de documentoscopia automatizada empregam algoritmos de visão computacional e aprendizado de máquina para analisar imagens de documentos capturadas por dispositivos móveis ou scanners. Esses algoritmos são treinados para:

- Classificar o tipo de documento: Identificar se é um RG, CNH, passaporte, etc.

- Verificar a presença e integridade de elementos de segurança: Comparar os elementos visuais do documento submetido com os padrões conhecidos para aquele tipo de documento.
- Detectar sinais de adulteração: Procurar por inconsistências na tipografia, alinhamento de texto, qualidade de impressão, ou indícios de manipulação digital da imagem do documento.
- Extrair e validar dados: Utilizar OCR para extrair informações textuais e cruzá-las com bases de dados governamentais ou outras fontes confiáveis para confirmar sua validade.

A combinação dessas análises permite um alto grau de confiança na determinação da autenticidade de um documento. Empresas como a Exato Digital, com sua solução DocCheck que integra tecnologias como as da FaceTec, exemplificam a aplicação prática desses conceitos para validação e combate a fraudes em operações digitais.

2.4.2 Biometria Facial para Verificação de Identidade

A biometria facial emergiu como uma das modalidades mais convenientes e amplamente adotadas para a verificação de identidade. Ela se baseia na premissa de que as características faciais de um indivíduo são únicas e podem ser usadas para distingui-lo de outros.

2.4.2.1 Reconstrução 3D da Face a Partir de Imagens 2D

A recuperação da forma tridimensional (3D) de uma face a partir de uma ou mais imagens bidimensionais (2D) é um problema clássico em visão computacional e de grande relevância para a biometria facial robusta. A informação 3D pode ajudar a superar limitações da análise 2D, como variações de pose e iluminação. Kamyab e Azimifar ([KAMYAB; AZIMIFAR, 2024](#)) abordam o problema da recuperação da forma 3D de um conjunto de *landmarks* faciais 2D extraídos de uma única imagem. Eles propõem um *framework* chamado DeepMDS, que utiliza uma abordagem de Escalonamento Multidimensional (*Multi-Dimensional Scaling* - MDS) para encontrar uma incorporação (*embedding*) 3D para os pontos 2D.

Tradicionalmente, o MDS é usado para redução de dimensionalidade, preservando as distâncias par a par entre os pontos. No entanto, para aumentar a dimensionalidade (de 2D para 3D), a matriz de dissimilaridade fornecida ao MDS precisa conter informações sobre as distâncias no espaço 3D desejado. Kamyab e Azimifar ([KAMYAB; AZIMIFAR, 2024](#)) propõem o uso de uma rede neural profunda para aprender essa matriz de dissimilaridade. A rede recebe pares de *landmarks* 2D (x_i, x_j) e estima a distância Euclidiana 3D

(D_{ij}^{3D}) entre seus correspondentes no espaço 3D. A matriz de dissimilaridade D , composta por essas estimativas D_{ij}^{3D} , é então utilizada pelo MDS para recuperar as coordenadas 3D dos *landmarks* (KAMYAB; AZIMIFAR, 2024). Para lidar com variações de pose e oclusões parciais que podem confundir a rede de dissimilaridade, um autoencoder é usado como um componente de calibração para transformar diferentes visualizações da face em uma visualização de perfil padronizada antes da estimativa de dissimilaridade (KAMYAB; AZIMIFAR, 2024). Essa abordagem destaca a sinergia entre métodos analíticos clássicos (MDS) e aprendizado profundo para resolver problemas complexos de visão computacional.

2.4.2.2 Uso de Landmarks Faciais e Aprendizado Profundo para Análise Facial

Os *landmarks* faciais, que são pontos-chave anatomicamente significativos na face (como cantos dos olhos, ponta do nariz, cantos da boca, contorno do maxilar, etc.), desempenham um papel crucial em diversas aplicações de análise facial, incluindo reconhecimento facial, rastreamento de expressões e, no contexto de fraude, verificação de identidade. A detecção precisa e robusta desses *landmarks* é um pré-requisito para muitas dessas tarefas.

Conforme discutido por Kamyab e Azimifar (KAMYAB; AZIMIFAR, 2024), a utilização de *landmarks* como características geométricas é vantajosa por ser menos sensível a variações de iluminação e outras condições radiométricas em comparação com a análise baseada em toda a informação fotométrica da imagem. Diversos algoritmos e métodos, muitos deles baseados em aprendizado profundo, foram desenvolvidos para a detecção automática de *landmarks* faciais a partir de imagens 2D (KAMYAB; AZIMIFAR, 2024).

Uma solução proeminente e de código aberto para detecção de *landmarks* faciais é o MediaPipe Face Landmarker, desenvolvido pelo Google. Esta ferramenta é capaz de detectar *landmarks* faciais em tempo real a partir de imagens ou vídeos. De acordo com sua documentação oficial, o MediaPipe Face Landmarker pode identificar 468 *landmarks* faciais 3D em uma face humana (Google for Developers, 2024). Além das coordenadas espaciais (x,y,z), a solução também pode fornecer matrizes de transformação para realizar o rastreamento de efeitos faciais e *blendshapes* (coeficientes de formas de mistura) que representam expressões faciais comuns, como sorriso, fechamento dos olhos, etc. (Google for Developers, 2024). Esses *blendshapes* são pontuações que indicam a intensidade de diferentes unidades de ação facial, permitindo uma representação quantitativa das expressões.

A capacidade de detectar um conjunto denso de *landmarks* 3D e *blendshapes* em tempo real, como oferecido pelo MediaPipe Face Landmarker, tem aplicações diretas na verificação de identidade e detecção de fraude. Por exemplo, a configuração geométrica dos *landmarks* pode ser usada para criar um descritor facial robusto para comparação

(FaceMatch, como mencionado na Seção 2.1). As informações de *blendshapes* podem ser cruciais para a detecção de vivacidade, analisando se as expressões faciais apresentadas são naturais ou correspondem a padrões de ataques de apresentação (como o uso de uma foto estática ou um vídeo pré-gravado). A tecnologia subjacente ao MediaPipe Face Landmarker envolve modelos de aprendizado de máquina otimizados para execução eficiente em diversas plataformas, incluindo dispositivos móveis e web (Google for Developers, 2024).

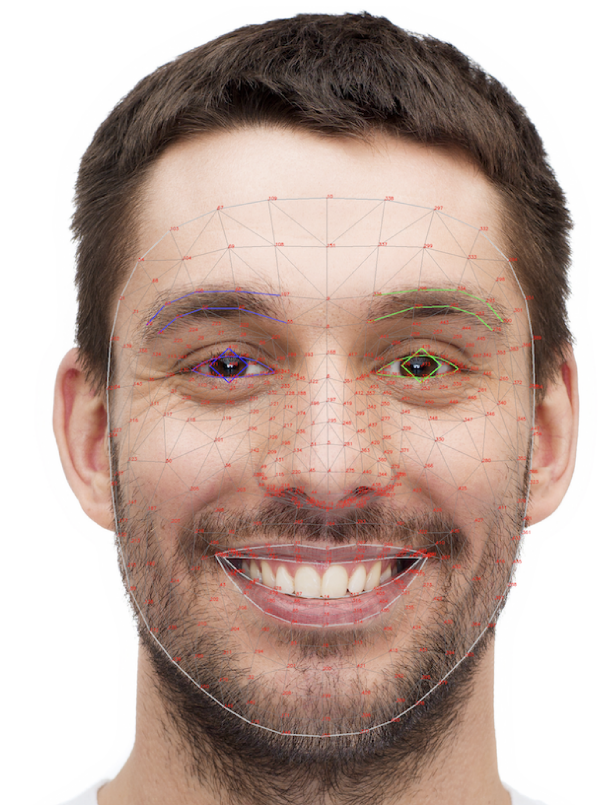


Figura 3 – *Landmarks* faciais utilizados pelo Mediapipe

2.4.2.3 Detecção de Vivacidade (*Liveness Detection*)

A simples correspondência de uma face apresentada com uma face armazenada em um banco de dados (*FaceMatch*) não é suficiente para garantir a segurança em muitos cenários, pois um fraudador poderia utilizar uma fotografia, um vídeo ou mesmo uma máscara 3D da pessoa legítima para tentar se passar por ela. A detecção de vivacidade (*liveness detection*) é, portanto, uma componente essencial dos sistemas de verificação biométrica facial, projetada para assegurar que a face sendo apresentada ao sensor pertença a um ser humano vivo e presente no momento da verificação, e não a uma representação artificial.

Conforme introduzido na Seção 2.1, a detecção de vivacidade envolve a análise de múltiplos fatores. Soluções como as da FaceTec, iProov e Amazon Rekognition Face Liveness utilizam abordagens sofisticadas que vão além da simples detecção de um rosto. A FaceTec, por exemplo, emprega um sistema patenteado de escaneamento facial 3D

(ZoOm-in FaceScan®) que cria um mapa facial 3D (FaceMap®) a partir de um curto vídeo-selfie. Este mapa captura a estrutura geométrica única do rosto, permitindo uma análise que considera textura da pele, movimentos faciais sutis e reflexos oculares para diferenciar uma face real de tentativas de *spoofing* com fotos, vídeos ou máscaras 2.2.2.

A iProov também se especializa em verificação biométrica de identidade com foco em *liveness*, utilizando uma combinação de biometria facial, iluminação controlada (como a técnica "*Flashmark*" que analisa reflexos fisiológicos) e IA para assegurar que o indivíduo é "a pessoa certa, uma pessoa real, autenticando-se no momento" 2.2.2. A Amazon Rekognition Face Liveness, por sua vez, utiliza IA e Visão Computacional, incluindo a emissão de luzes coloridas na tela durante um vídeo-selfie para observar a reação dos olhos e da pele, a fim de diferenciar usuários reais de vetores de ataque como vídeos pré-gravados, *deepfakes* e máscaras 3D 2.2.3.

O uso de *blendshapes* faciais, como os fornecidos pelo MediaPipe Face Landmarker (Google for Developers, 2024), pode complementar essas técnicas, permitindo a análise da dinâmica e naturalidade das expressões faciais como um indicador adicional de vivacidade. Essas tecnologias de detecção de vivacidade são cruciais para mitigar os riscos de ataques de apresentação e garantir a integridade dos processos de verificação de identidade baseados em biometria facial.

2.5 Desafios e Limitações Atuais na Detecção de Fraudes Inteligentes

Apesar dos avanços significativos nas tecnologias de detecção de fraudes, especialmente aquelas impulsionadas por inteligência artificial, o campo enfrenta uma série de desafios e limitações persistentes. A natureza dinâmica e adaptativa das técnicas de falsificação, juntamente com as complexidades inerentes aos próprios modelos de detecção e as implicações éticas e legais, configura um cenário onde a vigilância e a inovação contínuas são indispensáveis.

2.5.1 Evolução Constante das Técnicas de Manipulação

Um dos desafios mais proeminentes na detecção de fraudes visuais é a rápida e contínua evolução das próprias técnicas de manipulação. O advento dos *deepfakes*, impulsionado principalmente por Redes Adversariais Generativas (GANs), exemplifica essa dinâmica. As GANs, através de seu processo de treinamento adversarial entre um gerador e um discriminador, são intrinsecamente projetadas para aprimorar o realismo das amostras sintéticas, tornando-as cada vez mais difíceis de distinguir do conteúdo autêntico (LANGGUTH et al., 2021). Isso cria um ciclo contínuo, muitas vezes descrito como

um "jogo de gato e rato" ou uma "corrida armamentista", onde os avanços nas técnicas de detecção são rapidamente contrapostos por melhorias nas metodologias de falsificação (LANGGUTH et al., 2021).

Langguth et al. (LANGGUTH et al., 2021) destacam que, embora a manipulação de vídeo não seja um fenômeno novo, a mudança fundamental trazida pelos *deepfakes* reside na sua acessibilidade e no baixo custo de produção, permitindo que indivíduos com habilidades técnicas moderadas e hardware de consumo criem vídeos manipulados convincentes. As primeiras gerações de software *deepfake* baseadas em autoencoders exigiam um grande volume de imagens de treinamento, limitando seu uso a figuras públicas. No entanto, a incorporação de GANs na segunda e terceira gerações de software reduziu essa dependência, possibilitando a criação de *deepfakes* com menos dados de origem e com maior realismo (LANGGUTH et al., 2021). Essa constante evolução implica que os modelos de detecção treinados com base em artefatos de uma geração específica de *deepfakes* podem se tornar rapidamente obsoletos à medida que novas arquiteturas e técnicas de treinamento de GANs surgem.

2.5.2 Generalização de Modelos para Tipos de Fraude Não Vistos

A capacidade de um modelo de detecção generalizar para tipos de manipulação ou fontes de dados que não foram explicitamente incluídos em seu conjunto de treinamento é crucial para sua aplicabilidade no mundo real. Muitas abordagens de detecção, especialmente aquelas que aprendem a identificar artefatos específicos de um determinado método de manipulação, podem apresentar uma queda significativa de desempenho quando confrontadas com técnicas novas ou variações sutis de métodos conhecidos. Wu et al. (WU; CHEN; ZHOU, 2023) apontam que os métodos de classificação tradicionais podem ter dificuldades devido à definição relativa de "forjado" e "prístino", que varia de imagem para imagem. Sua abordagem FOCAL, baseada em aprendizado contrastivo imagem por imagem e agrupamento não supervisionado na inferência, visa justamente melhorar a generalização ao aprender representações mais fundamentais das anomalias forenses, em vez de se especializar em artefatos de um conjunto de treinamento específico.

Stehouwer et al. (STEHOUWER et al., 2019) também reconhecem a importância de lidar com a diversidade de tipos de falsificação facial, que incluem troca de identidade, troca de expressão, manipulação de atributos e síntese completa da face. Desenvolver um detector que seja robusto a todas essas variações, bem como a futuras técnicas ainda não concebidas, é um desafio considerável. A criação de *datasets* abrangentes e diversificados, como o DFFD proposto por eles (STEHOUWER et al., 2019) e outros como o FaceForensics++ (LANGGUTH et al., 2021), é um passo importante, mas a generalização para o "desconhecido" permanece uma área ativa de pesquisa. Nesse sentido, métodos como o proposto por Jin et al. (JIN; YE; CHEN, 2021), que se baseiam em sinais biológicos intrín-

secos (como a frequência cardíaca estimada por PPG), argumentam ter uma vantagem, pois a ausência ou inconsistência desses sinais em vídeos falsos pode ser uma característica mais universal e menos dependente do método específico de geração do *deepfake*. Os autores afirmam que sua abordagem "não se limita a um método de geração específico, não sendo afetada por métodos sintéticos e possuindo boa adaptabilidade" (JIN; YE; CHEN, 2021).

2.5.3 Robustez Contra Ataques Adversariais

Modelos de aprendizado profundo, incluindo aqueles usados para detecção de *deepfakes*, são conhecidos por sua vulnerabilidade a ataques adversariais. Um ataque adversarial consiste na introdução de perturbações mínimas e frequentemente imperceptíveis ao olho humano em uma entrada (por exemplo, uma imagem ou vídeo), com o objetivo de enganar o modelo e induzi-lo a uma classificação incorreta (LANGGUTH et al., 2021). No contexto da detecção de *deepfakes*, um ator mal-intencionado poderia aplicar tais perturbações a um vídeo falso para que ele seja classificado como autêntico por um detector.

Langguth et al. (LANGGUTH et al., 2021) discutem essa vulnerabilidade, citando pesquisas que demonstram que mesmo os melhores detectores de *deepfake* podem ser enganados por estratégias adversariais. A existência de "exemplos adversariais" é uma característica bem documentada de redes neurais profundas. O problema é exacerbado se o atacante tiver acesso, mesmo que parcial (caixa-cinza) ou apenas à capacidade de consultar (caixa-preta), o modelo de detecção. Com tal acesso, o atacante pode iterativamente refinar as perturbações para maximizar a chance de evasão. Langguth et al. (LANGGUTH et al., 2021) traçam um paralelo com a detecção de *malware*, onde os autores de *malware* alteram continuamente seus códigos para evitar a detecção por software antivírus. Da mesma forma, a disponibilidade pública de um detector de *deepfakes* pode permitir que os criadores de conteúdo manipulado o utilizem para testar e "blindar" suas criações contra ele, tornando o detector menos eficaz na prática (LANGGUTH et al., 2021). Desenvolver defesas robustas contra ataques adversariais, ou criar detectores que sejam inerentemente menos suscetíveis a eles, é um campo de pesquisa ativo e essencial para a confiabilidade a longo prazo das tecnologias de detecção.

2.5.4 Interpretabilidade dos Modelos de Deep Learning

Muitos modelos de aprendizado profundo, especialmente as redes neurais com múltiplas camadas, operam de maneira que é frequentemente descrita como uma "caixa-preta" (*black box*). Embora possam alcançar alta precisão em suas previsões, o processo exato pelo qual chegam a uma determinada decisão pode não ser transparente ou facilmente compreensível para um observador humano. Essa falta de interpretabilidade representa

um desafio significativo, particularmente em domínios críticos como a análise forense, onde a capacidade de explicar o raciocínio por trás de uma conclusão é muitas vezes tão importante quanto a própria conclusão.

No contexto da detecção de fraudes, poder explicar por que um modelo classificou uma imagem como manipulada, quais características específicas ou regiões da imagem levaram a essa decisão, é crucial para a validação do resultado, para a identificação de possíveis vieses no modelo e para a sua aceitação em contextos legais ou investigativos. Abordagens que incorporam mecanismos de atenção, como a proposta por Stehouwer et al. (STEHOUWER et al., 2019), buscam mitigar esse problema ao gerar mapas de atenção que destacam as regiões da imagem que mais contribuíram para a decisão da rede. Esses mapas fornecem uma forma de visualização e explicação, indicando, por exemplo, que uma face foi considerada manipulada devido a anomalias detectadas na região dos olhos ou da boca (STEHOUWER et al., 2019). No entanto, mesmo com esses mecanismos, a interpretabilidade completa do funcionamento interno de redes complexas permanece um objetivo de pesquisa. A confiança em sistemas automatizados de detecção de fraude depende não apenas de sua precisão, mas também da capacidade de entender e auditar suas decisões.

2.5.5 Questões Éticas, Legais e de Privacidade

A disseminação de *deepfakes* e outras manipulações digitais levanta profundas questões éticas, legais e de privacidade. Langguth et al. (LANGGUTH et al., 2021) ressaltam que, embora muitas aplicações de *deepfakes* possam ser humorísticas ou benignas, existe um potencial inerente para uso malicioso. Exemplos incluem a criação de pornografia não consensual (onde o rosto de uma pessoa é inserido em conteúdo explícito), *cyberbullying*, difamação e a disseminação de desinformação com o objetivo de influenciar a opinião pública ou processos políticos (LANGGUTH et al., 2021). O caso de vídeos falsos do líder do partido BJP em Delhi, que foram compartilhados em milhares de grupos de WhatsApp, ilustra o potencial de *deepfakes* para enganar eleitores (LANGGUTH et al., 2021).

Do ponto de vista legal, a resposta ao fenômeno dos *deepfakes* é complexa. Proibir completamente a tecnologia é problemático, pois inviabilizaria usos legítimos em entretenimento, sátira e arte. Por outro lado, a ausência de regulamentação pode facilitar abusos. Langguth et al. (LANGGUTH et al., 2021) discutem a pressão legislativa nos Estados Unidos para restringir o uso prejudicial de *deepfakes*, mas também alertam para os riscos de leis excessivamente amplas levarem à autocensura por parte das plataformas de mídia social. Uma alternativa seria aplicar leis existentes contra fraude, assédio e difamação aos casos de uso malicioso de *deepfakes* (LANGGUTH et al., 2021).

Questões de privacidade também são proeminentes. O treinamento de modelos

para gerar *deepfakes* muitas vezes requer grandes quantidades de imagens faciais, que podem ser coletadas sem o consentimento explícito dos indivíduos. Da mesma forma, os *datasets* usados para treinar detectores de *deepfakes* também podem conter imagens de pessoas que não consentiram com esse uso específico de seus dados biométricos. Encontrar um equilíbrio entre a necessidade de dados para pesquisa e desenvolvimento de contramedidas e a proteção da privacidade individual é um desafio contínuo.

Finalmente, Langguth et al. ([LANGGUTH et al., 2021](#)) sugerem que, dadas as limitações das contramedidas tecnológicas e legais, as sociedades precisarão desenvolver uma maior resiliência à desinformação visual. Isso implica em fomentar a alfabetização midiática e o pensamento crítico, para que os cidadãos estejam mais aptos a questionar a autenticidade do conteúdo que consomem e menos suscetíveis a serem enganados por manipulações, sejam elas *deepfakes* ou outras formas de "notícias falsas".

3 Metodologia de Desenvolvimento da Solução de Prova de Vida (*Liveness Detection*)

A crescente digitalização de serviços e transações financeiras tem impulsionado a necessidade por sistemas de autenticação de identidade cada vez mais robustos e seguros. Nesse contexto, a biometria facial surge como uma alternativa promissora, oferecendo um método de verificação conveniente e intuitivo. Contudo, a simples detecção de uma face não é suficiente para garantir a presença real de um indivíduo, dada a proliferação de ataques de apresentação, como o uso de fotografias, vídeos ou máscaras tridimensionais. Para mitigar tais vulnerabilidades, as técnicas de Prova de Vida, ou *Liveness Detection*, tornam-se indispensáveis, visando atestar a presença genuína e em tempo real do usuário perante o sensor de captura.

Este capítulo detalha a metodologia empregada no desenvolvimento de um protótipo de solução de prova de vida. O trabalho concentrou-se na criação de uma sequência lógica de análises e desafios interativos, utilizando a linguagem de programação Python e predominantemente as bibliotecas OpenCV e MediaPipe. A abordagem metodológica percorre desde a aquisição e o pré-processamento de imagens faciais, passando pela extração de características biométricas dinâmicas, como a posição da cabeça e o comportamento ocular, até a integração de diferentes módulos para uma decisão final sobre a vivacidade do usuário. As seções subsequentes descreverão cada etapa do desenvolvimento, as técnicas exploradas, as estratégias adotadas e, quando pertinente, as fundamentações matemáticas que embasam as funcionalidades implementadas. O foco reside no processo de construção e nas decisões de projeto que moldaram o protótipo, preparando o terreno para a análise dos resultados que será apresentada em um capítulo posterior.

3.1 Introdução à Arquitetura do Sistema

A concepção de um sistema de prova de vida eficaz requer uma arquitetura bem definida, capaz de orquestrar múltiplos processos de análise de forma coesa e eficiente. Nesta seção, apresentamos a arquitetura geral do protótipo desenvolvido, detalhando seus principais módulos componentes, o fluxo de processamento das informações e as tecnologias fundamentais que sustentam sua implementação. O objetivo é fornecer uma visão macroscópica do sistema, antes de adentrar nas especificidades de cada técnica de análise de vivacidade nas seções subsequentes.

3.1.1 Visão Geral do Protótipo e Seus Módulos Componentes

O protótipo de prova de vida foi projetado como um sistema modular, onde cada módulo é responsável por uma etapa específica da análise de vivacidade. Esta modularidade facilita o desenvolvimento, teste e futuras expansões do sistema.

O sistema inicia sua operação com o módulo de **Aquisição de Imagem**, responsável por capturar o fluxo de vídeo da câmera do dispositivo. Subsequentemente, o módulo de **Pré-processamento e Validação Facial** entra em ação, realizando a detecção de faces na imagem e assegurando que apenas uma face esteja presente no campo de visão para análise, uma funcionalidade crucial para evitar ambiguidades e garantir a correta associação dos dados biométricos ao usuário. Esta etapa utiliza algoritmos para identificar a região facial e preparar os dados para as análises subsequentes.

Uma vez validada a presença de uma única face, o módulo de **Análise Facial Estática e Dinâmica** é acionado. Este é um dos núcleos do sistema, encarregado de extrair pontos fiduciais da face utilizando tecnologias como o MediaPipe Face Mesh ([LUGARESI et al., 2019](#)). Estes pontos são cruciais para as análises subsequentes, incluindo:

- **Estimativa de Posição da Cabeça:** Determina a orientação tridimensional da cabeça (angulações de *yaw*, *pitch* e *roll*), permitindo não apenas guiar o usuário a manter uma pose adequada, mas também servindo como base para desafios de movimento.
- **Análise Ocular Detalhada:** Engloba a detecção e o rastreamento da íris, a detecção de piscadas (baseada, por exemplo, no *Eye Aspect Ratio* - EAR) e a estimativa da direção do olhar. Esses comportamentos oculares são indicadores importantes de vivacidade.

Paralelamente ou em sequência, o módulo de **Mecanismos Adicionais de Anti-spoofing** aplica técnicas complementares para dificultar ataques de apresentação. Um exemplo implementado é a análise de cor da esclera, que busca diferenciar um olho real de uma representação artificial, conforme disparos luminosos coloridos na tela do usuário.

Opcionalmente, um **Motor de Desafios de Vivacidade** pode ser ativado, solicitando ao usuário a execução de ações específicas, como movimentos de cabeça ou sequências de piscadas, cujas respostas são validadas pelos módulos de análise correspondentes.

Finalmente, o módulo de **Integração e Tomada de Decisão**, refletido conceitualmente no script `pilot.py`, agrega as evidências coletadas por todos os módulos anteriores. Este componente aplica uma lógica de decisão, que pode ser baseada em um sistema de pontuação ou em regras predefinidas, para emitir um veredito final sobre a autenticidade e vivacidade do usuário.

Esta arquitetura modular e multifacetada visa criar um sistema de prova de vida resiliente, capaz de analisar diversos aspectos do comportamento humano para uma autenticação segura.

3.1.2 Fluxograma Geral do Processo de Liveness Detection

O processo de verificação de vivacidade implementado no protótipo segue um fluxo lógico de operações, projetado para maximizar a acurácia e a eficiência da análise. A Figura 4 apresenta um fluxograma detalhado deste processo.

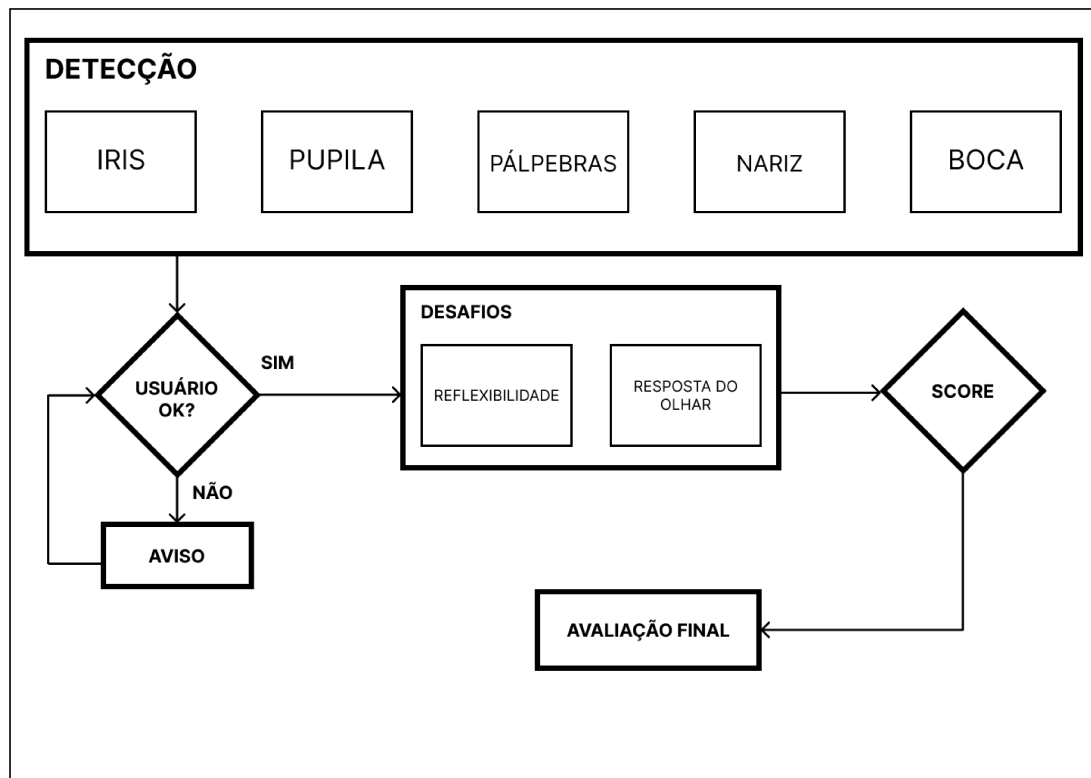


Figura 4 – Fluxograma geral do processo de verificação de vivacidade.

A sequência de operações inicia-se com a **Captura do Fluxo de Vídeo** (1) proveniente da webcam ou outro dispositivo de imagem. Cada quadro (*frame*) do vídeo é então submetido à etapa de **Detecção de Faces** (2). Nesta fase, o sistema verifica não apenas a presença de uma face, mas também se certifica de que há uma única face detectada, garantindo que a análise subsequente seja focada no usuário correto.

Se uma única face é identificada, o sistema prossegue para a **Extração de Pontos Fiduciais Faciais** (3). Utilizando a solução MediaPipe Face Mesh (LUGARESI et al., 2019), uma malha densa de *landmarks* 3D é sobreposta à face detectada. Estes pontos são a base para a maioria das análises de vivacidade.

Com os *landmarks* disponíveis, diversas análises podem ocorrer em paralelo ou em uma sequência otimizada (4):

- **Análise da Posição da Cabeça:** Os ângulos de *yaw*, *pitch* e *roll* são calculados. Se o sistema incluir uma fase de **Guia de Posicionamento** (4a), mensagens podem ser exibidas ao usuário para ajustar sua pose antes de prosseguir com os testes de vivacidade.
- **Análise Ocular:** Inclui o rastreamento da íris, a detecção de piscadas (e.g., via cálculo do EAR) e, potencialmente, a estimativa da direção do olhar.
- **Verificações Antispoofing Iniciais:** Técnicas como a análise da cor da esclera foram aplicadas nesta fase.

Dependendo da configuração do sistema, a etapa de **Apresentação de Desafios Interativos** (5) pode ser iniciada. Estes desafios podem incluir solicitar movimentos específicos da cabeça (e.g., virar para a esquerda/direita, assentir) ou ações oculares (e.g., piscar X vezes, seguir um ponto na tela).

A **Análise das Respostas aos Desafios** (6) é realizada pelos respectivos módulos (posição da cabeça, análise ocular), que verificam se as ações do usuário correspondem ao esperado.

Em conjunto com os desafios, ou como uma camada contínua de verificação, **Mecanismos Antispoofing Adicionais** (7) são empregados. Estes podem incluir a busca por indícios de apresentação de fotos ou vídeos (como reflexos anormais ou falta de micro-movimentos), e a avaliação de pistas de profundidade para diferenciar rostos 2D de rostos 3D reais.

Finalmente, todas as evidências coletadas (qualidade do posicionamento, piscadas, respostas aos desafios, resultados dos testes antispoofing) são enviadas ao módulo de **Agregação e Decisão Final** (8). Este módulo utiliza uma lógica predefinida, um sistema de pontuação simples, para determinar se o usuário é genuíno. O resultado é uma **Saída da Decisão de Vivacidade** (9) (aprovado/reprovado).

Este fluxo estruturado permite uma análise compreensiva e robusta, onde múltiplas características são avaliadas para formar uma conclusão confiável sobre a vivacidade do usuário.

3.1.3 Principais Tecnologias e Bibliotecas Empregadas

O desenvolvimento do protótipo de prova de vida foi realizado primariamente na linguagem de programação Python, devido à sua vasta gama de bibliotecas especializadas em visão computacional e aprendizado de máquina, bem como sua flexibilidade e rápida prototipagem. Dentre o ecossistema de ferramentas disponíveis, duas bibliotecas se destacaram como pilares fundamentais do projeto: OpenCV e MediaPipe.

O **OpenCV (Open Source Computer Vision Library)** ([BRADSKI, 2000](#)) foi extensivamente utilizado para operações basilares de processamento de imagem e vídeo. Suas funcionalidades incluem a captura de *streams* de vídeo da webcam, manipulação de matrizes de pixels (representando imagens), conversões de espaço de cores, desenho de elementos gráficos sobre as imagens (como retângulos delimitadores, pontos fiduciais e mensagens de texto para o usuário), e a implementação de algoritmos clássicos de visão computacional. Embora o MediaPipe ofereça suas próprias soluções para detecção facial, o OpenCV também provê alternativas robustas, como os classificadores baseados em Haar Cascades, que podem ser utilizados em etapas iniciais ou complementares de detecção. A versatilidade do OpenCV o torna uma ferramenta indispensável para a manipulação de dados visuais em tempo real.

A biblioteca **MediaPipe** ([LUGARESI et al., 2019](#)), desenvolvida pela Google, representou um avanço significativo para a implementação de funcionalidades complexas de percepção computacional. Em particular, a solução *MediaPipe Face Mesh* foi crucial para este projeto, fornecendo uma detecção robusta e em tempo real de uma malha tridimensional de 468 *landmarks* faciais (ou 478 para versões mais recentes). Estes pontos fiduciais detalhados são a espinha dorsal para diversas análises, incluindo a estimativa precisa da pose da cabeça, o rastreamento dos olhos e da íris, e a detecção de expressões faciais sutis como o piscar dos olhos. A capacidade do MediaPipe de executar tais tarefas com alta performance, mesmo em dispositivos com recursos computacionais limitados, foi determinante para a viabilidade do protótipo. Além do *Face Mesh*, outras soluções do MediaPipe, como as dedicadas ao rastreamento de mãos ou pose corporal, demonstram o potencial da ferramenta para futuras expansões do sistema de prova de vida com novos desafios interativos.

Em conjunto com estas duas bibliotecas principais, o **NumPy (Numerical Python)** ([HARRIS et al., 2020](#)) foi implicitamente e explicitamente utilizado para operações numéricas eficientes, especialmente na manipulação de vetores e matrizes que representam coordenadas de *landmarks*, transformações geométricas e outros dados numéricos inerentes ao processamento de imagens e à análise biométrica.

A escolha estratégica destas tecnologias permitiu a construção de um protótipo com capacidades avançadas de análise facial, aproveitando soluções otimizadas e amplamente testadas pela comunidade científica e pela indústria, ao mesmo tempo em que oferece flexibilidade para a customização e implementação de lógicas específicas para a detecção de vivacidade.

Tabela 1 – Principais tecnologias e bibliotecas utilizadas no desenvolvimento do protótipo.

Tecnologia/Biblioteca	Versão	Principal Utilização no Projeto
Python	3.9+	Linguagem principal de desenvolvimento
OpenCV	4.5+	Processamento de imagem/vídeo, captura de câmera, desenho de HUD
MediaPipe	0.8+	Deteção facial, extração de landmarks faciais (Face Mesh), rastreamento de íris
NumPy	1.20+	Operações numéricas eficientes com arrays e matrizes

3.2 Aquisição e Pré-processamento de Imagens Faciais

A etapa inicial de qualquer sistema de visão computacional, e particularmente crucial em aplicações biométricas como a prova de vida, consiste na aquisição e no pré-processamento adequados das imagens. A qualidade dos dados de entrada tem um impacto direto e substancial na performance e na confiabilidade de todas as análises subsequentes, desde a detecção facial até a extração de características biométricas complexas. Nesta seção, são detalhados os procedimentos adotados no protótipo para a captura de frames de vídeo, a validação da presença de uma face única e as considerações sobre os atributos de qualidade da imagem que asseguram uma análise robusta.

3.2.1 Captura e Seleção de Frames de Vídeo

A interface primária do sistema com o mundo real ocorre através da captura de imagens provenientes de um dispositivo de vídeo, sendo uma webcam integrada ou conectada ao computador do usuário, ou a câmera do *smartphone* virtualizada no computador. O acesso ao fluxo de vídeo é gerenciado por meio de bibliotecas de software especializadas, que abstraem a complexidade da comunicação com o hardware da câmera e fornecem um fluxo contínuo de quadros (*frames*) de imagem.

No contexto deste protótipo, cada *frame* capturado é tratado como uma imagem individual a ser processada. A taxa de captura de *frames* (frames por segundo - FPS) é uma consideração importante; taxas mais elevadas podem fornecer uma maior fluidez na análise de movimentos dinâmicos, mas também impõem uma carga computacional maior. Para aplicações em tempo real, como a que se propõe, busca-se um equilíbrio entre a capacidade de processamento disponível e a necessidade de uma resposta rápida do sistema. O sistema foi projetado para processar os *frames* à medida que são disponibilizados pela câmera, visando uma interação imediata com o usuário.

Após a aquisição, os *frames* podem passar por conversões preliminares de formato ou espaço de cores, conforme exigido pelas bibliotecas ou algoritmos subsequentes. Por

exemplo, é comum a conversão do formato BGR (Blue-Green-Red), frequentemente fornecido por padrão em algumas bibliotecas de captura, para o formato RGB (Red-Green-Blue), que é o esperado por outras ferramentas de análise facial, como certas configurações do MediaPipe. Tais transformações asseguram a compatibilidade e a otimização do processamento nas etapas seguintes. Nenhuma estratégia complexa de seleção ou amostragem de *frames* foi implementada nesta fase inicial, optando-se por analisar o fluxo de vídeo de forma contínua para não perder eventos relevantes de vivacidade.

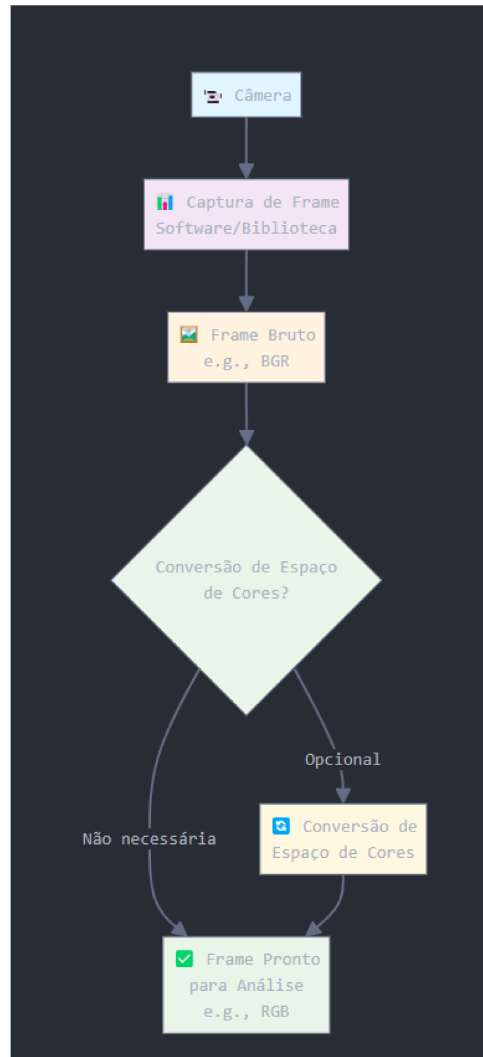


Figura 5 – Fluxo simplificado da captura e preparação inicial de frames de vídeo.

3.2.2 Detecção e Validação da Presença de Face Única

Uma vez que um *frame* de vídeo é capturado e preparado, a primeira tarefa analítica consiste em localizar a face do usuário na imagem. Esta etapa é fundamental, pois delimita a região de interesse (ROI) para todas as operações subsequentes de extração de características e análise de vivacidade. Para a detecção facial inicial, podem ser empregados diversos algoritmos, sendo os classificadores em cascata baseados em características

Haar (VIOLA; JONES, 2001) uma abordagem clássica e eficiente para detecção de objetos em tempo real, e cuja estrutura é compatível com os modelos de detecção facial preliminar que podem ser utilizados.

Além da simples detecção, é crucial para a integridade do sistema de prova de vida garantir que apenas uma face esteja presente e sendo analisada. A presença de múltiplas faces na cena poderia levar a ambiguidades, como a extração de características de um indivíduo que não é o usuário pretendido, ou mesmo tentativas de fraude onde uma pessoa tenta autenticar outra. Para mitigar este risco, o sistema incorpora uma lógica de validação que conta o número de faces detectadas no *frame*.



(a) Detecção bem-sucedida de face única.



(b) Detecção de múltiplas faces.

Figura 6 – Exemplos de cenários na etapa de detecção facial: (a) detecção bem-sucedida de face única e (b) detecção de múltiplas faces.

Se nenhuma face for detectada, o sistema pode instruir o usuário a se posicionar corretamente em frente à câmera ou verificar as condições de iluminação. Caso mais de uma face seja identificada, uma mensagem de alerta pode ser exibida, solicitando que apenas o usuário principal permaneça no campo de visão da câmera. O processamento só avança para as etapas mais complexas de análise de vivacidade quando uma, e somente uma, face é consistentemente detectada na imagem. Esta validação contribui para a robustez e a segurança do processo de autenticação, assegurando que o foco da análise biométrica recaia exclusivamente sobre o indivíduo correto.

3.2.3 Considerações sobre Qualidade da Imagem para Análise Robusta

A performance de algoritmos de análise facial, incluindo a extração de pontos fiduciais e a subsequente interpretação de movimentos e texturas, é intrinsecamente dependente da qualidade da imagem de entrada. Variações em condições de aquisição podem degradar significativamente a precisão dessas análises, levando a falsas rejeições ou, em

cenários piores, a uma menor capacidade de detecção de fraudes. Portanto, algumas considerações sobre a qualidade da imagem são essenciais.

A **iluminação** é, talvez, o fator mais crítico. Uma iluminação insuficiente resulta em imagens escuras, com perda de detalhes nas características faciais. Por outro lado, uma iluminação excessiva ou mal distribuída pode causar superexposição em certas áreas do rosto ou gerar sombras profundas que obscurecem *landmarks* importantes. O ideal é uma iluminação frontal, difusa e uniforme, que minimize sombras e realce as feições do usuário de forma clara.

A **resolução** da imagem também desempenha um papel importante. Imagens de baixa resolução podem não conter detalhes suficientes para uma localização precisa dos pontos fiduciais, especialmente para características menores como os cantos dos olhos ou da boca. Embora algoritmos modernos sejam cada vez mais robustos a variações de resolução, existe um limiar mínimo abaixo do qual a performance decai consideravelmente.

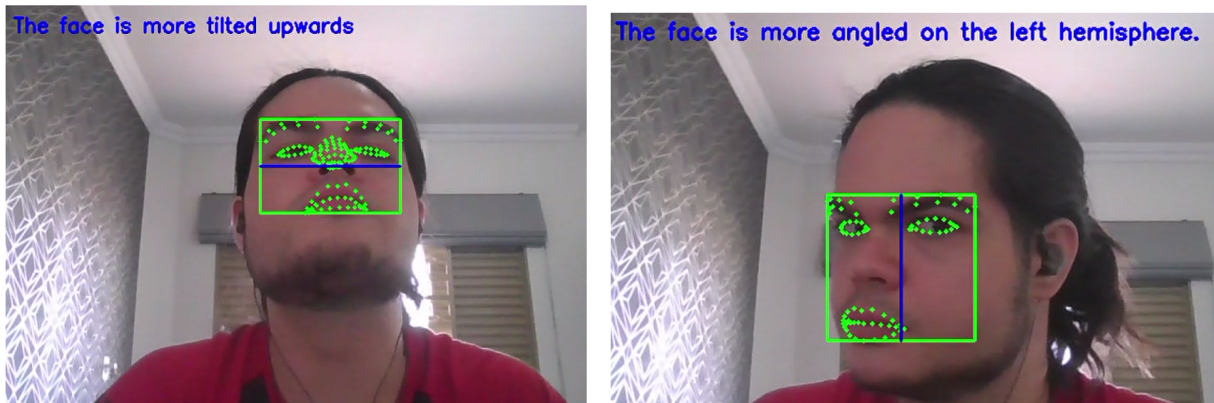
O **foco** da imagem deve ser nítido na face do usuário. Imagens desfocadas perdem detalhes de alta frequência que são importantes para a análise de textura e para a precisa localização de contornos faciais e *landmarks*.

Outros fatores como **ruído** excessivo na imagem, geralmente associado a sensores de baixa qualidade ou condições de pouca luz, e a **pose** excessivamente angulada da cabeça, também podem comprometer a análise. Embora o sistema de estimativa de pose da cabeça (a ser detalhado adiante) possa lidar com certas variações, poses extremas dificultam a visibilidade de todas as características faciais necessárias.

Fator de Qualidade	Condição Ideal	Impacto Negativo da Má Qualidade
Iluminação	Uniforme, difusa, sem sombras fortes ou superexposição	Perda de detalhes, erros na detecção de landmarks, inviabiliza análise de textura.
Resolução	Suficiente para distinguir feições e landmarks (e.g., VGA ou superior)	Perda de precisão na localização de landmarks, dificuldade em análises finas.
Foco	Nítido sobre a face do usuário	Perda de detalhes de alta frequência, contornos imprecisos, afeta análise de textura.
Ruído	Mínimo possível	Pode introduzir artefatos, confundir detectores de características.
Pose Facial	Frontal ou levemente angulada	Poses extremas ocultam landmarks e dificultam a normalização.

Para mitigar alguns desses problemas, o protótipo pode, idealmente, incorporar

mecanismos de *feedback* em tempo real, instruindo o usuário a melhorar as condições de captura (e.g., "Mova-se para um local mais iluminado", "Aproxime-se da câmera", "Centralize seu rosto"). Tais interações proativas não apenas melhoram a qualidade dos dados de entrada, mas também a experiência do usuário, guiando-o para uma interação bem-sucedida com o sistema de prova de vida. A atenção a esses fatores durante a fase de aquisição é um investimento que se reflete na robustez e confiabilidade de todo o processo de verificação.



(a) Detecção de angulação vertical do rosto. (b) Detecção de angulação horizontal do rosto.

Figura 7 – Exemplos de *feedback* em tempo real ao usuário: (a) detecção de angulação vertical do rosto e (b) detecção de angulação horizontal do rosto.

3.3 Análise Facial Estática e Dinâmica

Superada a fase de aquisição e validação da imagem facial, o sistema avança para uma análise mais aprofundada da face do usuário. Esta etapa é fundamental, pois transforma os dados brutos de pixels em representações estruturadas e significativas, permitindo a extração de características biométricas tanto estáticas quanto dinâmicas. A análise facial detalhada serve como alicerce para as subseqüentes verificações de vivacidade, como a estimativa da pose da cabeça, o rastreamento ocular e a detecção de expressões. Nesta seção, são descritos os métodos empregados para a detecção inicial da face, com foco na sua função de entrada para a extração de pontos fiduciais, e o processo de mapeamento desses pontos, que constitui a base para a compreensão da geometria e da dinâmica facial.

3.3.1 Detecção Facial de Precisão para Mapeamento

Conforme discutido na Seção 3.2.2, a identificação da presença e da localização de uma única face na imagem é um pré-requisito. Uma vez que a região facial principal é isolada, esta informação, na forma de uma caixa delimitadora (*bounding box*), serve como entrada para algoritmos mais refinados de análise facial. Embora a detecção inicial possa ser realizada por métodos como classificadores em cascata (VIOLA; JONES, 2001),

muitas soluções avançadas de percepção facial, a exemplo do MediaPipe (LUGARESI et al., 2019), integram seus próprios detectores de face de alta performance, otimizados para fornecer uma localização precisa que beneficie diretamente a etapa subsequente de extração de pontos fiduciais.

A precisão desta detecção inicial é vital, pois um erro na delimitação da face pode propagar-se e comprometer a exatidão do mapeamento dos pontos fiduciais. Portanto, o sistema busca empregar ou confiar em detectores que sejam robustos a variações comuns de iluminação, pequenas oclusões e diferentes tamanhos de face na imagem, garantindo que a etapa de extração de *landmarks* receba a região de interesse mais acurada possível.

3.3.2 Extração e Mapeamento de Pontos Fiduciais da Face

A extração de pontos fiduciais, ou *landmarks* faciais, é uma das pedras angulares da análise facial moderna e do presente protótipo de prova de vida. Esses *landmarks* são um conjunto de pontos anatomicamente significativos na face, como os cantos dos olhos, a ponta do nariz, os contornos dos lábios e da mandíbula, e as sobrancelhas. O mapeamento preciso desses pontos permite uma representação rica e detalhada da geometria facial do indivíduo.

Para esta tarefa crucial, o protótipo emprega a solução *Face Mesh* disponibilizada pela biblioteca MediaPipe (LUGARESI et al., 2019). Esta tecnologia é capaz de detectar e rastrear, em tempo real, uma malha densa composta por 468 *landmarks* tridimensionais na face humana. Cada *landmark* possui coordenadas (x, y, z) , onde (x, y) representam as coordenadas de pixel na imagem e z fornece uma estimativa da profundidade relativa do ponto em relação ao centro geométrico da cabeça, com o valor de z sendo menor quanto mais próximo da câmera o ponto estiver. Esta capacidade de inferir uma terceira dimensão, mesmo que relativa, a partir de uma imagem 2D, enriquece significativamente a análise. A Figura 8 ilustra a densidade e a distribuição desses pontos sobre uma face.

A robustez do *MediaPipe Face Mesh* frente a variações de pose, oclusões parciais (como o uso de óculos), diferentes condições de iluminação e expressões faciais dinâmicas é um dos seus principais trunfos. Essa capacidade garante que os *landmarks* sejam extraídos de forma consistente, o que é essencial para a confiabilidade das análises de vivacidade subsequentes.

Os *landmarks* faciais extraídos são a base para uma vasta gama de cálculos e inferências:

- **Estimativa da Posição da Cabeça (Head Pose Estimation):** A posição tridimensional de vários *landmarks* permite calcular os ângulos de rotação da cabeça (guinada, inclinação e rotação axial – *yaw*, *pitch*, *roll*).



Figura 8 – Exemplo da malha de 468 pontos fiduciais faciais (Face Mesh) detectada pela biblioteca MediaPipe sobre uma face humana.

- **Análise Ocular:** *Landmarks* específicos ao redor dos olhos são usados para calcular métricas como o *Eye Aspect Ratio* (EAR) para detecção de piscadas, e para localizar com precisão as pupilas ou íris para rastreamento do olhar.
- **Análise de Expressões Faciais:** Embora não seja o foco principal deste trabalho para a prova de vida, a deformação da malha facial permite a identificação de expressões como sorriso, surpresa, entre outras, que podem ser exploradas em cenários de interação mais complexos.
- **Definição de Regiões de Interesse (ROIs) Detalhadas:** Permitem isolar com precisão regiões como a boca, o nariz ou os olhos para análises específicas.

A manipulação das coordenadas desses pontos, incluindo o cálculo de distâncias euclidianas entre pares de *landmarks*, ângulos formados por trios de pontos, ou a normalização dessas coordenadas para mitigar efeitos de escala e translação da face na imagem, é realizada por módulos de processamento geométrico. Essas operações matemáticas sobre a nuvem de pontos fiduciais são essenciais para derivar as métricas quantitativas que alimentam os algoritmos de decisão de vivacidade. A estabilidade relativa de certas distâncias interoculares, por exemplo, pode ser contrastada com a variabilidade das distâncias ao redor da boca durante a fala, fornecendo pistas sobre a autenticidade da face.

Em suma, a extração detalhada e robusta dos pontos fiduciais da face é um passo transformador no fluxo de processamento, convertendo uma imagem facial em um conjunto rico de dados geométricos prontos para sustentar as diversas técnicas de verificação de vivacidade exploradas neste projeto.

3.4 Estimativa e Monitoramento da Posição da Cabeça

A análise da posição e dos movimentos da cabeça do usuário desempenha um papel multifacetado em sistemas de prova de vida. Primeiramente, assegura que o usuário esteja adequadamente posicionado em frente à câmera, otimizando a qualidade da captura para outras análises biométricas. Em segundo lugar, a capacidade de rastrear a orientação da cabeça em tempo real permite a implementação de desafios interativos baseados em movimento, que são indicadores robustos de vivacidade. Finalmente, a detecção de movimentos cefálicos naturais pode, por si só, contribuir para diferenciar um usuário real de uma representação estática, como uma fotografia. Esta seção descreve a metodologia empregada para determinar os ângulos de rotação da cabeça e como essa informação é utilizada para classificar a orientação facial e interagir com o usuário, fundamentando-se nos pontos fiduciais extraídos conforme detalhado na Seção 3.3.2.

3.4.1 Determinação dos Ângulos de Euler (Yaw, Pitch, Roll)

A orientação tridimensional da cabeça pode ser descrita por três rotações em torno dos eixos principais, conhecidas como ângulos de Euler: guinada (*yaw*, ψ), inclinação (*pitch*, θ) e rotação axial (*roll*, ϕ). A Figura 9 ilustra esses eixos de rotação. A guinada representa a rotação da cabeça para a esquerda ou direita, a inclinação descreve o movimento de abaixar ou levantar a cabeça, e a rotação axial corresponde à inclinação lateral da cabeça.

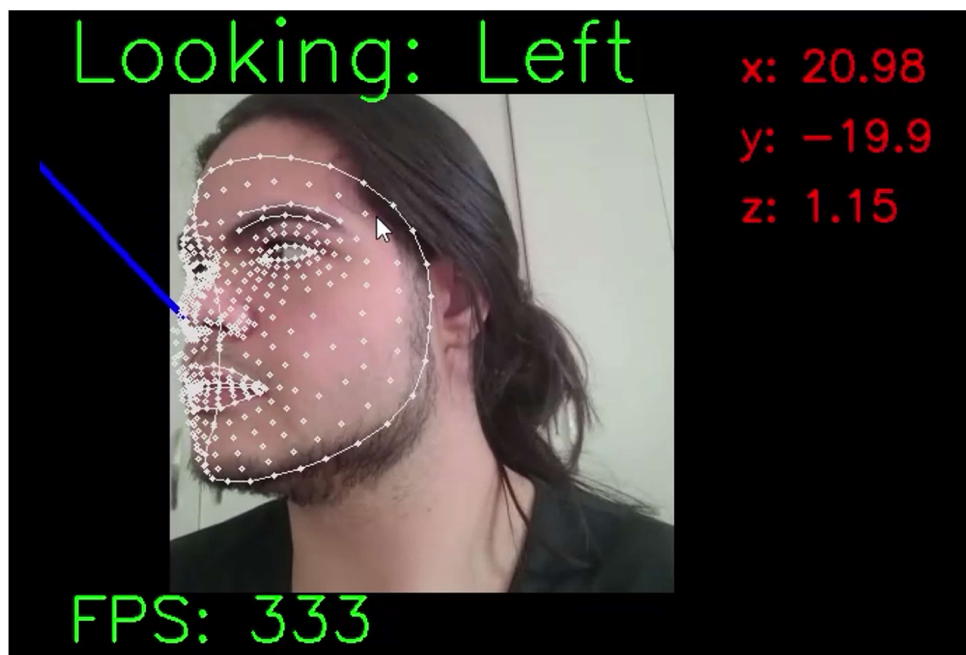


Figura 9 – Ilustração dos eixos de rotação da cabeça: Guinada (ψ , Yaw), Inclinação (θ , Pitch) e Rotação Axial (ϕ , Roll).

A estimativa desses ângulos a partir de uma imagem 2D é um problema clássico

em visão computacional, frequentemente abordado através de algoritmos que resolvem o problema *Perspective-n-Point* (PnP) (MURPHY-CHUTORIAN; TRIVEDI, 2009). O algoritmo PnP calcula a pose (posição e orientação) de um objeto tridimensional em relação a uma câmera, dadas as coordenadas 2D de n pontos do objeto na imagem e as coordenadas 3D correspondentes desses mesmos pontos em um sistema de referência do objeto.

No contexto deste protótipo, os insumos para o algoritmo PnP são:

1. **Pontos 2D na Imagem:** Um subconjunto selecionado dos pontos fiduciais faciais (por exemplo, ponta do nariz, queixo, cantos internos e externos dos olhos, cantos da boca) extraídos pela solução *MediaPipe Face Mesh*, conforme descrito na Seção 3.3.2.
2. **Modelo 3D Canônico da Cabeça:** As coordenadas 3D correspondentes aos *landmarks* 2D selecionados, definidas em um sistema de coordenadas genérico de uma cabeça humana. Este modelo 3D não precisa ser exato para o usuário específico, pois um modelo antropometricamente médio geralmente é suficiente para estimar a orientação relativa.
3. **Parâmetros Intrínsecos da Câmera:** Incluem a distância focal (f_x, f_y) e o ponto principal (c_x, c_y) da câmera. Em uma implementação ideal, esses parâmetros são obtidos através de um processo de calibração da câmera. Na ausência de calibração específica, podem ser utilizados valores aproximados, como considerar a distância focal igual à largura da imagem e o ponto principal no centro da imagem, embora isso possa introduzir imprecisões na estimativa absoluta da pose.

Dado esses dados, uma função como `cv2.solvePnP` da biblioteca OpenCV, que implementa algoritmos eficientes como o EPnP (LEPETIT; MORENO-NOGUER; FUA, 2009), pode ser utilizada. Esta função retorna um vetor de translação e um vetor de rotação. O vetor de rotação, que representa a orientação da cabeça no formato de Rodrigues (eixo-ângulo), pode ser convertido em uma matriz de rotação R de 3×3 . A partir desta matriz de rotação R , os ângulos de Euler (ψ, θ, ϕ) podem ser extraídos através de decomposições matriciais. Por exemplo, se a matriz de rotação R for definida como:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

Os ângulos de Euler podem ser calculados (assumindo uma ordem de rotação específica, e.g., YXZ):

$$\psi(\text{yaw}) = \text{atan2}(r_{13}, r_{33})$$

$$\theta(\text{pitch}) = \text{asin}(-r_{23})$$

$$\phi(\text{roll}) = \text{atan2}(r_{21}, r_{22})$$

É importante notar que a extração de ângulos de Euler de uma matriz de rotação pode sofrer de singularidades (conhecidas como *gimbal lock*) em certas configurações, mas para as faixas de movimento da cabeça humana em aplicações de prova de vida, geralmente são adequados. A precisão da estimativa depende da qualidade dos *landmarks* detectados, da acurácia do modelo 3D canônico e da exatidão dos parâmetros da câmera.

3.4.2 Classificação da Orientação da Cabeça para Interação com o Usuário

Uma vez que os ângulos de Euler (ψ, θ, ϕ) são determinados em tempo real para cada *frame* de vídeo, eles podem ser utilizados para diversas finalidades interativas e analíticas dentro do sistema de prova de vida. Um módulo de classificação de orientação da cabeça foi desenvolvido para interpretar esses ângulos contínuos e traduzi-los em estados discretos que facilitam a interação com o usuário e a lógica de decisão.

Primeiramente, essa classificação é empregada para **guiar o posicionamento do usuário**. Faixas angulares são definidas para caracterizar uma pose considerada "frontal" ou "ideal" para a análise (e.g., $|\psi| < 10^\circ$, $|\theta| < 10^\circ$, $|\phi| < 5^\circ$). Se a orientação da cabeça do usuário desviar significativamente dessas faixas, o sistema pode fornecer feedback visual ou textual, como "Vire o rosto para o centro", "Incline a cabeça ligeiramente para cima", ou "Mantenha a cabeça reta". Este mecanismo ajuda a garantir que as demais análises biométricas sejam realizadas sob condições de visualização facial ótimas.

Em segundo lugar, a classificação da orientação da cabeça é fundamental para a implementação de **desafios de vivacidade baseados em movimento cefálico**. O sistema pode solicitar ao usuário que execute movimentos específicos, como virar a cabeça para a esquerda, depois para a direita, assentir (movimento de *pitch*) ou negar (movimento de *yaw*). Para validar esses desafios, o módulo de classificação monitora se os ângulos de Euler atingem certas faixas predefinidas que correspondem ao movimento solicitado (e.g., para "virar à esquerda", o ângulo de *yaw* ψ deve exceder um limiar positivo, como $+20^\circ$). A sequência e a temporização corretas desses movimentos são então verificadas para atestar a vivacidade.

A definição dos limiares para cada classe de orientação (e.g., "Centro", "Esquerda", "Direita", "Para Cima", "Para Baixo") é uma etapa de configuração importante, que busca um equilíbrio entre a naturalidade do movimento para o usuário e a robustez na detecção da intenção do movimento. Um feedback claro e intuitivo ao usuário durante essas interações é crucial para o sucesso e a usabilidade do sistema de prova de vida. A capacidade de monitorar e classificar dinamicamente a pose da cabeça transforma uma simples captura de imagem em um processo interativo e mais seguro contra fraudes.

Tabela 2 – Exemplo de faixas angulares para classificação da orientação da cabeça.

Classificação da Pose	Yaw (ψ)	Pitch (θ)	Roll (ϕ)
Centro / Frontal	$[-10^\circ, +10^\circ]$	$[-10^\circ, +10^\circ]$	$[-5^\circ, +5^\circ]$
Virado(a) à Esquerda	$(+15^\circ, +45^\circ]$	$[-15^\circ, +15^\circ]$	$[-10^\circ, +10^\circ]$
Virado(a) à Direita	$[-45^\circ, -15^\circ)$	$[-15^\circ, +15^\circ]$	$[-10^\circ, +10^\circ]$
Inclinado(a) para Cima	$[-15^\circ, +15^\circ]$	$(+15^\circ, +30^\circ]$	$[-10^\circ, +10^\circ]$
Inclinado(a) para Baixo	$[-15^\circ, +15^\circ]$	$[-30^\circ, -15^\circ)$	$[-10^\circ, +10^\circ]$

3.5 Análise Ocular Detalhada para Prova de Vida

Os olhos são frequentemente descritos como "janelas da alma" e, no contexto da biometria e da prova de vida, eles são, sem dúvida, janelas para a autenticidade. A região ocular oferece uma riqueza de informações dinâmicas e comportamentais que são difíceis de serem replicadas por artefatos de fraude. Movimentos da íris, piscadas espontâneas e a capacidade de direcionar o olhar são características inerentemente humanas e vitais. Esta seção detalha as metodologias empregadas no protótipo para a detecção e rastreamento da íris, a identificação de piscadas e a análise da direção do olhar, todos fundamentados nos pontos fiduciais faciais previamente extraídos (Seção 3.3.2).

3.5.1 Detecção e Rastreamento da Íris

O movimento da íris é um forte indicador de vivacidade, pois olhos estáticos são uma característica comum em ataques de apresentação que utilizam fotografias ou vídeos simples. O rastreamento da posição da íris ao longo do tempo permite verificar a presença de micromovimentos sacádicos naturais ou validar respostas a desafios que envolvam o direcionamento do olhar.

O processo inicia-se com a localização precisa da região de cada olho, utilizando os *landmarks* faciais fornecidos pela solução *MediaPipe Face Mesh* (LUGARESI et al., 2019). Esses *landmarks* definem contornos precisos ao redor dos olhos, permitindo isolar Regiões de Interesse (ROIs) específicas para cada olho. Dentro dessas ROIs, algoritmos são aplicados para detectar o centro da pupila ou da íris. O MediaPipe, por si só, já oferece *landmarks* específicos para a íris, o que simplifica consideravelmente essa tarefa.

No entanto, para aumentar a robustez ou para explorar características específicas da íris, técnicas adicionais de processamento de imagem podem ser empregadas sobre a ROI ocular. Estas podem incluir a aplicação de filtros para realçar o contraste entre a pupila/íris e a esclera, seguida por algoritmos de detecção de círculos (como a Transformada de Hough) ou métodos baseados em contornos e ajuste de elipses para localizar o centro da pupila/íris com maior precisão. Módulos dedicados à detecção pupilar foram desenvolvidos para refinar essa localização. Em alguns casos, a criação de uma máscara da íris, isolando-a do restante do olho, pode auxiliar na análise de sua textura ou cor,



Figura 10 – Exemplo de localização da íris utilizando landmarks faciais para delimitar a ROI ocular e algoritmos de processamento de imagem para refinar a detecção do centro pupilar.

embora o foco principal aqui seja a sua posição.

Uma vez que o centro da íris é detectado em um *frame*, sua posição é rastreada ao longo dos *frames* subsequentes. Este rastreamento é crucial para analisar a dinâmica do movimento ocular. Desafios como reflexos na córnea ou nos óculos, baixa resolução da imagem na região dos olhos, ou oclusões parciais pelas pálpebras podem dificultar a detecção e o rastreamento contínuo, exigindo algoritmos robustos e, estratégias de recuperação em caso de perda temporária do alvo.

3.5.2 Detecção de Piscada de Olhos

A piscada é um reflexo fisiológico, em grande parte involuntário, essencial para a lubrificação e proteção dos olhos. A presença, frequência e naturalidade das piscadas são fortes indicadores de vivacidade, pois sua ausência ou um padrão de piscada artificial são difíceis de simular realisticamente por longos períodos ou em resposta a estímulos.

Uma abordagem amplamente adotada e eficaz para a detecção de piscadas, implementada neste protótipo, é o cálculo do *Eye Aspect Ratio* (EAR) (SOUKUPOVÁ; CECH, 2016). O EAR é uma métrica simples, porém robusta, calculada a partir das coordenadas dos *landmarks* faciais que contornam o olho. Para um único olho, são tipicamente utilizados seis *landmarks*, como ilustrado na Figura 11: dois pontos horizontais (P_1, P_4) nas extremidades do olho, e quatro pontos verticais (P_2, P_3, P_5, P_6) nas pálpebras superior e inferior.

O EAR é definido pela razão entre a distância vertical entre as pálpebras e a distância horizontal entre os cantos do olho:

$$\text{EAR} = \frac{||P_2 - P_6|| + ||P_3 - P_5||}{2 \cdot ||P_1 - P_4||}$$

onde $\|P_i - P_j\|$ denota a distância euclidiana entre os *landmarks* P_i e P_j .



Figura 11 – Localização dos seis landmarks faciais (P_1 a P_6) utilizados para o cálculo do Eye Aspect Ratio (EAR) para um olho.

Quando o olho está aberto, o valor do EAR é relativamente constante. Durante uma piscada, as pálpebras se aproximam, causando uma queda abrupta no valor do EAR. Quando o olho se abre novamente, o EAR retorna ao seu valor normal. Uma piscada é, portanto, detectada quando o valor do EAR cai abaixo de um limiar predefinido por um curto período e, em seguida, sobe novamente acima do limiar. A Figura 13 pode ilustrar essa variação do EAR ao longo do tempo durante uma sequência de piscadas.



Figura 12 – Detecção de piscada dos olhos direito e esquerdo.

Além da detecção baseada no EAR, análises mais diretas do movimento vertical e horizontal das pálpebras podem ser realizadas utilizando os *landmarks* palpebrais. Módulos específicos para essa análise foram desenvolvidos, permitindo, por exemplo, medir a velocidade de fechamento e abertura das pálpebras ou detectar micro-movimentos palpebrais que também podem ser indicativos de vivacidade. A combinação dessas abordagens pode levar a um sistema de detecção de piscadas ainda mais robusto e discriminativo.

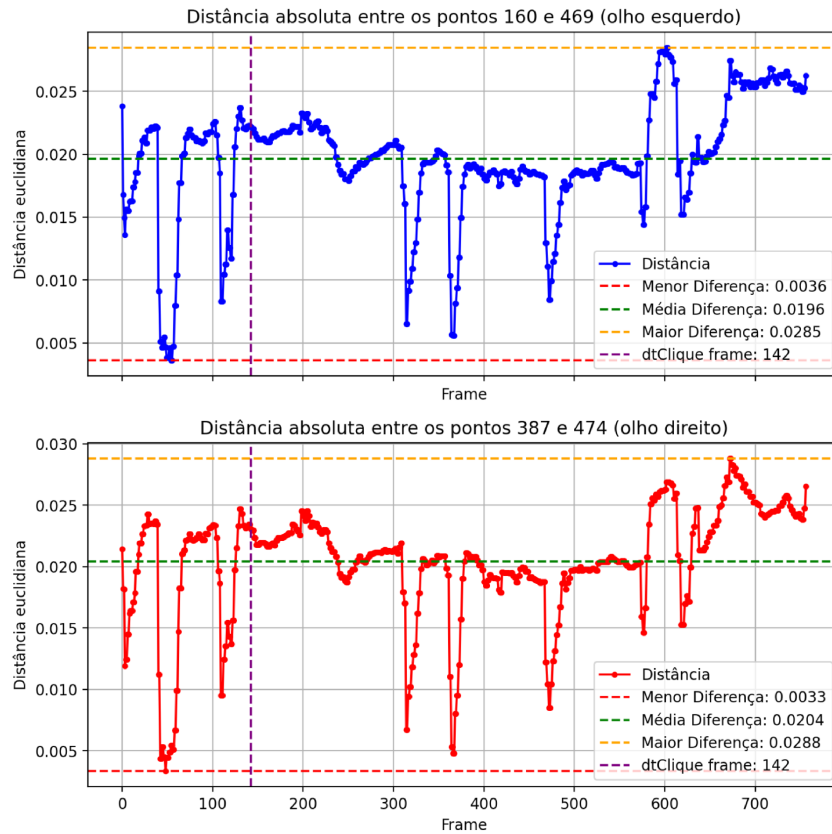


Figura 13 – Variação típica do valor do Eye Aspect Ratio (EAR) ao longo do tempo durante uma sequência de piscadas. A queda abaixo do limiar indica o fechamento do olho.

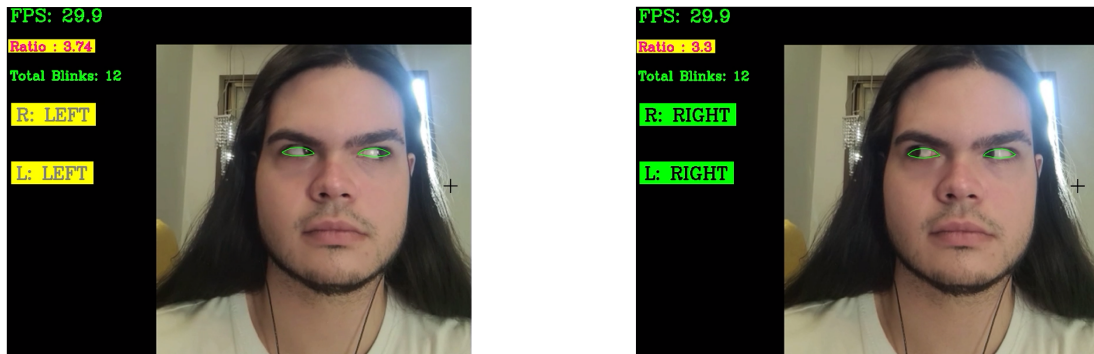
3.5.3 Análise de Direção do Olhar (Gaze Estimation)

A capacidade de estimar a direção para onde o usuário está olhando (*gaze estimation*) é uma funcionalidade poderosa para sistemas de prova de vida interativos. Ela permite verificar se o usuário consegue seguir instruções visuais, como focar em um ponto específico na tela ou acompanhar um objeto em movimento, o que é extremamente difícil de ser fraudado por ataques de apresentação estáticos ou mesmo por vídeos pré-gravados que não estejam sincronizados com o desafio proposto.

A estimativa da direção do olhar é um problema desafiador, com abordagens que variam em complexidade e precisão (HANSEN; JI, 2010). No contexto deste protótipo, que visa uma solução executável em tempo real e com recursos computacionais limitados, foram exploradas abordagens que se baseiam na posição relativa da pupila/íris dentro da ROI do olho, em conjunto com a orientação da cabeça.

Após a detecção do centro da pupila/íris (Seção 3.5.1), sua posição pode ser normalizada em relação às dimensões da ROI do olho ou em relação aos cantos do olho, que são definidos pelos *landmarks* faciais. Por exemplo, uma pupila significativamente deslocada para a esquerda dentro da ROI do olho esquerdo, enquanto a cabeça permanece relativamente frontal, indica que o usuário está olhando para a sua esquerda. Módulos de

dicados foram desenvolvidos para calcular essa posição relativa e classificá-la em direções gerais (e.g., "centro", "esquerda", "direita", "para cima", "para baixo").



(a) Exemplo de direção do olhar para a esquerda, com a pupila deslocada para o lado esquerdo da Região de Interesse (ROI) ocular.

(b) Exemplo de direção do olhar para a direita, com a pupila deslocada para o lado direito da Região de Interesse (ROI) ocular.

Figura 14 – Ilustração esquemática da inferência da direção do olhar (esquerda e direita) baseada na posição relativa da pupila dentro da Região de Interesse (ROI) do olho.

A orientação da cabeça, estimada conforme a Seção 3.4, é um fator importante a ser considerado, pois o olhar é frequentemente coordenado com os movimentos da cabeça. A informação da direção do olhar pode ser utilizada para compensar os movimentos da cabeça e obter uma estimativa do ponto de foco do usuário no ambiente ou na tela do dispositivo.

Embora a precisão da estimativa de olhar possa ser limitada sem calibração específica do usuário ou modelos 3D detalhados do olho, mesmo uma classificação mais grosseira da direção do olhar já é suficiente para muitos desafios de vivacidade. Por exemplo, solicitar ao usuário que olhe sequencialmente para os quatro cantos da tela e verificar se a pupila se move de forma correspondente constitui um teste de vivacidade robusto. A dificuldade em simular esses movimentos coordenados e responsivos do olhar representa uma barreira significativa para fraudadores.

Tabela 3 – Exemplo de classificação da direção do olhar baseada na posição normalizada da pupila (X_{norm} , Y_{norm}) dentro da ROI ocular.

Direção do Olhar	Condição em X_{norm} (Horizontal)	Condição em Y_{norm} (Vertical)
Centro	[0.4, 0.6]	[0.4, 0.6]
Esquerda	[0.0, 0.3]	[0.0, 1.0]
Direita	[0.7, 1.0]	[0.0, 1.0]
Para Cima	[0.0, 1.0]	[0.0, 0.3]
Para Baixo	[0.0, 1.0]	[0.7, 1.0]

Nota: Coordenadas normalizadas variam de 0.0 a 1.0. Os intervalos são exemplificativos.

3.6 Implementação de Interações e Verificações de Prova de Vida

As análises passivas e a extração de características faciais dinâmicas, detalhadas nas seções anteriores, são complementadas por interações e verificações que visam confirmar a presença ativa e responsiva do usuário. Diferentemente de desafios explícitos que demandam uma sequência complexa de ações, a abordagem adotada neste protótipo combina pré-condições de pose com interações oculares específicas, buscando um equilíbrio entre segurança, usabilidade e naturalidade. Esta seção descreve como a orientação da cabeça é validada de forma contínua e como as interações oculares, incluindo piscadas e o direcionamento implícito do olhar, são empregadas para reforçar a prova de vida (GALBALLY; MARCEL; FIERREZ, 2014).

3.6.1 Validação Contínua da Posição da Cabeça como Pré-requisito

No sistema desenvolvido, não são emitidas instruções explícitas para que o usuário realize movimentos específicos com a cabeça, como virar para a esquerda ou direita sob comando. Em vez disso, a correta orientação da cabeça é tratada como uma pré-condição contínua para o prosseguimento da análise de vivacidade. O sistema monitora em tempo real os ângulos de Euler (guinada ψ , inclinação θ , e rotação axial ϕ), calculados conforme descrito na Seção 3.4.

Se a cabeça do usuário permanecer significativamente angulada em relação à câmera – ou seja, fora de um intervalo considerado "frontal" ou "central" (ver Tabela 2 para exemplos de faixas) – por um período prolongado, o processo de avaliação de prova de vida pode não avançar ou não ser finalizado com sucesso. Uma mensagem explícita é apresentada ao usuário solicitando a correção da pose.

Esta validação da centralização da cabeça cumpre dois propósitos:

1. **Garantia de Qualidade para Outras Análises:** Assegura que a face esteja bem posicionada para a extração confiável de *landmarks* e para as análises oculares subsequentes.
2. **Deteção de Anomalias Grosseiras:** Uma cabeça excessivamente angulada e imóvel por longos períodos pode ser um indicativo de uma tentativa de fraude com um objeto estático ou uma fotografia mal posicionada.

Embora sutil, essa verificação contínua da pose cefálica age como um portão de controle, assegurando que apenas os usuários que demonstram um comportamento postural minimamente adequado prossigam na avaliação de vivacidade.

3.6.2 Interações Oculares para Confirmação de Vivacidade

As interações oculares constituem um componente mais ativo e explícito na verificação de vivacidade, combinando instruções diretas para ações reflexas e a indução de movimentos oculares através de tarefas interativas.

3.6.2.1 Verificação de Piscadas Naturais

Em um determinado momento do processo de verificação, o sistema instrui explicitamente o usuário a "PISQUE os olhos naturalmente". Esta instrução é crucial, pois a ausência de piscadas ou um padrão de piscada muito artificial pode indicar uma tentativa de fraude. A validação desta ação não se limita a detectar uma única piscada forçada, mas sim a observar a presença de um comportamento de piscar que seja consistente com a naturalidade esperada, utilizando o método de detecção baseado no *Eye Aspect Ratio* (EAR) (Seção 3.5.2) ou outras análises de movimento palpebral. A análise pode considerar a frequência e a duração das piscadas dentro de uma janela de tempo após a instrução, buscando um padrão que denote uma resposta genuína e não um artefato.

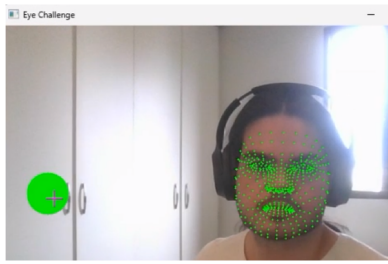
3.6.2.2 Direcionamento Implícito do Olhar através de Interação com Alvos

Para verificar a capacidade de direcionamento do olhar de forma interativa e intuitiva, o sistema emprega uma estratégia de indução implícita. Em vez de comandar explicitamente "Olhe para a esquerda" ou "Olhe para a direita", são apresentados ao usuário alvos visuais – tipicamente círculos animados – que aparecem em diferentes posições na tela (e.g., à esquerda, à direita e ao centro). O usuário é instruído a "Toque no círculo" ou "Clique no círculo" assim que este aparece.

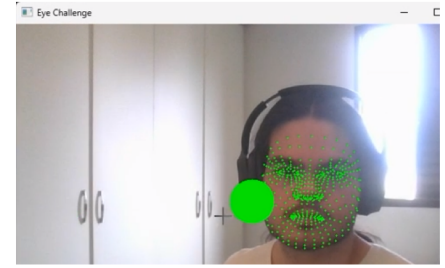
Esta interação possui um duplo propósito:

1. **Tarefa Explícita de Interação:** O usuário realiza uma ação motora (toque ou clique), que por si só já requer atenção e presença.
2. **Indução e Verificação Implícita do Olhar:** Para tocar ou clicar no círculo com precisão, o usuário é naturalmente induzido a direcionar seu olhar para a localização do alvo. Concomitantemente à espera da ação de toque/clique, o sistema utiliza as técnicas de estimativa da direção do olhar (Seção 3.5.3) para verificar se o movimento ocular do usuário é congruente com a aparição e a posição do círculo na tela.

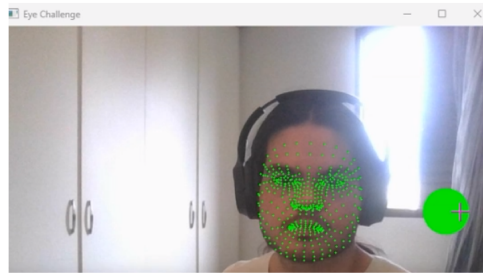
A validação considera a correlação entre a trajetória estimada do olhar e a sequência de aparição dos alvos. Um movimento ocular que acompanha consistentemente os alvos antes ou durante o ato de tocar/clicar é um forte indicador de vivacidade e de interação genuína com a interface. Esta abordagem torna o teste de vivacidade mais orgânico e menos propenso a ser decifrado ou simulado por respostas pré-programadas em um ataque.



(a) Alvo à esquerda.



(b) Alvo ao centro.



(c) Alvo à direita.

Figura 15 – Sequência de apresentação de alvos visuais (círculos) em diferentes posições da tela para indução do movimento ocular e interação por toque/clique.

A combinação dessas verificações – a manutenção de uma pose cefálica adequada, a demonstração de piscadas naturais sob instrução e a capacidade de interagir visual e motoramente com alvos dinâmicos – compõe um conjunto robusto de mecanismos para a prova de vida.

Tabela 4 – Desafios baseados em ações oculares e critérios de validação.

Desafio	Ação Esperada do Usuário	CrITÉRIOS de Validação (Exemplos)
Pisque os Olhos	Fechar e abrir os olhos uma vez	Detecção de queda no EAR abaixo do limiar, seguida de retorno.
Seguir Alvo Visual	Mover os olhos acompanhando um ponto na tela	Correspondência entre a direção estimada do olhar e a posição do alvo.
Olhar para Canto	Direcionar o olhar para um canto específico	Posição da pupila na ROI ocular compatível com o canto indicado.

3.7 Mecanismos Adicionais de Antispoofing

As interações e verificações descritas anteriormente, que se baseiam na análise de movimentos e respostas a estímulos, formam uma camada robusta de defesa contra ataques de apresentação. Contudo, para aumentar ainda mais a confiabilidade do sistema de prova de vida, é vantajoso incorporar mecanismos adicionais de *antispoofing*. Estes

mecanismos frequentemente operam de forma passiva ou semi-passiva, analisando características intrínsecas da imagem facial ou respostas fisiológicas sutis que são difíceis de serem forçadas. Esta seção detalha algumas dessas abordagens complementares implementadas, especificamente a análise da cor da esclera, a verificação da reação pupilar à luz e outras considerações gerais para detecção de fraude.

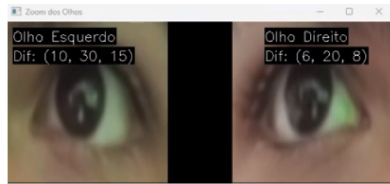
3.7.1 Análise da Cor da Esclera

A esclera, popularmente conhecida como a "parte branca do olho", possui características de cor que podem ser exploradas como um indicador de vivacidade. Olhos humanos reais apresentam uma esclera com tonalidades específicas, que podem diferir significativamente daquelas encontradas em representações artificiais, como olhos impressos em fotografias ou renderizados em modelos 3D de baixa fidelidade. Uma coloração atípica da esclera, como tons excessivamente brancos (comuns em impressões), amarelados (em certas condições de saúde ou em fotos antigas) ou com padrões de cor não naturais, pode sinalizar uma tentativa de fraude.

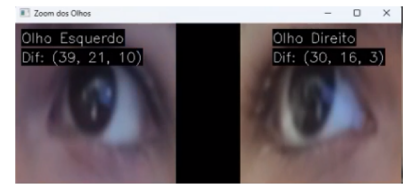
A metodologia para análise da cor da esclera no protótipo envolve as seguintes etapas:

1. **Segmentação da Região da Esclera:** Utilizando os *landmarks* faciais que delimitam o contorno dos olhos e, potencialmente, os *landmarks* da íris (Seção 3.5.1), as regiões correspondentes à esclera são isoladas em cada olho. É crucial excluir a íris, a pupila, as pálpebras e os cílios para uma análise precisa da cor da esclera. Módulos de software específicos, como o desenvolvido para a aplicação de análise de cores, são responsáveis por esta segmentação.
2. **Extração de Características de Cor:** Uma vez segmentada, as características de cor da região da esclera são extraídas. Isso envolve a conversão da ROI para um espaço de cores adequado (e.g., RGB, HSV ou Lab) e o cálculo de estatísticas descritivas dos canais de cor, como a média, mediana, desvio padrão ou histogramas de cor. Além disso, uma etapa de calibração é aplicada, para se adaptar às condições iniciais de iluminação que impactam diretamente na extração da cor.
3. **Classificação e Decisão:** As características de cor extraídas são comparadas com um modelo de referência que define as faixas de cor esperadas para uma esclera humana saudável e em condições normais de iluminação. Se os valores medidos desviarem significativamente desses padrões, um alerta de possível *spoofing* pode ser gerado, e prova de vida falha nesta etapa. Por exemplo, uma esclera com saturação de cor muito baixa e brilho muito alto pode indicar uma impressão.

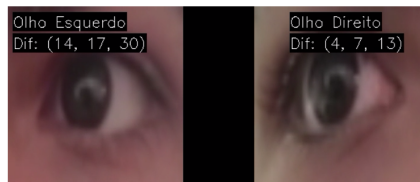
A eficácia desta técnica depende da qualidade da segmentação da esclera e da robustez do modelo de cor de referência frente a variações naturais na tonalidade da esclera entre indivíduos e diferentes condições de iluminação ambiente.



(a) Aparência da esclera sob iluminação com luz verde.



(b) Aparência da esclera sob iluminação com luz azul.



(c) Aparência da esclera sob iluminação com luz vermelha.

Figura 16 – Análise comparativa da aparência e reflexibilidade da esclera ocular quando iluminada por fontes de luz de diferentes cores (verde, azul e vermelha), visando identificar padrões que auxiliem na distinção entre um olho real e uma representação artificial.

3.7.2 Análise da Reação Pupilar à Luz (Reflexo Fotomotor)

O reflexo fotomotor (RFM), ou reação pupilar à luz, é a constrição involuntária da pupila em resposta a um aumento na intensidade luminosa que atinge a retina, seguida por sua dilatação quando a intensidade luminosa diminui. Este é um reflexo neural complexo, mediado pelo sistema nervoso autônomo, e sua presença e dinâmica são indicadores extremamente fortes de vivacidade, sendo muito difíceis de serem replicados artificialmente (SUN; TAN; WANG, 2005).

O protótipo implementa uma verificação baseada neste reflexo da seguinte maneira:

1. **Medição Basal do Diâmetro Pupilar:** Inicialmente, o diâmetro da pupila é medido em condições de iluminação ambiente normais ou controladas pela tela do dispositivo. Isto requer um algoritmo robusto de detecção de pupila e estimativa de seu diâmetro, conforme explorado em módulos específicos de detecção pupilar.
2. **Estímulo Luminoso:** Em um momento apropriado, o sistema emite um estímulo luminoso repentino. Isto é realizado fazendo com que a tela do dispositivo exiba uma cor branca brilhante por um curto período.

3. **Monitoramento da Resposta Pupilar:** Durante e imediatamente após o estímulo luminoso, o diâmetro da pupila é continuamente monitorado. Espera-se que ocorra uma constrição (redução do diâmetro) da pupila.

4. **Validação do Reflexo:** A validação envolve verificar se:

- Ocorreu uma redução significativa no diâmetro da pupila após o estímulo.
- A magnitude e a latência (tempo de resposta) da constrição estão dentro de faixas fisiologicamente plausíveis.
- Após a remoção do estímulo (a tela retorna a uma luminosidade normal), a pupila gradualmente se dilata de volta ao seu diâmetro basal ou próximo a ele.

A ausência de uma reação pupilar detectável, ou uma reação que não se enquadre nos parâmetros esperados (e.g., uma mudança de diâmetro muito lenta, muito rápida ou de magnitude implausível), pode ser considerada uma forte evidência de tentativa de fraude, como o uso de olhos protéticos, fotografias ou vídeos. Os desafios desta técnica incluem a necessidade de uma estimativa precisa e robusta do diâmetro pupilar em tempo real e o controle adequado do estímulo luminoso para eliciar uma resposta clara sem causar desconforto excessivo ao usuário.

3.7.3 Outras Abordagens e Heurísticas de Detecção de Fraude

Além das técnicas específicas mencionadas, um sistema de *antispoofing* pode se beneficiar de um conjunto de heurísticas e verificações adicionais, implementadas em módulos de software dedicados à detecção de anomalias e artefatos comuns em ataques de apresentação. Estas podem incluir:

- **Detecção de Padrões de Moiré:** O padrão de Moiré pode surgir quando uma tela digital (onde um vídeo de *spoofing* pode estar sendo exibido) é recapturada por outra câmera. A detecção desses padrões característicos pode indicar um ataque de repetição.
- **Análise de Reflexos Especulares:** Reflexos nos olhos ou na superfície da face podem fornecer pistas. Reflexos planos e uniformes, típicos de uma fotografia ou tela, diferem dos reflexos curvos e dinâmicos esperados em uma face 3D real e olhos úmidos.
- **Consistência Temporal de Características:** A análise da consistência de características faciais (como a própria malha de *landmarks* ou a textura em regiões específicas) ao longo do tempo pode revelar anomalias. Faces reais exibem micro-movimentos e deformações sutis que podem estar ausentes em *spoofs* estáticos ou vídeos em loop.



Figura 17 – Sequência ilustrativa da verificação do reflexo fotomotor: (a) diâmetro pupilar basal, (b) aplicação de estímulo luminoso pela tela, (c) constrição pupilar observada.

- **Detecção de Bordas de Dispositivos:** Em ataques onde uma foto ou um celular exibindo um vídeo são apresentados à câmera, pode ser possível detectar as bordas do dispositivo de apresentação.

A implementação dessas heurísticas geralmente envolve algoritmos de processamento de imagem e aprendizado de máquina para identificar os padrões suspeitos. Embora nenhuma heurística isolada seja infalível, a combinação de múltiplas verificações contextuais e baseadas em artefatos pode contribuir significativamente para a robustez geral do sistema de prova de vida.

A incorporação desses mecanismos adicionais de *antispoofting* visa criar um sistema de defesa em profundidade, onde diferentes técnicas atuam em conjunto para dificultar ao máximo o sucesso de tentativas de fraude.

3.8 Integração dos Módulos e Tomada de Decisão

As diversas técnicas de análise facial, ocular e de *antispoofing* detalhadas nas seções anteriores, embora individualmente capazes de fornecer indícios de vivacidade, atingem seu pleno potencial quando integradas em um sistema coeso e orquestrado. Esta seção descreve como os diferentes módulos do protótipo são combinados e gerenciados por um algoritmo central, culminando na geração de um escore de vivacidade e na tomada de uma decisão final sobre a autenticidade do usuário. O desenvolvimento de um "algoritmo piloto" foi fundamental para essa integração, permitindo não apenas o fluxo controlado das etapas de verificação, mas também a apresentação de diferentes níveis de informação através de uma "visão do desenvolvedor" e uma "visão do usuário".

3.8.1 Orquestração das Diferentes Etapas de Verificação pelo Algoritmo Piloto

O núcleo da integração do sistema de prova de vida reside em um "algoritmo piloto" principal. Este algoritmo é responsável por coordenar a execução sequencial e, em alguns casos, paralela, de todos os módulos de análise: desde a aquisição da imagem e detecção facial, passando pela extração de *landmarks*, estimativa de pose da cabeça, análise ocular (incluindo piscadas e direcionamento do olhar induzido), até a aplicação dos mecanismos adicionais de *antispoofing* e a gestão das interações com o usuário.

O fluxo de orquestração geralmente segue uma lógica progressiva, como ilustrado na Figura 4:

1. **Verificações Iniciais e Pré-condições:** O algoritmo piloto inicia com a captura da imagem, detecção de face única e a validação contínua da centralização da cabeça, conforme descrito na Seção 3.6.1. Se estas condições não forem satisfeitas, o processo pode ser pausado ou interrompido.
2. **Extração de Características Fundamentais:** Em seguida, procede-se à extração dos pontos fiduciais da face, que são a base para a maioria das análises subsequentes.
3. **Análises Passivas e Ativas:** Com os *landmarks* disponíveis, são disparadas as análises de pose da cabeça (para monitoramento contínuo), detecção de piscadas (em resposta à instrução explícita), análise da reação pupilar à luz e a interação com alvos visuais para indução do olhar. Mecanismos como a análise da cor da esclera também podem ser executados.
4. **Coleta de Evidências:** O algoritmo piloto coleta os resultados e as métricas parciais de cada um desses módulos.

Uma característica importante do desenvolvimento deste algoritmo piloto é a implementação de duas interfaces de visualização distintas, alternáveis conforme a necessidade:

- **Visão do Desenvolvedor:** Esta interface é projetada para fins de depuração, calibração e análise detalhada do comportamento do sistema. Ela sobrepõe à imagem de vídeo uma vasta gama de informações em tempo real, como os *landmarks* faciais detectados, os ângulos de pose da cabeça, valores de EAR, o estado da detecção de piscadas, a classificação da direção do olhar resultados parciais dos testes de *antispoofing*. Esta visão é instrumental para identificar falhas, ajustar parâmetros e compreender a contribuição de cada módulo.
- **Visão do Usuário:** Em contraste, esta interface é minimalista e focada na experiência do usuário final. Ela exibe apenas as instruções necessárias para a interação (e.g., "PISQUE os olhos naturalmente", "Toque no círculo"), feedback simples sobre o progresso da verificação (e.g., "Analisando...") e, ao final, o resultado da verificação de vivacidade. O objetivo é tornar o processo o mais intuitivo e menos intrusivo possível, ocultando a complexidade subjacente.

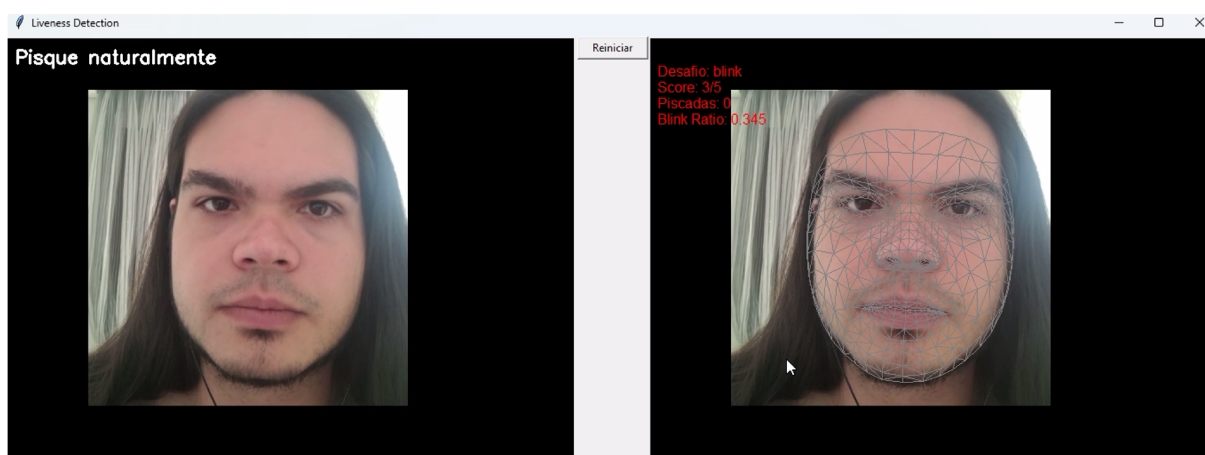


Figura 18 – Exemplo da "Visão do Desenvolvedor" e "Visão do Usuário".

A capacidade de alternar entre essas visões durante o desenvolvimento e teste do protótipo demonstrou ser de grande valia.

3.8.2 Lógica de Decisão Final para Aprovação/Rejeição da Vivacidade

Após a execução de todas as etapas de verificação relevantes orquestradas pelo algoritmo piloto, as evidências coletadas de cada módulo de análise são agregadas para formar uma decisão final sobre a vivacidade do usuário. No protótipo desenvolvido, esta agregação culmina na geração de um **escore básico de vivacidade**.

A formulação deste escore é um aspecto crucial do sistema. Embora uma variedade de abordagens possa ser adotada, desde modelos de aprendizado de máquina complexos até sistemas baseados em regras, o protótipo inicial focou em uma lógica mais direta. O escore pode ser constituído, por exemplo, por:

- Uma contagem ponderada de testes de vivacidade bem-sucedidos.
- A soma de pontuações parciais atribuídas a cada verificação, onde verificações consideradas mais robustas (e.g., reação pupilar à luz, interação com múltiplos alvos) podem ter um peso maior.
- Uma combinação de resultados booleanos (passou/falhou) de diferentes testes, processados através de uma árvore de decisão simples.

Por exemplo, a observação consistente de piscadas naturais, a correta interação com os alvos visuais (implicando movimento ocular adequado), uma reação pupilar fisiologicamente correta e uma cor de esclera dentro dos padrões esperados contribuiriam positivamente para o escore final. Falhas em múltiplos testes, ou em testes considerados críticos, resultariam em um escore baixo.

A tomada de decisão final é então realizada comparando este escore de vivacidade com um ou mais limiares predefinidos. Um único limiar pode ser usado para uma decisão binária:

- Se $\text{Escore} \geq \text{Limiar_Vivacidade}$: Usuário considerado genuíno (prova de vida bem-sucedida).
- Se $\text{Escore} < \text{Limiar_Vivacidade}$: Suspeita de fraude ou falha na verificação (prova de vida malsucedida).

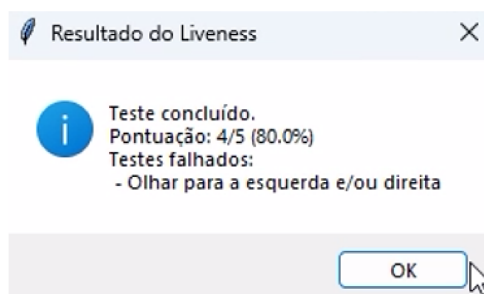


Figura 19 – Processo de agregação de evidências dos diferentes módulos de análise para a formação do escore básico de vivacidade e a subsequente tomada de decisão.

A determinação do valor ótimo para este limiar é tipicamente um processo empírico, realizado através de testes extensivos com dados de usuários reais e tentativas de fraude simuladas, buscando um equilíbrio adequado entre a Taxa de Falsa Aceitação

(FAR) e a Taxa de Falsa Rejeição (FRR). Para um protótipo, o escore pode ser mais um indicador para o desenvolvedor, e o limiar pode ser ajustado progressivamente.

Tabela 5 – Exemplo ilustrativo de contribuição de diferentes verificações para o escore de vivacidade.

Verificação de Vivacidade	Resultado Positivo (Pontos/Peso)	Resultado Negativo (Pontos/Peso)
Centralização da Cabeça	(pré-requisito)	N/A (processo não avança)
Piscadas Naturais	+10	0
Interação com Alvos (Gaze)	+10 por alvo	0
Reação Pupilar à Luz	+10	0
Cor da Esclera	+10	0

4 Resultados

Este capítulo é dedicado à apresentação e análise dos resultados experimentais obtidos com o protótipo de solução de prova de vida (*liveness detection*) desenvolvido. Conforme detalhado no Capítulo 3, o sistema integra múltiplos módulos de análise facial e ocular, além de interações específicas e mecanismos de *antispoofing*, culminando em um escore de vivacidade. O objetivo dos testes realizados foi avaliar a eficácia do protótipo tanto em cenários de uso legítimo, por usuários reais, quanto em tentativas de fraude (*spoofing attacks*) utilizando diferentes artefatos de apresentação. Serão apresentados os dados coletados, as métricas de desempenho calculadas e uma discussão inicial sobre os achados.

4.1 Configuração Experimental e Métricas de Avaliação

Para avaliar o desempenho do protótipo de detecção de vivacidade, foi definida uma metodologia experimental que abrangeu testes com usuários legítimos e simulações de ataques por apresentação. Esta seção descreve brevemente o ambiente de teste, os materiais utilizados para os ataques e as métricas empregadas para quantificar a performance do sistema.

4.1.1 Ambiente e Materiais de Teste

Os testes de uso legítimo foram conduzidos utilizando duas configurações de câmera principais:

1. **Webcam Dedicada/Integrada:** Uma webcam padrão conectada a um computador notebook, representando um cenário comum de uso em aplicações web ou de desktop.
2. **Câmera de Smartphone via DroidCam:** Um smartphone Android teve sua câmera conectada ao computador através do aplicativo DroidCam, simulando o uso de dispositivos móveis ou configurações onde a câmera do celular é preferida ou a única disponível.

Para os testes de ataque por apresentação, foram preparados os seguintes artefatos:

1. **Imagem Estática em Tela de Smartphone:** Uma fotografia de um rosto foi exibida na tela de um smartphone, e este dispositivo foi então apresentado à webcam

do sistema de prova de vida. Este tipo de ataque visa simular a tentativa de fraude utilizando uma foto impressa ou digital de um usuário autorizado.

2. **Vídeos Deepfake:** Foram gerados vídeos sintéticos utilizando a plataforma de inteligência artificial Kling AI. O *prompt* utilizado para a geração foi: "Gere um vídeo hiper-realista de uma pessoa olhando à frente e piscando naturalmente". Foram criados dois modelos distintos de *deepfakes* a partir deste *prompt*, e cada modelo foi utilizado em múltiplas tentativas de ataque. Estes vídeos representam uma ameaça mais sofisticada, pois podem exibir movimentos faciais como piscadas.

Todas as interações e análises foram gerenciadas pelo "algoritmo piloto" descrito na Seção 3.8. Conforme a lógica de pontuação definida para o protótipo, um escore acumulado de 100% foi considerado como "aprovação" (vivacidade confirmada), enquanto qualquer escore inferior a 100% foi interpretado como "reprovação" (suspeita de fraude ou falha na verificação).

4.1.2 Métricas de Avaliação de Desempenho

Para quantificar o desempenho do sistema de detecção de vivacidade, foram utilizadas as seguintes métricas padrão em avaliação de sistemas biométricos e de *antispoofing*:

- **Taxa de Aceitação Verdadeira (TAR)** ou *True Acceptance Rate (TAR)*: Representa a proporção de tentativas de acesso de usuários legítimos que foram corretamente aceitas pelo sistema. É calculada como:

$$TAR = \frac{VP}{VP + FN}$$

onde VP (Verdadeiros Positivos) é o número de tentativas legítimas corretamente aceitas, e FN (Falsos Negativos) é o número de tentativas legítimas incorretamente rejeitadas.

- **Taxa de Falsa Rejeição (TFR)** ou *False Rejection Rate (FRR)*: Indica a proporção de tentativas de acesso de usuários legítimos que foram incorretamente rejeitadas pelo sistema. É calculada como:

$$TFR = \frac{FN}{VP + FN} = 1 - TAR$$

- **Taxa de Rejeição Verdadeira (TRR)** ou *True Rejection Rate (TRR)*: Refere-se à proporção de tentativas de ataque por apresentação (fraudes) que foram corretamente rejeitadas pelo sistema. É calculada como:

$$TRR = \frac{VN}{VN + FP}$$

onde VN (Verdadeiros Negativos) é o número de tentativas de ataque corretamente rejeitadas, e FP (Falsos Positivos) é o número de tentativas de ataque incorretamente aceitas.

- **Taxa de Falsa Aceitação (TFA)** ou *False Acceptance Rate (FAR)*: Denota a proporção de tentativas de ataque por apresentação que foram incorretamente aceitas pelo sistema como se fossem de um usuário legítimo. É calculada como:

$$\text{TFA} = \frac{\text{FP}}{\text{VN} + \text{FP}} = 1 - \text{TRR}$$

Idealmente, um sistema de prova de vida robusto deve apresentar altos valores de TAR e TRR, e baixos valores de FRR e FAR.

4.2 Resultados dos Testes de Uso Legítimo

A avaliação do desempenho do sistema de prova de vida em cenários de uso legítimo é fundamental para aferir sua usabilidade e a capacidade de corretamente autenticar usuários genuínos. Foram conduzidas séries de testes utilizando diferentes configurações de câmera para simular condições de uso variadas. Conforme a metodologia de pontuação adotada, um escore de 100% foi necessário para a aprovação da tentativa de prova de vida; qualquer escore inferior resultou em reprovação.

4.2.1 Desempenho com Webcam Dedicada/Integrada

Nesta primeira bateria de testes, foram realizadas 10 (dez) tentativas de prova de vida por um usuário legítimo utilizando uma webcam padrão conectada a um computador. O objetivo era avaliar a performance do sistema em uma configuração de hardware convencional.

Das 10 tentativas, 9 (nove) foram concluídas com sucesso, alcançando um escore de 100% e sendo, portanto, corretamente aprovadas como interações genuínas. Estes casos representam Verdadeiros Positivos (VP). Em uma única tentativa, ocorreu uma falha específica no sub-teste de direcionamento do olhar para os alvos circulares. Embora os demais testes de vivacidade tenham sido superados, a falha neste componente resultou em um escore final de 80%. De acordo com o critério estabelecido de que apenas um escore de 100% configuraria uma aprovação, esta tentativa foi considerada uma reprovação, caracterizando um Falso Negativo (FN).

Com base nesses resultados (9 VP e 1 FN em 10 tentativas), as métricas de desempenho para o uso com webcam foram calculadas da seguinte forma:

- Taxa de Aceitação Verdadeira (TAR):

$$\text{TAR}_{\text{webcam}} = \frac{9}{9 + 1} = \frac{9}{10} = 90\%$$

- Taxa de Falsa Rejeição (TFR):

$$\text{TFR}_{\text{webcam}} = \frac{1}{9 + 1} = \frac{1}{10} = 10\%$$

A ocorrência de uma Falsa Rejeição, atribuída à sensibilidade do teste de direcionamento do olhar naquela instância particular, indica um ponto de atenção para calibração ou robustez deste módulo específico em certas condições de uso com webcam.

4.2.2 Desempenho com Câmera de Smartphone (via DroidCam)

Para avaliar o sistema em um cenário que simula o uso de câmeras de dispositivos móveis, foram realizadas 10 (dez) tentativas de prova de vida por um usuário legítimo utilizando a câmera de um smartphone conectada ao computador via o aplicativo DroidCam.

Nesta configuração, todas as 10 (dez) tentativas foram concluídas com sucesso, alcançando um escore de 100% e sendo corretamente aprovadas. Portanto, foram registrados 10 Verdadeiros Positivos (VP) e 0 Falsos Negativos (FN). As métricas de desempenho para o uso com DroidCam são:

- Taxa de Aceitação Verdadeira (TAR):

$$\text{TAR}_{\text{DroidCam}} = \frac{10}{10 + 0} = \frac{10}{10} = 100\%$$

- Taxa de Falsa Rejeição (TFR):

$$\text{TFR}_{\text{DroidCam}} = \frac{0}{10 + 0} = \frac{0}{10} = 0\%$$

O desempenho nesta configuração é um resultado promissor, indicando boa adaptabilidade do sistema à qualidade e características de imagem tipicamente fornecidas por câmeras de smartphone. É interessante notar que, embora tenha sido observada durante a fase de desenvolvimento uma maior dificuldade na estimativa precisa do olhar do usuário no sentido horizontal com a câmera do smartphone – devido a fatores como a proporção da tela do celular e o ângulo de captura quando o dispositivo é segurado –, nos testes formais conduzidos, esta dificuldade não se traduziu em falhas de aprovação.

Tabela 6 – Resultados consolidados dos testes de prova de vida com usuários legítimos.

Configuração da Câmera	Nº de Tentativas	VP	FN	TAR (%)	TFR (%)
Webcam Dedicada/Integrada	10	9	1	90,0	10,0
Smartphone (via DroidCam)	10	10	0	100,0	0,0
Total/Média Ponderada	20	19	1	95,0	5,0

4.2.3 Sumário dos Resultados de Uso Legítimo

Os resultados dos testes de uso legítimo, comparando as duas configurações de câmera, são consolidados na Tabela 6.

No geral, o sistema demonstrou um alto índice de sucesso na autenticação de usuários legítimos, com uma Taxa de Aceitação Verdadeira global de 95%. A única falha de rejeição incorreta ocorreu em um cenário específico com webcam, sugerindo que a robustez do teste de direcionamento do olhar pode ser um foco para refinamentos futuros, visando manter a segurança sem comprometer a experiência do usuário.

4.3 Resultados dos Testes de Ataques por Apresentação (Spoofing)

A robustez de um sistema de prova de vida é verdadeiramente testada pela sua capacidade de resistir a tentativas de ataque por apresentação, onde artefatos fraudulentos são utilizados na tentativa de enganar os sensores biométricos. Foram conduzidos experimentos com dois tipos comuns de ataques de apresentação: o uso de uma imagem estática exibida em uma tela e o uso de vídeos *deepfake*. O objetivo foi avaliar a eficácia dos mecanismos de *antispoofing* implementados no protótipo.

4.3.1 Resistência a Ataques com Imagem Estática

O primeiro cenário de ataque consistiu na apresentação de uma imagem facial estática, exibida na tela de um smartphone, diretamente à webcam do sistema de prova de vida. Este tipo de ataque é um dos mais básicos e comuns, visando explorar sistemas que não verificam adequadamente a tridimensionalidade ou os sinais vitais dinâmicos.

Foram realizadas 10 (dez) tentativas de ataque utilizando este método. Em todas as 10 (dez) tentativas, o sistema de prova de vida corretamente identificou a apresentação como fraudulenta e rejeitou o acesso. Isso resultou em 10 Verdadeiros Negativos (VN) e 0 Falsos Positivos (FP).

As métricas de desempenho do sistema contra ataques de imagem estática foram:

- Taxa de Rejeição Verdadeira (TRR): $TRR_{img_estatica} = \frac{10}{10+0} = \frac{10}{10} = 100\%$
- Taxa de Falsa Aceitação (TFA): $TFA_{img_estatica} = \frac{0}{10+0} = \frac{0}{10} = 0\%$

O desempenho (100% TRR e 0% TFA) contra este tipo de ataque demonstra a eficácia dos mecanismos básicos de detecção de vivacidade implementados, como a exigência de piscadas naturais (ausentes em uma imagem estática), a interação com alvos visuais (que uma imagem não pode realizar) e a análise da reação pupilar à luz (impossível para uma imagem). A Figura 20 pode ilustrar a configuração deste tipo de ataque.

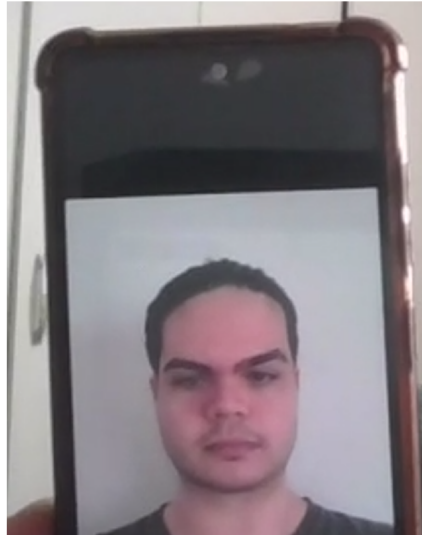


Figura 20 – Ilustração da configuração do teste de ataque por apresentação utilizando uma imagem estática exibida na tela de um smartphone e apresentada à webcam do sistema.

4.3.2 Resistência a Ataques com Vídeos Deepfake

Os vídeos *deepfake* representam uma ameaça consideravelmente mais sofisticada, pois podem simular movimentos faciais dinâmicos, como piscadas e leves movimentos da cabeça, que podem enganar sistemas de prova de vida menos robustos. Para este teste, foram gerados vídeos hiper-realistas utilizando a plataforma Kling AI, com o *prompt* "Gere um vídeo hiper-realista de uma pessoa olhando à frente e piscando naturalmente". Foram criados dois modelos (variações) de *deepfakes* a partir deste *prompt*, e foram realizadas 5 (cinco) tentativas de ataque com cada modelo, totalizando 10 (dez) tentativas.

Das 10 (dez) tentativas de ataque com vídeos *deepfake*, o sistema conseguiu rejeitar corretamente 9 (nove) delas como fraudulentas, caracterizando 9 Verdadeiros Negativos (VN). No entanto, em 1 (uma) das tentativas, o vídeo *deepfake* foi incorretamente aceito pelo sistema como uma interação legítima, resultando em 1 Falso Positivo (FP).

As métricas de desempenho contra os ataques de vídeo *deepfake* foram:

- Taxa de Rejeição Verdadeira (TRR):

$$\text{TRR}_{\text{deepfake}} = \frac{9}{9 + 1} = \frac{9}{10} = 90\%$$

- Taxa de Falsa Aceitação (TFA):

$$\text{TFA}_{\text{deepfake}} = \frac{1}{9 + 1} = \frac{1}{10} = 10\%$$

Embora uma TRR de 90% contra *deepfakes* gerados por IA seja um resultado considerável para um protótipo, a ocorrência de uma Falsa Aceitação (10% TFA) evidencia a complexidade e o desafio contínuo que esses tipos de ataques representam. A análise preliminar da falha sugere que o vídeo *deepfake* que logrou êxito pode ter conseguido simular piscadas de forma convincente o suficiente para passar no teste de EAR e apresentado características que foram interpretadas como reações fisiológicas mínimas nos testes de análise de cor da esclera ou reação pupilar à luz. Conforme mencionado na concepção do protótipo, para esses dois últimos testes, uma pequena variação nos valores medidos já era suficiente para considerar o sub-teste como aprovado. Essa parametrização, embora possa ter contribuído para uma boa taxa de aceitação de usuários legítimos, pode ter sido explorada pelo *deepfake* que conseguiu uma Falsa Aceitação. A Figura 21 apresenta frames dos vídeos *deepfake* utilizados.



(a) Primeiro vídeo gerado por *prompt*.



(b) Segundo vídeo gerado por *prompt*.

Figura 21 – Exemplos de frames dos vídeos deepfake gerados pela plataforma Kling AI e utilizados nos testes de ataque por apresentação.

4.3.3 Sumário dos Resultados de Testes de Ataque

A Tabela 7 consolida os resultados obtidos nos testes de ataque por apresentação.

Tabela 7 – Resultados consolidados dos testes de ataques por apresentação (spoofing).

Tipo de Ataque	Nº de Tentativas	VN	FP	TRR (%)	TFA (%)
Imagem Estática (em tela)	10	10	0	100,0	0,0
Vídeo Deepfake (Kling AI)	10	9	1	90,0	10,0
Total/Média Ponderada	20	19	1	95,0	5,0

Os resultados indicam que o protótipo foi eficaz contra ataques básicos de imagem estática. Contra ataques mais elaborados com vídeos *deepfake*, o sistema demonstrou alta capacidade de rejeição, embora uma vulnerabilidade tenha sido identificada, resultando em uma taxa de falsa aceitação de 10% para essa modalidade específica.

4.4 Análise e Discussão Consolidada dos Resultados

A apresentação dos resultados nas seções anteriores permitiu quantificar o desempenho do protótipo de detecção de vivacidade em diferentes cenários de teste. Nesta seção, os resultados são analisados de forma consolidada, buscando interpretar as métricas de performance, identificar os pontos fortes e as limitações do sistema desenvolvido, e relacionar os achados com as metodologias e decisões de projeto discutidas no Capítulo 3.

4.4.1 Interpretação Geral do Desempenho

Avaliando o conjunto dos testes, o protótipo demonstrou um desempenho geral promissor. Para usuários legítimos (Tabela 6), obteve-se uma Taxa de Aceitação Verdadeira (TAR) de 90% com webcam dedicada/integrada e um resultado de 100% TAR utilizando a câmera de smartphone via DroidCam. Isso resultou em uma TAR ponderada de 95%, indicando uma boa capacidade de reconhecer interações genuínas. A Taxa de Falsa Rejeição (TFR) global de 5% é majoritariamente atribuída a uma única falha no teste de olhar com a webcam.

No que tange à resistência contra ataques por apresentação (Tabela 7), o sistema exibiu uma performance excelente contra imagens estáticas, com 100% de Taxa de Rejeição Verdadeira (TRR) e 0% de Taxa de Falsa Aceitação (TFA). Contra os vídeos *deepfake*, mais desafiadores, o sistema alcançou uma TRR de 90%, com uma TFA de 10%. A TRR ponderada para todos os ataques foi de 95%.

Esses números sugerem um bom equilíbrio inicial entre a segurança contra fraudes comuns e a conveniência para o usuário legítimo, embora pontos específicos de melhoria tenham sido identificados.

4.4.2 Pontos Fortes do Protótipo

Diversos aspectos positivos puderam ser observados a partir dos resultados:

- **Alta Performance com Câmeras de Smartphone (DroidCam):** A obtenção de 100% TAR nos testes com DroidCam é um indicativo de que o sistema pode ser bem adaptado para uso com câmeras de dispositivos móveis, que são prevalentes em muitas aplicações de autenticação.
- **Eficácia Total Contra Imagens Estáticas:** A rejeição de todas as 10 tentativas de ataque com imagem estática demonstra que a combinação de desafios interativos (como a indução do olhar e a solicitação de piscadas) e a análise de sinais vitais (como a reação pupilar à luz, mesmo que com limiares lenientes) é eficaz contra este tipo de fraude básica. A ausência de movimento e de respostas fisiológicas em uma imagem estática é prontamente detectada.
- **Resiliência Considerável a Vídeos Deepfake:** Uma TRR de 90% contra vídeos *deepfake* gerados por IA, embora não perfeita, é um resultado significativo para um protótipo. Indica que a combinação de múltiplas análises (piscadas naturais, interação com alvos, possível detecção sutil de artefatos na esclera ou na reação pupilar) oferece uma barreira considerável mesmo para ataques dinâmicos.

4.4.3 Limitações Identificadas e Análise de Falhas

Os testes também revelaram áreas onde o protótipo pode ser aprimorado:

- **Taxa de Falsa Rejeição com Webcam (10%):** A única Falsa Rejeição ocorreu devido a uma falha no sub-teste de direcionamento do olhar para os alvos circulares, utilizando a webcam. Dado que o critério de aprovação exigia 100% de acerto no score, a falha em um único componente levou à reprovação do usuário legítimo. Isso sugere que a sensibilidade ou a calibração do teste de olhar para a configuração de webcam pode precisar de ajustes para aumentar sua robustez sem comprometer a segurança.
- **Taxa de Falsa Aceitação com Vídeos Deepfake (10%):** A aceitação de um dos vídeos *deepfake* é o ponto mais crítico em termos de segurança. O *prompt* utilizado para gerar os *deepfakes* ("pessoa olhando à frente e piscando naturalmente") focou em simular piscadas, um dos principais indicadores de vivacidade. Além disso, a estratégia de prototipagem adotada, onde para os casos de análise da cor da esclera ou a reação pupilar à luz, a simples variação em valores, ou seja, o menor demonstrativo de reação fisiológica a essas avaliações, o algoritmo já interpretava como aprovado naquele teste, pode ter contribuído para essa Falsa Aceitação. Se o

deepfake apresentou mínimas variações de pixel na região ocular que foram erroneamente interpretadas como reações fisiológicas sutis pela esclera ou pupila (devido aos limiares lenientes desses sub-testes), isso, combinado com piscadas bem simuladas e a ausência de movimentos cefálicos complexos solicitados, pode ter levado ao escore suficiente para a aprovação indevida.

- **Desafios na Estimativa de Olhar Horizontal com Smartphone:** Foi uma observação durante o desenvolvimento que "a estimativa de olhar do usuário no sentido horizontal se torna mais difícil de observar com a câmera do smartphone, devido à proporção e tamanho da tela do celular". Embora os testes formais com DroidCam tenham resultado em 100% TAR, essa dificuldade intrínseca pode representar uma limitação em cenários de uso mais variados com dispositivos móveis, potencialmente exigindo algoritmos de estimativa de olhar mais sofisticados ou uma adaptação dinâmica dos alvos de interação para diferentes proporções de tela e distâncias de visualização.

4.4.4 Considerações Finais sobre os Resultados

Os resultados experimentais indicam que o protótipo de detecção de vivacidade desenvolvido possui uma base sólida, com bom desempenho tanto na aceitação de usuários legítimos quanto na rejeição da maioria das tentativas de fraude testadas. A arquitetura modular e a combinação de diferentes técnicas de análise (movimento da cabeça como pré-requisito, piscadas, interação com alvos visuais, reação pupilar e análise de cor da esclera) provaram ser uma estratégia válida.

A regra de decisão baseada em um escore de 100% para aprovação, embora rigorosa, ajudou a identificar rapidamente os sub-testes que necessitam de maior refinamento (como o teste de olhar na webcam). A leniência nos limiares de alguns testes passivos (esclera, pupila), adotada para fins de prototipagem, demonstrou ser uma faca de dois gumes: provavelmente auxiliou na alta taxa de aceitação de usuários legítimos, mas também pode ter sido um fator contribuinte para a única Falsa Aceitação por *deepfake*.

Os achados deste capítulo fornecem subsídios importantes para os próximos passos no desenvolvimento, incluindo o ajuste de parâmetros, o refinamento dos algoritmos de análise e a exploração de mecanismos de *antispoofing* adicionais para mitigar as vulnerabilidades identificadas, especialmente contra ataques de *deepfake* mais avançados.

5 Conclusão

Ao longo deste trabalho, foi abordado o desenvolvimento e a avaliação de um protótipo de sistema de detecção de vivacidade (*liveness detection*) com o objetivo de mitigar fraudes por apresentação em sistemas de autenticação biométrica facial. Desde a concepção teórica, passando pela implementação dos diversos módulos de análise e interação, até a realização de testes experimentais, buscou-se construir uma solução que integrasse múltiplas pistas de vivacidade para robustecer a verificação de presença real do usuário. Este capítulo final tem como propósito sintetizar as principais contribuições alcançadas por este projeto, bem como discutir, de forma crítica, as limitações identificadas e propor direções para trabalhos futuros que possam aprimorar e expandir a solução desenvolvida.

5.1 Principais Contribuições do Trabalho

Considera-se que este Projeto de Fim de Curso logrou êxito em apresentar contribuições relevantes tanto no âmbito do desenvolvimento prático de uma solução de prova de vida quanto na exploração de técnicas de visão computacional aplicadas à segurança biométrica. As principais contribuições podem ser resumidas da seguinte forma:

- **Desenvolvimento de um Protótipo Multifacetado de Detecção de Vivacidade:** Foi implementado um sistema funcional que integra um conjunto diversificado de técnicas de análise, incluindo a detecção e mapeamento de pontos fiduciais faciais, a validação contínua da pose cefálica como pré-requisito, a detecção de piscadas, a interação com alvos visuais para indução e verificação do movimento ocular, a análise da reação pupilar à luz e a verificação da cor da esclera.
- **Implementação de Mecanismos de Interação Específicos:** Destaca-se a concepção de interações como a solicitação explícita de piscadas naturais e, particularmente, o método de indução implícita do olhar através de alvos clicáveis na tela, que se mostrou uma forma intuitiva de verificar o direcionamento da atenção visual do usuário.
- **Criação de um Algoritmo Piloto Integrador com Múltiplas Visões:** O desenvolvimento de um algoritmo centralizador que orquestra todos os módulos de análise e que oferece uma "visão do desenvolvedor" (para depuração e análise detalhada) e uma "visão do usuário" (para interação simplificada) foi uma contribuição metodológica importante para a prototipagem e teste do sistema. Este algoritmo também implementou um sistema de escore básico para a tomada de decisão.

- **Obtenção de Resultados Experimentais Promissores:** Os testes iniciais, embora com escopo limitado, indicaram um bom desempenho do protótipo:
 - Elevada Taxa de Aceitação Verdadeira (TAR) para usuários legítimos, atingindo 100% com o uso de câmera de smartphone (via DroidCam) e 90% com webcam dedicada.
 - Rejeição completa (100% TRR) de ataques básicos utilizando imagens estáticas apresentadas em telas.
 - Uma considerável Taxa de Rejeição Verdadeira (90% TRR) contra vídeos *deepfake* gerados por IA, indicando resiliência a ataques dinâmicos mais sofisticados.
- **Identificação de Pontos de Melhoria e Vulnerabilidades:** A própria identificação de falhas, como a Falsa Rejeição devido à sensibilidade do teste de olhar em um cenário e a Falsa Aceitação de um vídeo *deepfake*, constitui uma contribuição, pois direciona os esforços futuros para o fortalecimento de componentes específicos do sistema.

Em conjunto, essas contribuições demonstram a viabilidade da abordagem proposta e fornecem uma base sólida para futuras pesquisas e desenvolvimentos na área de detecção de vivacidade.

5.2 Trabalhos Futuros e Recomendações

Apesar dos resultados encorajadores obtidos com o protótipo, é fundamental reconhecer suas limitações e identificar as avenidas para aprimoramento e pesquisa futura. A natureza de um Projeto de Fim de Curso implica, frequentemente, um escopo restrito para exploração exaustiva. Com base nisso, são propostas as seguintes direções:

5.2.1 Desenvolvimento de uma Interface de Usuário Comercial e Amigável

As interfaces desenvolvidas ("visão do desenvolvedor" e "visão do usuário") foram adequadas para os propósitos de prototipagem e teste funcional. No entanto, para uma aplicação comercial ou de uso disseminado, seria imprescindível o desenvolvimento de uma Interface Gráfica do Usuário (GUI) profissional, intuitiva e amigável. Tal interface deveria focar em:

- **Clareza e Simplicidade:** Instruções minimalistas e feedback visual imediato e compreensível para o usuário final.
- **Experiência do Usuário (UX):** Um fluxo de interação suave, que minimize o tempo de verificação e o esforço cognitivo do usuário.

- **Adaptabilidade:** Possibilidade de customização para diferentes contextos de aplicação (e.g., mobile, desktop, quiosques) e para atender a requisitos específicos de identidade visual de uma solução comercial.
- **Acessibilidade:** Considerar princípios de design acessível para usuários com diferentes necessidades.

5.2.2 Validação Experimental Extensiva e Robusta com Amostragem Significativa

Um ponto crítico a ser destacado é a **limitação do escopo dos testes experimentais realizados** neste trabalho. As avaliações foram conduzidas com um número reduzido de tentativas e, possivelmente, com uma diversidade limitada de participantes e condições. Embora os resultados forneçam indicativos iniciais da performance, eles carecem de significância estatística robusta para generalizações amplas sobre a eficácia do sistema. Para superar essa limitação, recomenda-se:

- **Aumento do Espaço Amostral:** Realizar testes com um número consideravelmente maior de usuários legítimos, abrangendo diferentes faixas etárias, gêneros, etnias (para verificar possíveis vieses nos algoritmos de detecção facial e de *landmarks*), e características como o uso de óculos, barbas, etc.
- **Diversificação das Condições de Teste:** Avaliar o sistema sob um espectro amplo de condições ambientais, incluindo variações significativas de iluminação (baixa luz, luz excessiva, sombras, diferentes tipos de fontes luminosas), fundos complexos e diferentes qualidades e tipos de câmeras.
- **Expansão e Sofisticação dos Cenários de Ataque:**
 - Testar contra uma gama maior de artefatos de apresentação (PAIs), incluindo impressões fotográficas de alta qualidade, máscaras 2D e 3D, e ataques de repetição utilizando telas de alta resolução.
 - Utilizar vídeos *deepfake* gerados por diferentes plataformas e com maior variabilidade de características e comportamentos, incluindo aqueles que tentam simular interações mais complexas.
 - Avaliar a resistência a ataques onde o fraudador tenta manipular ativamente os sensores ou explorar vulnerabilidades específicas de cada módulo.

Somente através de uma validação experimental mais extensa e metodologicamente rigorosa será possível obter métricas de desempenho (TAR, FRR, TRR, TFA) com intervalos

de confiança adequados e que reflitam de forma mais fidedigna a performance do sistema em cenários de uso real. A pequena amostragem atual pode enviesar os resultados, superestimando ou subestimando a real capacidade do protótipo.

5.2.3 Refinamento e Aprofundamento dos Algoritmos de Análise Ocular Antifraude

Os mecanismos de análise da cor da esclera e da reação pupilar à luz, embora tenham contribuído para os resultados, foram implementados com uma lógica de aprovação baseada na "menor demonstração de reação fisiológica", visando facilitar a prototipagem. Esta abordagem, como discutido, pode representar uma vulnerabilidade. Sugere-se um aprofundamento significativo nestes algoritmos:

- **Análise Quantitativa da Cor e Reflexibilidade da Esclera:**

- Em vez de apenas detectar variações, desenvolver modelos quantitativos das propriedades de cor e reflexão da esclera humana sob diferentes condições de iluminação (incluindo as luzes coloridas mencionadas anteriormente).
- Explorar técnicas de aprendizado de máquina para classificar texturas e padrões de reflexão da esclera, distinguindo tecido real de materiais impressos ou exibidos em tela.
- Investigar a normalização de cor para compensar variações na iluminação ambiente.

- **Análise Detalhada do Reflexo Fotomotor (RFM):**

- Ir além da simples detecção de constrição/dilatação. Implementar a análise quantitativa da curva do RFM: amplitude da constrição, velocidade de constrição e dilatação, tempo de latência.
- Otimizar o estímulo luminoso (intensidade, duração, cor) para eliciar uma resposta mais clara e mensurável, minimizando o desconforto.
- Desenvolver modelos para classificar os padrões de RFM como genuínos ou anômalos, considerando a variabilidade interpessoal e os efeitos da idade ou fadiga.
- Estudar a robustez da medição do diâmetro pupilar em diferentes condições.

O refinamento desses módulos pode aumentar significativamente sua contribuição para a segurança do sistema, tornando-os mais difíceis de serem enganados.

5.2.4 Outras Direções e Recomendações

Adicionalmente, outras linhas de pesquisa e desenvolvimento podem ser exploradas:

- **Exploração de Novos Indicadores de Vivacidade:** Investigar a inclusão de outras pistas biométricas ou comportamentais, como a análise de microtexturas da pele (anteriormente considerada), a estimativa de profundidade 3D da face de forma mais robusta, ou mesmo sinais fisiológicos sutis como a variabilidade da frequência cardíaca inferida remotamente (rPPG).
- **Otimização de Desempenho Computacional:** Focar na otimização dos algoritmos para execução eficiente em dispositivos com recursos limitados, como smartphones e sistemas embarcados, considerando o consumo de CPU, memória e energia.
- **Mecanismos Avançados de Fusão de Dados e Decisão:** Substituir o sistema de escore básico por técnicas mais sofisticadas de fusão de informações provenientes dos múltiplos módulos, como redes Bayesianas, lógica fuzzy, ou modelos de aprendizado de máquina (e.g., SVM, Random Forest, Redes Neurais) treinados para classificar uma sessão como genuína ou fraudulenta com base no vetor de características de vivacidade.
- **Sistema de Aprendizado Contínuo e Adaptação:** Desenvolver mecanismos para que o sistema possa se adaptar a novos tipos de ataques à medida que estes surgem, através de plataformas de treinamento e reavaliação contínua.
- **Estudo de Usabilidade e Aceitabilidade em Larga Escala:** Conduzir estudos com um público mais amplo para avaliar a percepção, aceitabilidade e usabilidade dos diferentes desafios de vivacidade, visando otimizar a experiência do usuário sem comprometer a segurança.

Em conclusão, este trabalho estabeleceu uma fundação promissora. As direções futuras aqui propostas visam não apenas sanar as limitações identificadas, mas também impulsionar a solução para um nível de maturidade e robustez compatível com as exigências de aplicações reais e os desafios impostos pela constante evolução das técnicas de fraude.

Referências

- Amazon Web Services. **Amazon Rekognition Face Liveness**. 2023. Acesso em: 19 maio 2025. Disponível em: <<https://aws.amazon.com/pt/rekognition/face-liveness/>>. Citado na página 19.
- _____. **AWS AI Service Cards: Responsible AI Guidelines**. [S.l.], 2023. Acesso em: 19 maio 2025. Disponível em: <<https://docs.aws.amazon.com/rekognition/latest/dg/face-liveness.html>>. Citado na página 18.
- _____. **Detect Real Users with Face Liveness**. 2023. Acesso em: 19 maio 2025. Disponível em: <<https://aws.amazon.com/blogs/machine-learning/detect-real-users-with-amazon-rekognition-face-liveness/>>. Citado na página 19.
- APPALARAJU, S.; CHAOJI, V. Image similarity using deep cnn and curriculum learning. In: **Grace Hopper Celebration India (GHCI)**. [S.l.: s.n.], 2017. Citado na página 26.
- BRADSKI, G. R. The opencv library. In: **Dr. Dobb's Journal of Software Tools**. [S.l.: s.n.], 2000. v. 25, n. 11, p. 120–125. Citado na página 44.
- Exato Digital. **Validação Facial DocCheck**. 2025. Acesso em: 19 maio 2025. Disponível em: <<https://blog.exato.digital/validacao-facial-doccheck/>>. Citado na página 17.
- FaceTec, Inc. **FaceTec.com | 3D Liveness, 3D Face Matching, UR Codes**. 2025. Acesso em: 19 maio 2025. Disponível em: <<https://www.facetec.com/>>. Citado 2 vezes nas páginas 16 e 17.
- FEBRABAN; DELOITTE. **Pesquisa FEBRABAN de Tecnologia Bancária - Volume 2**. 2024. <<https://www.febraban.org.br>>. Acesso em: 9 abr. 2025. Citado 2 vezes nas páginas 11 e 14.
- GALBALLY, J.; MARCEL, S.; FIERREZ, J. Biometric antispoofing methods: A survey. In: MARCEL, S.; NIXON, M. S.; FIERREZ, J.; EVANS, N. (Ed.). **Handbook of Biometric Anti-Spoofing: Trusted Biometrics under Spoofing Attacks**. [S.l.]: Springer London, 2014. p. 83–105. Citado 2 vezes nas páginas 11 e 60.
- Google for Developers. **MediaPipe Face Landmarker**. 2024. <https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker?hl=pt-br>. Acessado em: 31 de maio de 2025. Citado 3 vezes nas páginas 33, 34 e 35.
- HANSEN, D. W.; JI, Q. In the eye of the beholder: A survey of models for eyes and gaze. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, n. 3, p. 478–500, 2010. Citado na página 58.
- HARRIS, C. R.; MILLMAN, K. J.; WALT, S. J. van der; GOMMERS, R.; VIRTANEN, P.; COUNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; KERKWIJK, M. H. van; BRETT, M.; HALDANE, A.; RÍO, J. F. del; WIEBE, M.; PETERSON, P.; GÉRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E.

Array programming with numpy. **Nature**, v. 585, n. 7825, p. 357–362, 2020. Citado na página 44.

Instituto DataSenado. **Panorama Político 2024: apostas esportivas, golpes digitais e endividamento**. 2024. <<https://www12.senado.leg.br/noticias/materias/2024/10/01/golpes-digitais-atingem-24-da-populacao-brasileira-revela-datasenado>>. Acesso em: 14 abr. 2025. Citado na página 15.

iProov Limited. **iProov: Verificação Facial Segura e Autenticação Biométrica**. 2025. Acesso em: 19 maio 2025. Disponível em: <<https://www.iproov.com/pt-br/>>. Citado 2 vezes nas páginas 17 e 18.

_____. **Relatório de Inteligência de Ameaças 2025**. [S.l.], 2025. Acesso em: 19 maio 2025. Disponível em: <<https://www.iproov.com/press/annual-identity-verification-threat-intelligence-report>>. Citado na página 17.

JIN, X.; YE, D.; CHEN, C. Análise forense e segurança facial baseada em dados: Combatendo a falsificação: para detectar deepfake com sinais biológicos multidimensionais. **Security and Communication Networks**, Hindawi (Traduzido), v. 2021, 2021. Citado 7 vezes nas páginas 5, 23, 26, 27, 28, 36 e 37.

KAMYAB, S.; AZIMIFAR, Z. Deep-MDS framework for recovering the 3d shape of 2d landmarks from a single image. **Journal of Visual Communication and Image Representation**, Elsevier, v. 98, p. 104032, 2024. Citado 2 vezes nas páginas 32 e 33.

LANGGUTH, J.; POGORELOV, K.; BRENNER, S.; FILKUKOVÁ, P.; SCHROEDER, D. T. Don't trust your eyes: Image manipulation in the age of deepfakes. **Frontiers in Communication**, Frontiers Media SA, v. 6, p. 632317, 2021. Citado 11 vezes nas páginas 22, 23, 24, 25, 26, 28, 35, 36, 37, 38 e 39.

LEPETIT, V.; MORENO-NOGUER, F.; FUA, P. EPnP: An accurate O(n) solution to the PnP problem. In: **International Journal of Computer Vision**. [S.l.: s.n.], 2009. Citado na página 53.

LUGARESI, C.; TANG, J.; NASH, H.; MCCLANAHAN, C.; PINTO, E.; ADELI, M.; CHANG, Y. J. K. T. G. L. R. R. G. M. G. M. A. V. U. S. G. P. W. C.; YONG, K.; GLLVEZ-RLVA, G.; GRISHCHENKO, I.; GRUNDMANN, M. Mediapipe: A framework for building perception pipelines. **arXiv preprint arXiv:1906.08172**, 2019. Citado 5 vezes nas páginas 41, 42, 44, 50 e 55.

MEDEIROS, G. C. F. **Uso de texturas para o acompanhamento da evolução do tratamento de úlceras dermatológicas**. Dissertação (Dissertação de Mestrado) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2001. Acesso em: 31 maio 2025. Citado 2 vezes nas páginas 21 e 22.

MURPHY-CHUTORIAN, E.; TRIVEDI, M. M. Head pose estimation in computer vision: A survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 4, p. 607–626, 2009. Citado na página 53.

National Institute of Standards and Technology. **National Voluntary Laboratory Accreditation Program (NVLAP)**. 2023. Acesso em: 19 maio 2025. Disponível em: <<https://www.nist.gov/nvlap>>. Citado na página 16.

SALLES, L. A.; FILHO, P. L. de P.; JUNIOR, A. C. Estimação de idade em imagens digitais a partir de deep learning para apoiar análise pericial. **Anais do XV Workshop de Visão Computacional (WVC)**, p. 80–85, 2021. Citado 2 vezes nas páginas 25 e 26.

SOUKUPOVÁ, T.; CECHE, J. Real-time eye blink detection using facial landmarks. In: HANBURY, A. ; PULTZ, T. (Ed.). **Proceedings of the 21st Computer Vision Winter Workshop (CVWW)**. [S.l.: s.n.], 2016. p. 1–8. Citado na página 56.

STEHOUWER, J.; DANG, H.; LIU, F.; LIU, X.; JAIN, A. On the detection of digital face manipulation. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops**. [S.l.: s.n.], 2019. p. 0–0. Citado 6 vezes nas páginas 25, 29, 30, 31, 36 e 38.

SUN, Z.; TAN, T.; WANG, Y. Iris liveness detection based on pupillary light reflex. In: **Proceedings of the International Conference on Biometrics (ICB)**. [S.l.]: Springer Berlin Heidelberg, 2005. (Lecture Notes in Computer Science, v. 3832), p. 588–594. Citado na página 64.

VIOLA, P.; JONES, M. J. Rapid object detection using a boosted cascade of simple features. In: **Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001**. [S.l.: s.n.], 2001. v. 1, p. I–511–I–518 vol.1. Citado 2 vezes nas páginas 47 e 49.

WU, H.; CHEN, Y.; ZHOU, J. **Rethinking Image Forgery Detection via Contrastive Learning and Unsupervised Clustering**. 2023. Acesso em: [insira data]. Disponível em: <<https://arxiv.org/abs/2308.09307>>. Citado 3 vezes nas páginas 20, 21 e 36.

ZHOU, P.; HAN, X.; MORARIU, V. I.; DAVIS, L. S. **Learning Rich Features for Image Manipulation Detection**. 2018. Acesso em: [insira data]. Disponível em: <<https://arxiv.org/abs/1805.04953>>. Citado 4 vezes nas páginas 20, 21, 22 e 25.