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

LUCAS ROYES SCHARDOSIM

CONTRAMEDIDAS PARA EVITAR A FALSIFICAÇÃO DO USUÁRIO NO ACESSO A SISTEMAS BIOMÉTRICOS VIA IMAGENS E VÍDEOS

Porto Alegre
2018

LUCAS ROYES SCHARDOSIM

CONTRAMEDIDAS PARA EVITAR A FALSIFICAÇÃO DO USUÁRIO NO ACESSO A SISTEMAS BIOMÉTRICOS VIA IMAGENS E VÍDEOS

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Engenharia da Computação

ORIENTADOR: Prof. Dr. Jacob Scharcanski

Porto Alegre
2018

LUCAS ROYES SCHARDOSIM

CONTRAMEDIDAS PARA EVITAR A FALSIFICAÇÃO DO USUÁRIO NO ACESSO A SISTEMAS BIOMÉTRICOS VIA IMAGENS E VÍDEOS

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

Orientador: _____
Prof. Dr. Jacob Scharcanski, UFRGS
Doutor pela University of Waterloo – Waterloo, Canadá

Banca Examinadora:

Prof. Dr. André Lima Férrer de Almeida, DETI-UFC
Doutor pela Université de Nice Sophia Antipolis – França

Prof. Dr. Maciel Zortea, Natural Resources Analytics - IBM Research Brazil
Doutor pela Universidade de Gênova – Itália

Prof. Dr. Alexandre Balbinot, PPGEE-UFRGS
Doutor pela UFRGS – Porto Alegre, Brasil

Prof. Dr. Valner João Brusamarello, PPGEE-UFRGS
Doutor pela UFSC – Florianópolis, Brasil

Prof. Dr. Edison Pignaton de Freitas, PPGEE-UFRGS
Doutor pela UFRGS – Porto Alegre, Brasil

Coordenador do PPGEE: _____
Prof. Dr. João Manoel Gomes da Silva

Porto Alegre, dezembro de 2018.

DEDICATÓRIA

Dedico este trabalho aos meus pais Caetano e Maria, irmã Chris, esposa Kellen, primos, amigos e tias, em especial, pela dedicação e apoio em todos os momentos.

AGRADECIMENTOS

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

Aos colegas do PPGEE e PPGC pelo seu auxílio nas tarefas desenvolvidas durante o curso.

Ao Prof. Dr. Jacob Scharcanski.

Ao pessoal da secretaria do PPGEE e em especial à Miriam Rosek.

À Natália Cecília Rebelo pelas inúmeras revisões das referências.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

RESUMO

A autenticação de usuário é um passo importante para proteger as informações e neste campo a biometria da face é vantajosa. A biometria do rosto é natural, fácil de usar e menos invasiva. Esta tese complementa a biometria facial e visa desenvolver métodos de contramedidas capaz de detectar tentativas de acesso por usuário não autorizado fraudando a identidade de um usuário autorizado. Apesar do grande sucesso alcançado em biometria da face nas últimas décadas, pouca atenção tem sido dada ao problema crítico de ataque de apresentação, *spoofing attack*. Somente nos últimos anos que gradualmente os sistemas de reconhecimento facial estão cientes da vulnerabilidade dos ataques de apresentação. Fraudadores não autorizados podem tentar falsificar os sistemas de reconhecimento facial exibindo cópias falsificadas do rosto de um cliente autorizado, tais como fotos ou vídeos. Embora simples, estes ataques são geralmente muito bem sucedidos. Nesta tese foram desenvolvidas contramedidas capazes de detectar ataques de apresentação à sistemas biométricos faciais, sendo criadas três abordagens distintas para a detecção de ataques de apresentação. As metodologias utilizam informações de movimento, energia de deformação das faces, descritores de texturas e esteganoanálise adequadamente projetados para um espaço de dimensão reduzido através de projeções aleatórias, análise de componentes principais e análise discriminante linear. As abordagem são capazes de detectar ataques a cada quadro de um vídeo e, para o estudo de caso, duas bases de dados foram avaliadas. Uma delas é a base de dados do Centro de Biometria e Pesquisa em Segurança da Academia Chinesa de Ciências, CASIA. E a outra, é a base de dados de fotografia de impostor da Universidade Nanjing de Aeronáutica e Astronáutica, NUAA. Através de um classificador de redes neurais artificiais a melhor metodologia desenvolvida alcança uma taxa de erro $HTER = 4.93\%$ para NUAA e $HTER = 6.51\%$ para CASIA.

Palavras-chave: Anti spoofing, contramedidas, falsificação, face, vídeo, imagem.

ABSTRACT

User authentication is an important step in protecting the information and in this field the face biometry is advantageous. Face biometry is natural, easy to use and less invasive. The proposal of this work complements facial biometrics and aims to develop a method of countermeasures capable of detecting unauthorized access attempts by fraudulating the identity of an authorized user. Despite the great success achieved in face biometrics in recent decades, little attention has been paid to the critical problem of spoofing attack. It is only in recent years that facial recognition systems are gradually aware of the vulnerability of presentation attacks. Unauthorized fraudsters may attempt to falsify facial recognition systems by displaying fake copies of an authorized customer's face, such as photos or videos. Though simple, these attacks are usually very successful. In this thesis were developed countermeasures capable of detecting presentation attacks to biometric facial systems. Three distinct approaches have been created for the detection of presentation attacks. The methodologies use motion information, face deformation energy, texture descriptors and steganoanalysis properly projected for a small dimension space through random projections, principal component analysis and linear discriminant analysis. The approaches are able to detect spoof attacks on each frame of a video and, for case study, two databases were evaluated. One of them is the database of the Center for Biometry and Security Research of the Chinese Academy of Sciences, CASIA. And the other, is the impostor photography database of Nanjing University of Aeronautics and Astronautics, NUAA. Through an artificial neural network classifier the best developed methodology achieves an error rate of $HTER = 4.93\%$ for NUAA and $HTER = 6.51\%$ for CASIA.

Keywords: anti spoofing, countermeasures, fake, face, image, video .

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS

CASIA *Chinese Academy of Sciences*, Academia Chinesa de Ciências Astronáutica

FAR *False Acceptance Rate*, Taxa de falsa aceitação

FRR *False Rejection Rate*, Taxa de falsa rejeição

FP *False Positive*, Falso Positivo

FN *False Negative*, Falso Negativo

HTER *half Total Error Rate*, Taxa da metade do erro total

IDIAP *Istituto Dalle Molle di Intelligenza Artificiale Percettiva*, Instituto Dalle Molle de Inteligência Artificial Preceptiva

NUAA *Nanjing University of Aeronautics and Astronautics*, Universidade Nanjing de Aeronáutica e Astronáutica

RGB Espaço de cor *Red, Gree, Blue*

TP *True Positive*, Verdadeiro Positivo

TN *True NEgative*, Verdadeiro Negativo

TPS Thin-Plate Spline

PAD *Presentation Attack Detection*, detecção de ataque de apresentação, *presentation attack detection*, detecção de ataque de apresentação.

ANOVA *Analysis of variance*, análise de variância

SUMÁRIO

1 INTRODUÇÃO

1.1 Motivação

Sistemas biométricos já constituem um componente significante das tecnologias de identificação atuais e emergentes (HADID *et al.*, 2015). Essas tecnologias visam determinar e/ou verificar a identidade de um indivíduo a partir de características de comportamento e/ou características biológicas.

Os autores BIGGIO *et al.* (2015) caracterizam ataques e contramedidas para sistemas biométricos de acordo com a técnica de ataque, localização do ataque e tipo de defesa, conforme a Tabela 1. Para cada técnica de ataque é reportado um componente alvo (localização do ataque) e a respectiva defesa.

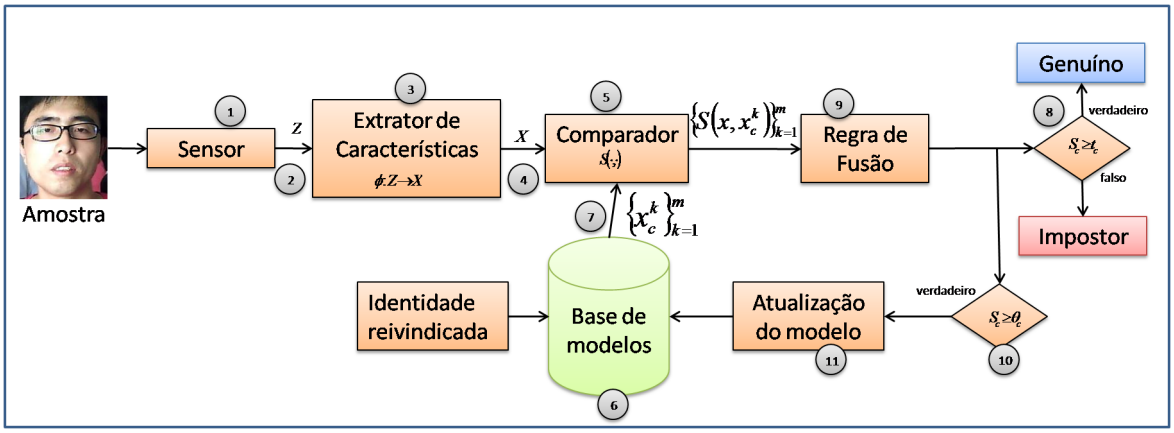
Um exemplo de arquitetura para um sistema de verificação biométrica e seus pontos correspondentes de ataque, destacados com números em círculos cinzas, está ilustrado no diagrama da Fig 1. Durante a verificação biométrica, uma imagem de face $z \in Z$ adquirida pelo sensor é processada por um extrator de características $\phi : Z \in X$ para obter uma representação compacta $x \in X$. Os modelos $\left\{x_c^k\right\}_{k=1}^m$ da identidade reivindicada c são recuperados da base de modelos, e comparados com x usando um algoritmo de correspondência $S : X \times X \in \Re$. As pontuações resultantes $\left\{S\left(x, x_c^k\right)\right\}_{k=1}^m$ são combinadas por uma regra de fusão, produzindo uma pontuação agregada $S_c(x)$ que expressa o grau em que x provavelmente pertence a c. A pontuação $S_c(x)$ é então comparada com uma limiar de decisão t_c para decidir se a amostra é genuína ou impostora. Se uma autoatualização do modelo for implementada e $S_c(x)$ for que um limiar θ_c, um dos modelos em $\left\{S\left(x, x_c^k\right)\right\}_{k=1}^m$ é atualizado.

Nesse exemplo de arquitetura e de acordo com a classificação sugerida pelos autores BIGGIO *et al.* (2015), ataques de apresentação (*spoofing attack*) consiste em fabricar um traço biométrico para representar um cliente inscrito. A localização ocorre ao nível do sensor (ponto 1 na Figura 1), este tipo de ataque também se refere como ataque direto. As defesas atuais são fundamentadas em métodos de detecção de vivacidade. Este tipo de método tem por objetivo verificar se o traço submetido ao sistema biométrico é 'vivo' ou 'falso'[1] verificando padrões característicos. Sistemas multibiométricos foram propostos para a defesa destes tipos de fraude. Utilizando uma regra de fusão segura entre diferentes feições, por exemplo reconhecimento facial + leitura de impressão digital (RODRIGUES; LING; GOVINDARAJU, 2009).

Ataques de repetição podem ocorrer nas interfaces entre os módulos repetindo uma imagem roubada da característica biométrica do cliente alvo para o extrator de característica (ponto 2), ou diretamente da característica correspondente para o comparador(ponto

[1]Tradução dos termos utilizados por Biggio *et al* (2015, p. 34) *alive* e *fake*, respectivamente.

Figura 1: Arquitetura de um sistema de verificação biométrico.



Fonte: (BIGGIO *et al.*, 2015)

4). Um atacante pode até mesmo repetir um sinal para substituir as características de um determinado modelo da identidade reivindicada (ponto 7). Este ataque pode ser claramente encenado se os correspondentes canais de comunicação são inseguros, mas também através de canais criptografados; desta forma o sinal codificado pode ser roubado e substituído no canal diretamente. Isto pode ser evitado através da encriptação de uma marca de tempo no sinal. Outra possível contramedida é o isolamento físico, para evitar o envio de dados através de canais inseguros (como a Internet). Um exemplo popular de isolamento físico é o uso de cartão inteligente executando operações *match-on-card*. No entanto, esta técnica tem as suas desvantagens, incluindo as limitações em termos de recursos computacionais e de memória, além do fato que o usuário deve sempre usar o cartão inteligente para ser autenticado (JAIN; NANDAKUMAR; NAGAR, 2008).

Ataques do tipo *hill climbing* são semelhantes ao método de repetição, pois afetam canais inseguros de comunicação, em particular os pontos 2 e 4 da Figura 1. Seu objetivo é reconstruir um modelo de imagem iterativamente, enviando diversas imagens ligeiramente alteradas ao extrator de características (ponto 2) ou as suas características enviadas diretamente ao comparador (ponto 4), retendo aquelas que maximizam a pontuação do casamento $S_c(x)$, onde x é a imagem atual ou um conjunto de características. Na prática, é uma técnica gradiente ascendente na qual aproxima o gradiente de $S_c(x)$ numericamente. Neste caso, assume-se que o atacante possa ser capaz de observar $S_c(x)$ para qualquer imagem consultada, o que só pode ser viável se o sistema oferece (ou vaza) tais informações. Além da proteção de canal acima mencionada, um mecanismo de defesa adicional consiste na quantização da pontuação de casamento pelo comparador capaz de prover informação menos precisas ao atacante.

A implementação dos algoritmos (pontos 3, 5, 8, 9, 10 e 11) pode exibir vulnerabilidades que podem ser exploradas por atacantes habilidosos através da instalação de programas maliciosos, isto é, *malware*, incluindo vermes, cavalos de troia e outros. Este problema pode ser evitado ou minimizado utilizando programação segura, ou *hardware* especializado para executar algumas operações críticas. Uma prática segura de programação é verificar a integridade do algoritmo, isto é, cada algoritmo ou função que manipula uma variável de entrada e nunca retorna um comportamento inesperado. Por exemplo, o comparador espera um vetor de entrada x, $x \in \mathbb{R}^3$, e ao contrário recebe uma entrada em um formato diferente. Neste tipo de situação é importante saber se o algoritmo irá responder ao erro de forma a permitir o acesso ou não.

Tabela 1: Categorização de ataques e contramedidas para Sistemas Biometricos, (BIGGIO *et al.*, 2015).

Técnica de Ataque	Localização do Ataque	Ponto(s) de ataque	Defesa
Apresentação	Sensor	1	Detecção de vivacidade, Multi biométrico
Repetição	Interface/Canais	2, 4, 7	Encriptação de Canal, Isolação Física
hill Climbing	Interface/Canais	2, 4	Encriptação de Canal, Isolação Física, Quantização da Pontuação
Infestação por *Malware*	Modulos/ Algoritmos	3,5,8-11	Código de Segurança, *hardware* especializado, Algoritmo de Integridade
Roubo, Substituição e Eliminação de Modelo	Modelo da Base de Dados	6	Encriptação de Modelo, Modelos Canceláveis/Revogáveis

Roubo, substituição e eliminação de modelo atacam diretamente a base de dados dos modelos (ponto 6 na Figura1). Se os modelos não estão propriamente protegidos, alguém pode ser capaz de roubá-los e usá-los para criar um modelo falso para realizar um ataque de repetição ou personalizar um cliente alvo em um sistema diferente e realizar diferentes operações, por exemplo, pesquisar em bases de dados protegidas. Uma outra possibilidade é o atacante adicionar suas próprias digitais e colocá-las no modelo de outro cliente. Adicionalmente, modelos de determinado cliente podem ser apagados para causar uma negação de serviço, ou seja, para evitar que o cliente alvo seja reconhecido corretamente, contramedidas incluem encriptação de modelos e também o uso de modelos canceláveis/revogáveis, que podem ser utilizados somente em sistemas específicos e reeditado se roubados.

Trabalhos recentes revelam que sistemas biométricos faciais são vulneráveis a ataques de falsificação de usuário, ou seja, ataques de apresentação (FREITAS PEREIRA *et al.*, 2013). Na literatura internacional (TAN *et al.*, 2010; ANJOS; MARCEL, 2011; CHINGOVSKA; ANJOS; MARCEL, 2012; FREITAS PEREIRA *et al.*, 2013; GALBALLY; MARCEL; FIERREZ, 2014; HADID *et al.*, 2015; MANJANI *et al.*, 2017), o termo *spoofing attack* é utilizado largamente para este tipo de fraude. Além do acrônimo PAD, do inglês *presentation attack detection*, detecção de ataque de apresentação. Um exemplo disso é um fraudador que tenta enganar um sistema de reconhecimento facial apresentando uma fotografia, um vídeo, ou mesmo uma máscara tridimensional buscando assemelhar-se ao indivíduo legítimo. A Figura 2 ilustra um ataque de apresentação utilizando uma foto impressa, a Figura 2(a) é um acesso genuíno a um sistema biométrico facial, enquanto a Figura 2(b) é uma tentativa de ataque utilizando uma fotografia de um cliente cadastrado. Embora simples, esse ataque é geralmente muito bem sucedido.

Esse tema sobre detecção de ataques de apresentação ilustra a capa da revista (FRONT COVER, 2015) mostrando a relevância do tema na atualidade, por questões que envolvem segurança ao acesso de dados. Por isso novas bases de dados estão sendo desenvolvidas, por exemplo a base OULU-NPU: *A mobile face presentation attack with real world variations* (BOULKENAFET *et al.*, 2017) foi criada para a competição *International Joint Conference on Biometrics* (IJCB) 2017. Já a base de dados REPLAY-MOBILE *Face presentation-attack* (COSTA-PAZO *et al.*, 2016) foi criada para capturar características específicas dos dispositivos móveis. Também, devido à importância desse tema, foi estabelecido uma ISO/IEC DIS 30107-3 que reúne os princípios e métodos para a avaliação do desempenho dos mecanismos de detecção de ataques de apresentação.

Isso evidencia a relevância do tema e a contínua busca por metodologias mais eficazes. Sendo assim, esta tese envolve levantamento bibliográfico, desenvolvimento de contramedidas e testes experimentais de detecção de ataques de apresentação. Tudo isso

Figura 2: Exemplo de um ataque de apresentação. (a) acesso genuíno e (b) tentativa de ataque.

(a) (b)

Fonte: (TAN *et al.*, 2010)

com o intuito de analisar, classificar e interpretar ataques a sistemas biométricos faciais via imagens e vídeos em larga escala, de maneira eficiente, robusta e automática.

1.2 Objetivos

Essa tese objetivou desenvolver três novas metodologias capazes de detectar ataques a sistemas biométricos faciais no contexto de ataques de apresentação.

A primeira abordagem deste trabalho teve como objetivo específico desenvolver um método de contra-ataques de apresentação específico por pessoa. Esta metodologia pode ser resumida da seguinte forma. Dado um vídeo da base de dados, define-se um quadro de referência I, escolhe-se uma face neutra e os quadros seguintes são definidos por J. A extração do rosto é realizada via *deep learning* (ZHANG *et al.*, 2016), mais informações sobre esse método pode-se encontrar na Seção 3.1.6.2. A seguir, a face é dividida em k blocos. Cada bloco J_k do quadro J tem o seu movimento estimado em relação aos blocos do quadro de referência I_k através do método proposto pelo autor em (SCHARDOSIM; SCHARCANSKI, 2017) e aplicando transformações não afins do tipo *Thin-Plate Spline* (TPS). A correspondência entre blocos sempre ocorre entre o quadro de referência e o quadro atual J. Para os blocos correspondentes de J_k são extraídas as texturas usando um extrator aqui desenvolvido chamado de Log-Radial Diff (LRD). As texturas extraídas são projetadas em uma dimensão reduzida com projeções aleatórias, *Random Projections* (RP) (ACHLIOPTAS, 2003) e classificadas utilizando *Support Vectors Machine* (SVMs), Máquina de vetores de suporte (CRISTIANINI; SHAWE-TAYLOR, 2000). Mais detalhes sobre SVMs está na Secão 3.1.6.1. Com a utilização do classificador SVMs são construídos os modelos de classe para as faces genuínas e classes de ataques: ataque de foto, ataque de foto com corte nos olhos e ataque de vídeo, conforme detalhado na Seção 3.1.

Já a segunda metodologia teve como objetivo específico abordar o problema de detecção de ataque de falsificação da face de um usuário como um problema de classificação binária (SOLDERA *et al.*, 2017), de modo que as estatísticas de todo o conjunto de imagens, consistindo de ambos os rostos humanos vivos e fotografias das faces, podem ser totalmente explorados. Inicialmente, cada face é alinhada usando uma transformação não linear aplicada aos pontos de referência da face detectada. Posteriormente, cada rosto é

dividido em blocos e características locais são extraídas de cada bloco representando sua cor e textura local. Usando uma função de perda e SVMs como um classificador supervisionado, o melhor conjunto de blocos para comparar imagens faciais são selecionados durante a fase de treinamento e utilizados na fase de testes. Mais detalhes estão descritos na Seção 3.2.

Enquanto que o terceiro objetivo específico consiste em detectar ataques de apresentação utilizando características de movimento, texturas e esteganoanálise. Essas características são projetadas um espaço discriminante e alimenta um classificador binário cuja resposta será uma etiqueta indicando se a face é genuína ou impostora. Dado um vídeo da base de dados, define-se um quadro de referência I, escolhe-se uma face neutra e os quadros seguintes são definidos por J. São extraídos 68 pontos característicos da face (olhos, boca, nariz, sobrancelha e perímetro) utilizando o método (KAZEMI; SULLIVAN, 2014). Uma vez que a face é delimitada com estes pontos, diversas características são extraídas. Para informação de movimentos é calculado a energia de deformação através da transformação *Thin Plate Spline* (TPS) entre cada quadro J e o quadro de I. Enquanto as informações de texturas são extraídos utilizando um descritor de Quantização de Fase Local (*Local Phase Quantization*) (LPQ) (OJANSIVU; HEIKKILÄ, 2008) e métodos de esteganografia (CHEN; SHI, 2008; PEVNY; BAS; FRIDRICH, 2010). As características de movimento e textura formam um vetor característico e podem ser projetadas em uma dimensão reduzida utilizando as técnicas de PCA e/ou LDA. Após isto são classificadas. Utilizando essa abordagem são construídos os modelos de classe para as faces genuínas e ataques. Essa metodologia está detalhada na Seção 3.3.

1.3 Desenho da pesquisa

Esta tese, quanto à natureza, parte de pesquisa aplicada na medida em que prevê gerar novos conhecimentos com finalidades imediatas, utilizando os dados da literatura especializada e as tecnologias existentes. Quanto aos objetivos (descritos na Seção 1.2), é pesquisa exploratória, com procedimento explicativa (PRODANOV; FREITAS, 2013). Esta pesquisa envolve levantamento bibliográfico, testes experimentais de detecção de ataques de apresentação com informações disponíveis em bases de dados. De forma a analisar, classificar e interpretar ataques a sistemas biométrico via imagens e vídeos.

Esta abordagem difere-se de propostas de validação de dados biométricos da face porque autentica a forma como os dados são apresentados. Para isto, estão sendo utilizadas bases de dados internacionais tais como a base de dados de fotografia de impostor da Universidade Nanjing de Aeronáutica e Astronáutica, *Nanjing University of Aeronautics and Astronautics* (NUAA) (TAN *et al.*, 2010) e a base de dados contra fraude de faces do Centro de Biometria e Pesquisa de Segurança, *Center for Biometrics and Security Research* (CBSR) (ZHANG *et al.*, 2012). Considerando esses dados, as perguntas de pesquisa são:

1. É possível detectar vivacidade e verificar fraudes a sistemas biométricos quando apresenta-se uma imagem ou vídeo?
2. É possível desenvolver contramedidas eficazes para ataques de apresentação?
3. É possível encontrar características capazes de discriminar as classes de faces genuínas e *spoof*?

A partir desses questionamentos são formadas as seguintes hipóteses:

1. As metodologia proposta na primeira, segunda e terceira abordagens serão adequadas para detectar ataques de apresentação;

2. O extrator de feição proposto LRD será adequado para detectar ataques de apresentação;

3. As características extraídas através da análise esteganográfica são capazes de discriminar as classes de faces genuínas e impostoras.

1.4 Contribuições

Considerando os objetivos, as perguntas de pesquisa e as hipóteses, as contribuições desta de tese são:

- Desenvolvimento de um método de compensação de movimento frente a deformações afins (SCHARDOSIM; SCHARCANSKI, 2017) utilizado parcialmente na contramedida 1 (Seção 3.1);

- Desenvolvimento de um novo extrator de textura chamado de Log-Radial Diff (LRD), utilizado na contramedida 1 (Seção 3.1);

- Desenvolvimento de um método de compensação de movimento frente a deformações não afins utilizando transformações do tipo *Thin-Plate Spline* (TPS), utilizado nos métodos 1 e 2 (Seções 3.1 e 3.2);

- Desenvolvimento de uma metodologia básica para detecção de ataques de apresentação (SOLDERA *et al.*, 2017), contramedida 2 (Seção 3.2);

- Utilização da TPS para calcular a energia de deformação das faces, constitui uma das características empregadas no método 3 (Seção 3.3);

- Utilização da esteganoanálise para PAD, constitui uma das características empregadas no método 3 (Seção 3.3);

- Totalizando o desenvolvimento de três novas metodologias para PAD;

- Participação em artigo de conferência aceito para a I2MTC 2019.

1.5 Organização do Texto

Essa tese está organizada em cinco capítulos. Na introdução há a motivação, os objetivos, hipóteses e contribuições, bem como a organização do texto. O segundo capítulo é da revisão bibliográfica e do estado da arte sobre contramedidas para evitar a falsificação de usuário. Já o capítulo três apresenta as metodologias desenvolvidas de PAD no acesso a sistemas biométricos faciais. No capítulo seguinte são apresentados os resultados e discussões. Na sequência são expostas as conclusões e trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA E ESTADO DA ARTE

Os autores (WEN; HAN; JAIN, 2015) relatam que os primeiros estudos sobre detecção de fraude em sistemas biométricos faciais foram reportados em 2004 por (LI *et al.*, 2004). Com o crescimento de popularidade no uso de reconhecimento facial para controle de acesso, este tópico tem atraído significativamente a atenção nos últimos quatro anos (TAN *et al.*, 2010; ANJOS; MARCEL, 2011; CHINGOVSKA; ANJOS; MARCEL, 2012; FREITAS PEREIRA *et al.*, 2013; GALBALLY; MARCEL; FIERREZ, 2014; HADID *et al.*, 2015). Diversas técnicas foram desenvolvidas e é possível compará-las de acordo com as técnicas utilizadas, bem como seus pontos fortes e limitações, conforme pode ser visto na Tabela 2.

Já os autores (CHAKKA *et al.*, 2011) propuseram uma classificação fundamentada nas seguintes pistas:

- Presença de vitalidade;
- Diferença nos padrões de movimento;
- Diferença na qualidade de imagem.

Presença de vitalidade ou detecção de vivacidade consiste em procurar características que somente faces vivas possuem. Por exemplo, (PAN *et al.*, 2007) explora a observação que humanos piscam uma vez entre 2s a 4s e propuseram uma contramedida baseada em na detecção de piscada.

Contramedidas fundamentadas nas diferenças de padrões de movimento dependem do fato que faces reais apresentam um comportamento de movimento diferente comparado a uma tentativa de fraude. Os autores (KOLLREIDER; FRONTHALER; BIGUN, 2009; ANJOS; CHAKKA; MARCEL, 2014) utilizam o método de fluxo óptico para avaliar o movimento das faces ao longo dos quadros de vídeo.

Contramedidas baseadas na avaliação das diferenças de qualidade das imagens dependem na presença de artefatos intrínsecos presentes na mídia de ataque. Tais características podem ser originadas da qualidade da mídia ou diferenças nas propriedades de refletância do objeto exposto à câmera. Por exemplo, em (WEN; HAN; JAIN, 2015) ataque de apresentação são detectados através da análise de distorção de imagem utilizando quatro características distintas (reflexão especular, borramento, momento cromático e diversidade de cores).

2.1 Métodos Fundamentados em Movimento e Textura

Métodos baseados em movimento são projetados principalmente para conter ataques por foto. A ideia principal é capturar pistas importantes de vitalidade, tais como movimen-

Tabela 2: Comparação entre Diferentes Métodos Contra Ataques de Apresentação (WEN; HAN; JAIN, 2015)

Técnica utilizada	Pontos fortes	Limitações
Movimentos	Boa habilidade de generalização	Pouca robustez (pode ser logrado com movimento falso)
Textura	Rápida resposta (<1s), Baixa complexidade computacional	Pouca habilidade de generalização (vulnerável para variações das condições na aquisição)
Análise da qualidade de imagem	Boa generalização, rápida resposta (<1s), baixa complexidade computacional	Precisa de diferentes classificadores para diferentes tipos de ataques
Outras pistas	Muito robusto	Sensor adicional ou técnica de processamento necessária (Infra vermelho, áudio, 3D), resposta lenta(>3s)

tos subconscientes dos órgãos e músculos da face. Por exemplo o piscar de olhos (SUN *et al.*, 2007), movimento da boca (KOLLREIDER *et al.*, 2007) e a rotação da cabeça (BAO *et al.*, 2009). Dado que o movimento é relativo a característica ao longo dos quadros do vídeo, esses métodos tendem a ter uma maior generalização do que métodos fundamentados em texturas, conforme será descrito a seguir. No entanto, métodos baseados em movimento podem ser facilmente confundidos por outros movimentos, por exemplo, movimento de fundo que é irrelevante à vivacidade da face. Adicionalmente o método pode ser burlado quando o ataque é a apresentação de um vídeo ao invés de uma foto.

A detecção de ataques de apresentação através da análise de movimento visa detectar pistas geradas quando falsificações bidimensionais são apresentadas à câmera do sistema biométrico, por exemplo, fotos ou videoclipes (ANJOS; MARCEL, 2011). Esses métodos são projetados principalmente para combater ataques utilizando fotografias, capturam importantes pistas para a vitalidade: o movimento subconsciente órgãos e músculos em um rosto vivo, como piscar de olhos (SUN *et al.*, 2007), movimento da boca (KOLLREIDER; FRONTHALER; BIGUN, 2009) e rotação da cabeça (BAO *et al.*, 2009). Um dado movimento é uma característica relativa aos quadros de um vídeo, espera-se que esses métodos tenham uma melhor habilidade de generalização do que os métodos baseados em textura. No entanto, as limitações dos métodos fundamentados em movimento são aparentes. A frequência de movimento facial é restringida pelo ritmo fisiológico humano, que varia de 0.2 a 0.5 hz (BHARADWAJ *et al.*, 2013).

Portanto, é preciso um tempo relativamente longo (geralmente> 3s) para acumular recursos de vitalidade para detecção de falsificação de rosto. Além disso, métodos baseados em movimento podem ser facilmente contornados ou confuso por outros movimentos, por exemplo, movimento de fundo, que são irrelevantes para a vivacidade facial ou movimento através de um ataque por meio de vídeo.

Objetos planares se moverão de maneira diferente do que faces humanas que são objetos 3-D. Esses padrões de deformação podem ser usados para detecção de falsificação (ANJOS; MARCEL, 2011). Por exemplo (TAN *et al.*, 2010) explora o modelo de refletividade de Lambert para derivar diferenças entre imagens 2-D da face apresentadas durante um ataque e um acesso real(3-D) da face. Ele faz isso derivando uma equação que estima as informações reflexão existentes das imagens capturadas em ambos os cenários usando um método baseado em *retinex* variacional ou uma abordagem muito mais

simples baseada na diferença dos gaussianas (LI; TAN, 2009) semelhante a (LI *et al.*, 2004). Este é o primeiro trabalho em literatura a propor um banco de dados publicamente disponível especificamente adaptado para o desenvolvimento de contra-medidas a ataques de apresentação. Já os autores (KOLLREIDER; FRONTHALER; BIGUN, 2009) apresentam uma técnica para avaliar a vivacidade baseada em uma pequena sequência de imagens usando um detector binário que avalia as trajetórias de partes selecionadas da face apresentadas ao sensor de entrada usando uma análise simplificada do fluxo óptico seguida por classificador heurístico. Os mesmos autores introduziram em (KOLLREIDER; FRONTHALER; BIGUN, 2008) um método para fundir pontuações de diferentes sistemas que observam, simultaneamente, o movimento da face em 3-D. Este esquema avalia a vivacidade através das propriedades como piscar os olhos ou movimentos da boca. Enquanto os autores (BAO *et al.*, 2009) propõem um método para detectar ataques produzidos com mídia planar (como papel ou telas) usando estimativa de movimento por fluxo ótico.

Os autores (WANG *et al.*, 2017) dividem os métodos fundamentados em movimentos em duas categorias. Uma é o padrão movimento do rosto, isto é, as expressões faciais e os movimentos. A outra é que o movimento relativo entre a face e o plano de fundo. Sun et al. (SUN *et al.*, 2007) introduziu uma abordagem baseada em piscar dos olhos usando *Conditional Random Fields* (CRFs) para modelar atividades da piscada, que é uma ação representada pela sequência da imagem do olho, inclui imagens com olhos fechados e abertos. Além disso, os autores compararam *AdaBoost* e *hidden Markov Model* (hMM) com o modelo CRF proposto. Pan et al. (PAN *et al.*, 2011) propuseram uma estratégia que funde pistas a partir de piscada dos olhos e pistas usando o contexto da cena de uma forma não intrusiva para o reconhecimento da face. Utilizam um modelo de grafo para representar as dependências contextuais em sequências de imagens de piscadas dos olhos.

Kollreider et al. KOLLREIDER; FRONTHALER; BIGUN (2005) sugerem combinar a detecção de partes faciais com a estimativa de fluxo óptico para estimar a vivacidade, e explorar a trajetória de partes da face de uma sequência de imagens para diferenciar faces reais das falsas. Os pontos-chaves do método proposto são que uma face 3D gera um movimento especial 2D maior nas partes centrais da face (por exemplo, nariz) em comparação com regiões mais externas da face (por exemplo, orelhas). Ou seja, quando uma fotografia é movimentada em frente a uma câmera gera movimentos em várias regiões da face. Partes mais próximas de uma câmera tem movimento diferente comparado às regiões mais longes em um face viva. Para estimativa de fluxo óptico, foi proposto um novo fluxo chamado Fluxo Óptico de Linhas (OFL) (KOLLREIDER; FRONTHALER; BIGUN, 2009) que só era especializado em movimentos de linhas. Note que OFL pode distinguir o movimentos de pontos de movimentos de linhas. Para os primeiros, eles não só empregaram uma classificação de características Gabor baseadas em modelos, mas também apresentaram padrões correspondentes de fluxo óptico.

Kollreider et al. (KOLLREIDER *et al.*, 2007) propuseram uma nova abordagem para a vivacidade utilizando a classificação dos movimentos labiais e leitura labial sem informações de áudio, o que exige a cumplicidade dos usuários para proferir uma sequência aleatória de dígitos através de uma caixa de diálogo solicitado por texto. Analisando a dinâmica dos lábios, extraindo as OFL em tempo real. Os autores consideram as bocas como ROIs em cada pessoa foi gravada durante a leitura da sequência 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 para extrair o OFL e treinar um classificadores SVMs para avaliar a dinâmica labial.

Métodos fundamentados em análise de texturas são desenvolvidos para tratar ataques por fotos ou vídeos. Esses métodos são propostos para extrair artefatos das imagens de

Tabela 3: Melhores Resultados, Base CASIA

Método	HTER(%)
DoG *baseline* (ZHANG *et al.*, 2012)	27.55
LBP-TOP u2 (FREITAS PEREIRA *et al.*, 2013)	9.05
DoG+LBP+SVMs *baseline* (WEN; HAN; JAIN, 2015)	7.85
IDA+SVMs (WEN; HAN; JAIN, 2015)	6.70
LBP+SVMs *baseline* (WEN; HAN; JAIN, 2015)	3.40
hOG (YANG *et al.*, 2015)[1]	2.01
Dicionário (MANJANI *et al.*, 2017)	1.3

ataques. Em (YANG *et al.*, 2013), os autores argumentam que feições de texturas (LBP, DoG, hOG) são capazes de diferenciar faces *spoof* de faces genuínas. Métodos baseados em texturas tem alcançado um significante sucesso nas bases de dados IDIAP (CHINGOVSKA; ANJOS; MARCEL, 2012) e CASIA (ZHANG *et al.*, 2012). A taxa HTER (ver Seção 3, Equação (1)) na base de dados IDIAP foi reduzida de 13.87% em (CHINGOVSKA; ANJOS; MARCEL, 2012) e 7.60% em (FREITAS PEREIRA *et al.*, 2013) para 6.62% em (BHARADWAJ *et al.*, 2013) por incorporar pistas de texturas. Ao contrário dos métodos que utilizam análise de movimento, métodos baseados em texturas utilizam um único quadro para detectar a fraude.

Porém a habilidade de generalização pode ser um pouco pobre. Um estudos reportado em (FREITAS PEREIRA *et al.*, 2013) mostra que para dois métodos de texturas, a taxa HTER aumenta drasticamente em um cenário de cruzamento de banco de dados, isto é, o sistema é treinado em um banco de dados e testado em um outro. Devido à natureza intrínseca dos métodos baseados em texturas, eles podem facilmente ser sobreajustados para uma determinada condição de iluminação e não generalizar bem para outras condições de aquisição de imagens.

A Tabela 3 ilustra taxas HTER para os melhores métodos da literatura para a base de dados CASIA (ZHANG *et al.*, 2011).

2.2 Abordagem distintas

A Tabela 4 organiza diferentes propostas em relação ao ano da publicação, base de dados e técnica utilizada. É possível observar a relevância do tema nos últimos anos. Além disso, evidencia-se a frequência do uso das bases de dados NUAA (TAN *et al.*, 2010), CASIA (ZHANG *et al.*, 2012) e IDIAP (CHINGOVSKA; ANJOS; MARCEL, 2012) para análise de métodos de contramedidas para evitar a a falsificação de usuário. E por último, a técnica utilizada para abordar esse problema.

Além dos tópicos levantados na Tabela 4, há um ponto em que todas as abordagens são idênticas exceto o trabalho proposto em (YANG *et al.*, 2015): método especifico por pessoa. Nos outros trabalhos, um classificador genérico para todas as pessoas é construído para detectar ataques a sistemas biométricos faciais. Porém, devido a diferenças individuais entre os sujeitos, um classificador genérico pode não generalizar bem. Em (YANG *et al.*, 2015) foi proposto um classificador especificamente treinado para cada indivíduo que diminui as interferências entre sujeitos. Desta forma, nesta tese foi desenvolvido um classificador específico por pessoa, ver Seção 3.1 e dois classificadores genéricos, ver Seções 3.2 e 3.3, utilizando técnicas fundamentadas em movimento, análise de texturas e deformações.

Tabela 4: Métodos e Técnicas

Referência	Ano	Base de Dados	Técnica
(HE; LU; SHI, 2008)	2008	IRIS	Qualidade da Imagem
(TAN *et al.*, 2010)	2010	NUAA	Qualidade da Imagem
(ZHANG *et al.*, 2011)	2010	Própria	Análise Multiespectral
(AKHTAR *et al.*, 2011)	2011	Própria	Multimodal
(ANJOS; MARCEL, 2011)	2011	Print-Attack	Movimento
(MAATTA; HADID; PIETIKAINEN, 2011)	2011	NUAA	Textura
(ZHANG *et al.*, 2012)	2012	CASIA	Textura
(KOSE; DUGELAY, 2012)	2012	NUAA	Textura
(MARSICO *et al.*, 2012)	2012	NUAA e hONDA	Movimento
(BHARADWAJ *et al.*, 2013)	2013	IDIAP e Print-Attack	Movimento e Textura
(FREITAS PEREIRA *et al.*, 2013)	2013	IDIAP e CASIA	Textura
(ALBU, 2015)	2015	NUAA	Textura
(BENLAMOUDI *et al.*, 2015)	2015	NUAA e CASIA	Textura
(WEN; HAN; JAIN, 2015)	2015	MSU, CASIA e IDIAP	Qualidade da Imagem
(YANG *et al.*, 2015)	2015	CASIA e IDIAP	Qualidade da Imagem
(BOULKENAFET; KOMULAINEN; HADID, 2016)	2016	CASIA, IDIAP e MSU	Textura

2.3 Base de dados NUAA

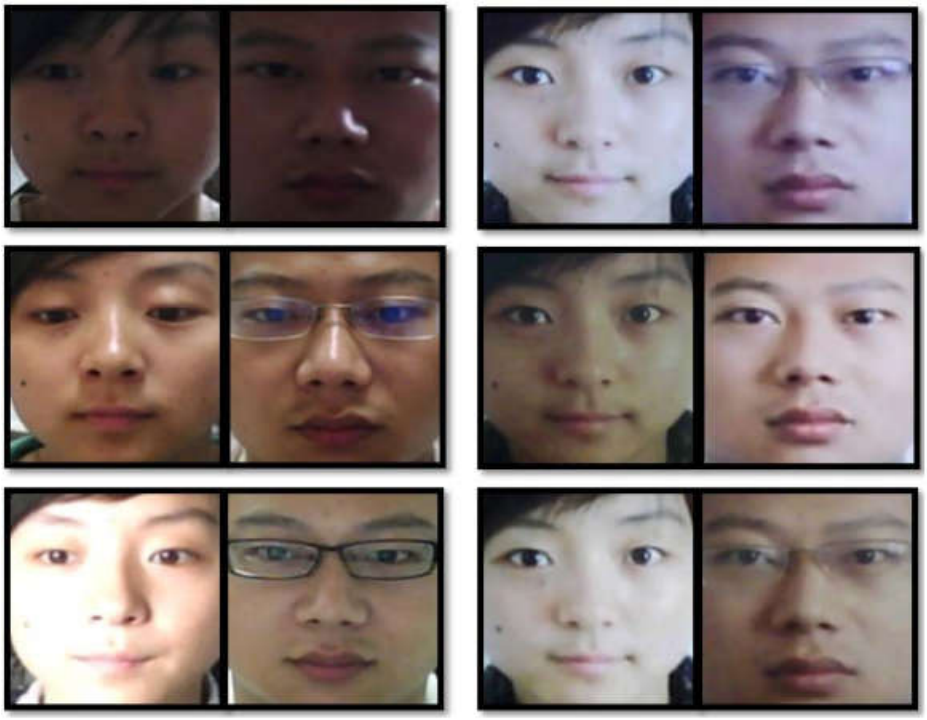
A base de dados NUAA disponibilizada pelos autores TAN *et al.* (2010) é uma grande coleção de fotografia-impostor publicamente disponível. Foi construída utilizando uma *webcam* comum. Esse banco de dados foi coletado em três sessões com cerca de 2 semanas de intervalo entre duas sessões, e o local e as condições de iluminação de cada sessão são diferentes também. Com um total de 15 sujeitos (numerados de 1 a 15) 4 foram convidados a participar deste trabalho. Em cada sessão, foram capturadas as imagens de ambos os sujeitos vivos e suas fotografias. Algumas amostra das imagens das três sessões estão na Figura 3 Os autores TAN *et al.* (2010) acreditam que o objetivo principal da PAD é distinguir um rosto real de uma fotografia, em vez de diferenciar pessoas diferentes como o caso do reconhecimento facial usual, a exigência de número de sujeitos é menos exigente em comparação com a riqueza de variações contidas no conjuntos de dados.

Em particular, para cada sujeito em cada sessão, foi utilizado a webcam para capturar uma série de imagens faciais (com taxa de 20 quadros por segundos e 500 imagens para cada sujeito). Durante a captura de imagens, cada sujeito foi solicitado a olhar para a *webcam* frontalmente, com expressão neutra e sem movimentos aparentes, tal como piscada dos olhos ou movimento da cabeça. Em outras palavras, foi sugerido que um ser humano vivo pareça uma foto o máximo possível (vice-versa para fotografia). Alguns exemplos das imagens capturadas estão ilustrados na Figura 3 (coluna da esquerda).

Os autores TAN *et al.* (2010) coletaram amostras de fotografias da seguinte forma, primeiro foi tirado uma foto de alta definição para cada sujeito usando uma câmera Canon comum de uma maneira que a área da face ocupasse pelo menos $2/3$ da área inteira da fotografia. Em seguida, imprimiram as fotos de duas maneiras. A primeira é usar o método tradicional para imprimí-las em papel fotográfico comum com tamanho de 6.8cm $\times$ 10.2cm (pequeno) e 8.9cm $\times$ 12.7cm (maior), respectivamente. A outra maneira, foi impresso cada foto em um papel A4 de 70g usando uma impressora hP colorida comum. Baseado nisso, três categorias de ataque por foto são simuladas frente à *webcam*, de uma forma semelhante a PAN *et al.* (2007), como mostrado na Figura 4.

Nessa base foi construído um conjunto de treinamento e um conjunto de teste a partir das fotos genuínas e ataques por fotos, ambos contendo várias imagens de clientes e im-

Figura 3: Ilustração das amostras do banco de dados NUAA. Em cada coluna (de cima para baixo) as amostras são respectivamente da sessão 1, sessão 2 e sessão 3. Em cada linha, o par da esquerda é de um humano vivo e o direito de uma foto. Note que ele contém várias mudanças de aparência comumente encontradas por um sistema de reconhecimento facial (por exemplo, sexo, iluminação, com ou sem óculos). Todas as imagens originais da base de dados são coloridas com a mesma definição de 640×480 pixels.



Fonte: (TAN *et al.*, 2010)

postores. O conjunto de treinamento é construído usando as imagens das duas primeiras sessões e o conjunto de testes da terceira sessão. Em particular, o conjunto de treinamento contém 889 imagens a partir da primeira sessão e 854 imagens da segunda sessão e todos os sujeitos disponíveis nas duas sessões estiveram envolvidos. Consequentemente foram adquiridas as 1743 imagens dos 9 sujeitos como traço biométrico válido e para as imagens impostoras do conjunto de treinamento foram selecionadas 855 e 893 da primeira e segunda sessão respectivamente conjunto, portanto temos 1748 imagens impostoras ao total. O conjunto de teste contém 3362 imagens de seres humanos vivos selecionados da sessão 3 e 5761 imagens de fotos (ataques) selecionadas da sessão 3 também. A Tabela 5 fornece algumas estatísticas disso. Note que não há sobreposição entre o conjunto de treinamento e o conjunto de teste. Além disso, alguns assuntos no conjunto de teste não aparecem no conjunto de treinamento, o que aumenta a dificuldade do problema.

Resumindo, esse banco de dados contêm mais de 12 mil imagens fotográficas de 15 indivíduos. Conforme o diagrama exposto na Figura 5, a base contém 12614 quadros divididos em 4 grupos, treino cliente, treino impostor, teste cliente, teste impostor. Consta com um conjunto de treino de 3491 quadros que compreende um subconjunto de imagens genuínas com 1744 quadros (15 vídeos) e outro subconjunto de 1747 imagens de ataques (18 vídeos). Enquanto o conjunto de teste corresponde a 9123 quadros que compreende

Figura 4: Ilustração de diferentes ataques por foto (da esquerda para a direita): (1) mova a foto horizontalmente, verticalmente, atrás e frente; (2) girar a foto em profundidade ao longo do eixo vertical; (3) o mesmo que (2) mas ao longo do eixo horizontal; (4) dobrar a foto para dentro e para fora ao longo do eixo vertical; (5) o mesmo que (4), mas ao longo do eixo horizontal.

Fonte: (TAN *et al.*, 2010)

Tabela 5: Os números de imagens nos conjuntos treino/teste da bases NUAA (TAN *et al.*, 2010)

	Seção 1	Seção 2	Seção 3	Total
Conjunto Treinamento				
Cliente	889	854	0	1743
Impostor	855	893	0	1748
Total	1744	1747	0	3491
Conjunto Teste				
Cliente	0	0	3362	3362
Impostor	0	0	5761	5761
Total	0	0	9123	9123

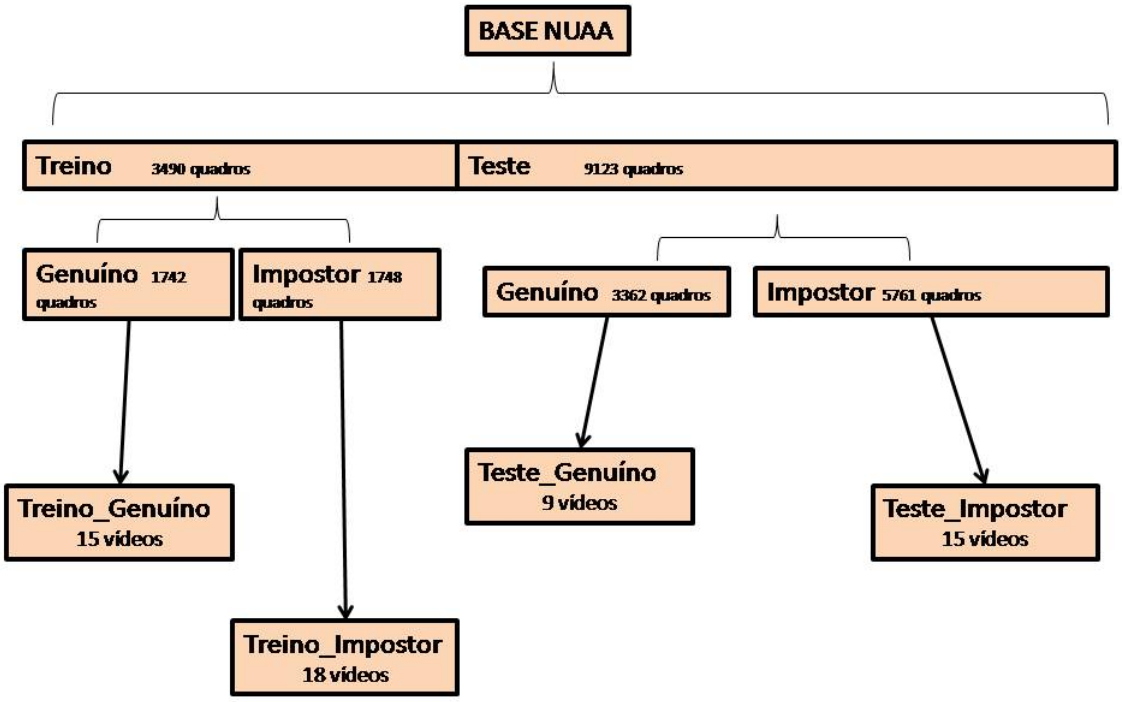
um subconjunto de imagens genuínas com 3362 quadros (9 vídeos) e outro subconjunto de 5761 imagens de ataques (15 vídeos). A Figura 6 ilustra as sequências de quadros do conjunto teste de imagens genuínas. Enquanto a Figura 7 exemplifica as sequências de quadros do conjunto teste de imagens impostoras.

Ao analisar a Tabela 5 e a Figura 5 é possível verificar que essa base está desbalanceada em relação as suas classes. Isto acontece no conjunto de teste onde 36.85% dos quadros são faces genuínas e os 63.15% dos quadros restantes são faces pertencentes a classe ataque. Enquanto que o conjunto de treino possui 49.93% dos quadros pertencentes a classe genuína e 50.07% dos quadros pertencentes a classe ataque.

2.4 Base de dados CASIA

O autores ZHANG *et al.* (2012) relatam que o problema da qualidade de imagem não havia sido discutido anteriormente em bases de dados para PAD. De um modo geral, o desempenho de um algoritmo depende da qualidade da imagem até certo ponto. Além disso, o requisito da qualidade também determina os dispositivos para coleta dos dados.

Figura 5: Composição da base de dados NUAA.

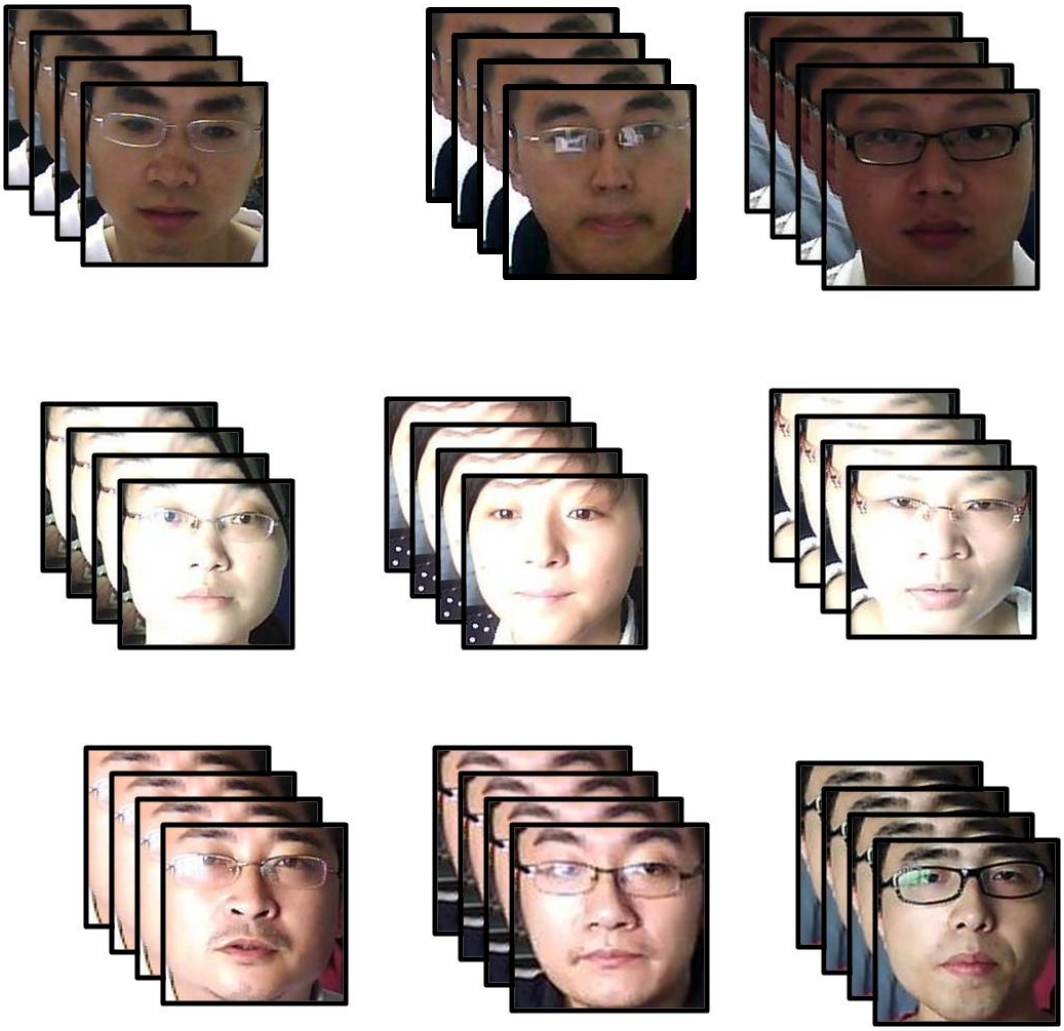


Por uma questão de simplicidade, os autores ZHANG *et al.* (2012) definem empiricamente a qualidade da imagem como a preservação das texturas faciais, pelas quais se nota mais atenção ao sentimento perceptivo do que medidas estritamente quantitativas.

Para construir a base de dados CASIA os autores ZHANG *et al.* (2012) utilizaram três câmeras diferentes para registrar os dados com distintas qualidades. O vídeo de baixa qualidade é capturado por uma câmera USB de longa data, já que o uso de longo tempo sempre degrada a qualidade de imagem. A altura e largura da imagem para vídeos de baixa qualidade é 640 e 480 respectivamente. A qualidade normal de vídeo é capturada por uma câmera USB nova que mantem a qualidade de imagem original. A resolução da qualidade normal é 480 $\times$ 640. Para os vídeos de alta qualidade, é utilizada uma câmera Sony $NEX-5$ de alta resolução para gravação, cuja resolução máxima é de 1920 $\times$ 1080. Foi tirado uma imagem com a resolução máxima de 1920 $\times$ 1080 para cada sujeito para confeccionar a foto de ataque. Para vídeos de faces genuínas e de ataques, no entanto, um vídeo de alta resolução é um fardo pesado demais para salvar e calcular. Portanto, os autores ZHANG *et al.* (2012) cortaram uma região dos quadros de 1280 $\times$ 720 que contêm as faces para a criação dos vídeos finais. Dessa forma, a carga e os cálculos foram reduzidos enquanto que a qualidade pode ser maximizada. Alguns exemplos podem ser vistos na Figura 8, em que apenas as partes da face são mostradas. Um vídeo completo pode ser visto na Figura 9.

Em relação às faces genuínas, 50 sujeitos foram utilizados para formar as faces genuínas. Todos os sujeitos são capturados em cenas naturais sem unificação do ambiente artificial. Durante a gravação, os participantes devem exibir um comportamento intermitente em vez de ficarem parados. Os autores ZHANG *et al.* (2012) argumentaram que o

Figura 6: Faces genuínas do conjunto teste.



Fonte: (TAN *et al.*, 2010)

movimento facial é uma pista crucial para a vivacidade na detecção de ataques de apresentação, e é necessário fornecê-los como uma estratégia de desafio-resposta usada em métodos de detecção de movimento facial. O tipo de movimento de piscar é escolhido porque é mais natural e fácil de usar do que outros tipos de movimento como movimento da cabeça e movimento da boca. Um processo de piscada pode ser visto na Figura 10.

Já as faces falsas, são os principais ingredientes do banco de dados CASIA (ZHANG *et al.*, 2012). Especificamente, foram projetados os três tipos de ataques de falsificação de faces. Ataque de foto deformada. Como mencionado, é utilizado uma câmera Sony NEX-5 para gravar uma imagem 1920×1080 e um vídeo de 1280×720 para cada sujeito. Essa imagem de alta resolução é utilizada para imprimir as fotos e essas fotos são impressas no papel de cobre, que tem qualidade muito maior do que papel normal de impressão A4. Em um ataque de foto deformado, o atacante distorce deliberadamente uma foto intacta, tentando simular o movimento facial. Intacto significa que não há regiões de corte na foto, em contraste com o ataque foto com corte, explicado a seguir. Os ataques por fotos distorcidas são muito semelhantes aos da Figura 4.

Ainda sobre as faces impostoras, há o ataque com fotos cortadas nos olhos. Nessa base de dados, foi exigido que os sujeitos participantes exibissem um comportamento in-

Figura 7: Faces impostores do conjunto teste.

Fonte: (TAN *et al.*, 2010)

termitente, aqui as regiões dos olhos são cortadas. A atacante se esconde atrás e exibe piscadas de olhos pelos buracos como mostrado na Figurafig:casia(c). Outra implementação possível é proposta em (KOLLREIDER; FRONTHALER; BIGUN, 2008) que uma foto intacta é firmemente colocada atrás de outra foto com cortes na região dos olhos, e movendo a foto de trás, o movimento de piscando pode ser simulado. Na base de dados Casia ambas as duas implementações existem.

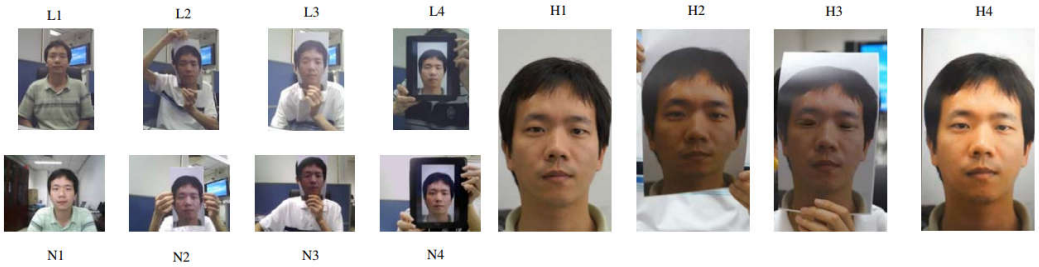
Já o último tipo de ataque é através da apresentação de vídeos. Neste caso, os vídeos de alta resolução das faces genuínas são exibidos usando um iPad. Observe que limitado pela resolução da tela do iPad, a resolução original vídeos (1280×720) serão inevitavelmente reduzidos pelo dispositivo. Um exemplo pode ser visto na Figurafig:casia(d).

Resumindo, a base de dados CASIA *Face Anti-Spoofing Database* (FASD) (ZHANG *et al.*, 2012), lançada em 2012, consiste em 600 gravações de vídeo de genuínos e tentativas de ataque de 50 sujeitos diferentes. Embora o tamanho da base de dados CASIA seja um pouco menor do que a base de dados IDIAP (CHINGOVSKA; ANJOS; MARCEL, 2012), contém amostras mais diversas em termos de qualidade de imagem. São disponíveis 3 qualidades de imagens e nesta tese os experimentos são realizados no conjunto de dados de alta resolução. Além da qualidade de imagem, há variações de rosto (pose e expressões variadas) e tentativas de ataques: deformação de foto, foto cortada nos olhos e vídeo em alta resolução exibido em um dispositivo móvel. Dessa forma, o conjunto de dados foi dividido por indivíduo, para cada um há 4 vídeos sendo 1 genuíno e 3 ataques (deformação, corte e video). A Figura 11 exemplifica essa divisão do conjunto de dados. A base CASIA é dividida (conforme os autores ZHANG *et al.* (2012)) da seguinte forma, para cada sujeito, os 4 vídeos tiveram os quadros extraídos e alocados aleatoriamente dentro dos conjuntos de treinamento e teste (70% e 30%). Por exemplo, para o sujeito #2, o video genuíno possui 351 quadros, assim o conjunto de treino recebe aleatoriamente 246 quadros e o conjunto de teste recebe 105 quadros.

Figura 8: Vídeos de baixa, normal e alta qualidade. Apenas as regiões da face são mostradas.

Fonte: (ZHANG *et al.*, 2012)

Figura 9: Um conjunto de vídeos completo para um sujeito. As quatro imagens superiores da esquerda representam os vídeos de baixa qualidade (L1, L2, L3 e L4), em baixo à esquerda é a qualidade normal dos vídeos (N1, N2, N3 e N4), e à direita são os vídeos de alta qualidade (h1, h2, h3 e h4). Para cada qualidade, da esquerda para a direita são genuínos, ataque por foto, ataque por foto cortada e ataque por vídeo.

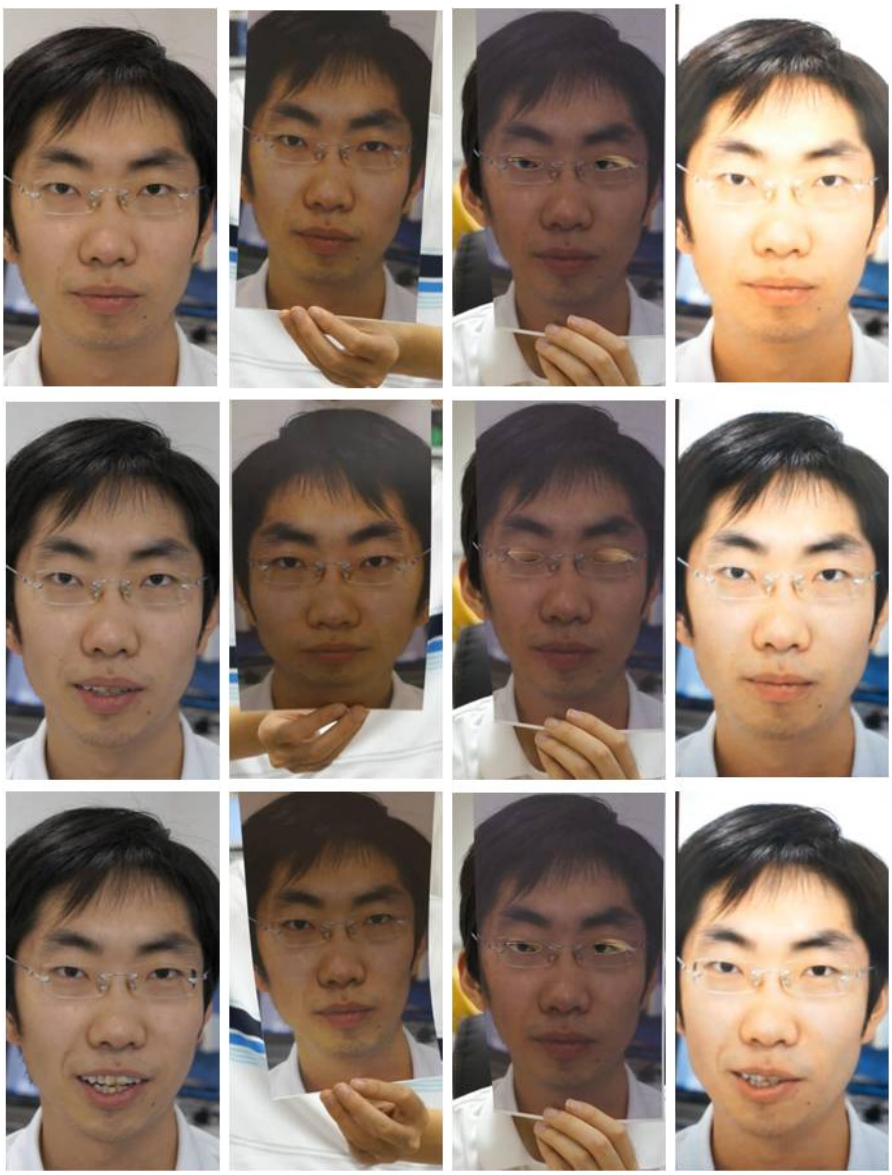


Fonte: (ZHANG *et al.*, 2012)

Figura 10: Processo de piscada dos olhos, (ZHANG *et al.*, 2012).

Fonte: (ZHANG *et al.*, 2012)

Figura 11: Ilustração da base de dados CASIA. Sujeito #1, (a) quadros genuínas, (b) ataque por deformações da foto, (c) ataque por fotos com buracos nos olhos e (d) ataque por reprodução de vídeo.



(a) genuíno (b) deformação (c) corte (d) vídeo

Fonte: (ZHANG *et al.*, 2012)

3 CONTRAMEDIDAS PARA EVITAR A FALSIFICAÇÃO DO USUÁRIO NO ACESSO A SISTEMAS BIOMÉTRICOS FACIAIS

3.1 Método 1

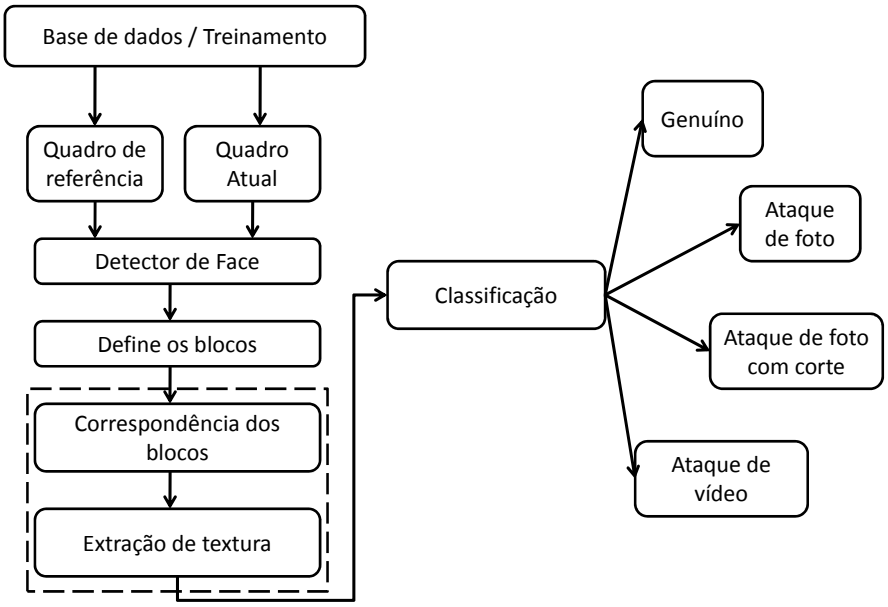
Embora tenha havido muito avanço nas últimas décadas para superar várias limitações do processo de reconhecimento facial, incluindo pesquisas na segmentação facial, estudos recentes revelam que sistemas biométricos faciais são vulneráveis a ataques de falsificação, ou seja, ataques de apresentação (HADID *et al.*, 2015), que consistem em gerar um traço biométrico falso para representar um cliente registrado. Um exemplo foi apresentado na Figura 2 (Seção 1.1), onde um fraudador tenta enganar um sistema de reconhecimento facial ao apresentar uma fotografia, um vídeo ou uma máscara 3D que se assemelha a um indivíduo legítimo

A proposta desta abordagem visa desenvolver um método de contra-ataques de apresentação específico por pessoa capaz de integrar as diversas qualidades ressaltadas na Tabela 2. Isso pode ser atingindo agrupando diferentes técnicas à metodologia proposta. Dado um vídeo da base de dados, define-se um quadro de referência $I_{t=1}$, escolhe-se uma face neutra e os quadros seguintes são definidos por J_t. A extração do rosto é realizada via *deep learning* (ZHANG *et al.*, 2016) (Ver a Seção 3.1.6.2). Divide-se a face em k blocos. Cada bloco k do quadro J_t é definido como J_t^k sendo k-ésimo bloco pertencente ao quadro J no instante t. Cada bloco tem o seu movimento estimado em relação aos blocos do quadro de referência $I_{t=1}^k$ utilizando o método proposto pelo autor em (SCHARDOSIM; SCHARCANSKI, 2017) e aplicando transformações não afins do tipo *Thin-Plate Spline* (TPS). A correspondência entre blocos sempre ocorre entre o quadro de referência $I_{t=1}$ e o quadro atual J_t com $t > 1$. Para os blocos correspondentes de J_t^k são extraídas as texturas utilizando um extrator aqui desenvolvido chamado de Log-Radial Diff (LRD). As texturas extraídas são projetadas em uma dimensão reduzida utilizando Random Projections (RP) (ACHLIOPTAS, 2003) e classificadas utilizando *Support Vector Machine* (SVMs), Máquina de vetores de suporte (CRISTIANINI; SHAWE-TAYLOR, 2000). Utilizando essa abordagem são construídos os modelos de classe para as faces genuínas e classes de ataques: ataque de foto, ataque de foto com corte nos olhos e ataque de vídeo.

Essa primeira abordagem desenvolvida está ilustrada na Figura 12. No qual a sequência de blocos ilustram as principais etapas do método. Nas caixas Correspondência de blocos e Extração de Textura encontram-se as principais contribuições desta tese.

A Tabela 6 relaciona as etiquetas conhecidas (conjunto verdade) para cada bloco com as respostas do classificador gerando 4 condições: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN) e *False Negative* (FN). No qual TP significa uma face genuína classi-

Figura 12: Diagrama da metodologia proposta.



ficada como genuína, TN significa uma face *spoof* classificada como *spoof*, FP significa uma face *spoof* classificada erroneamente como genuína e FN significa uma face genuína classificada erroneamente como *spoof*. As duas últimas condições são importantes, pois podem ser utilizadas para calcular as taxas de aceitação falsa e a taxa de rejeição falsa discutidas a seguir.

Para manter a consistência aos trabalhos anteriores (ANJOS; MARCEL, 2011; MARSICO *et al.*, 2012; FREITAS PEREIRA *et al.*, 2013; WEN; HAN; JAIN, 2015; BOULKENAFET; KOMULAINEN; HADID, 2016; MANJANI *et al.*, 2017) utiliza-se a *half Total Error Rate* (HTER) como métrica definida na Equação (1):

$$HTER(\tau, D_{treino}) = \frac{FRR + FAR}{2} \tag{1}$$

onde *False Rejection Rate* (FRR) é a taxa de rejeição falsa, ou seja, é o número de faces genuínas classificadas erroneamente como *spoof* dividido pelo número total de faces genuínas. Enquanto *False Acceptance Rate* (FRR) é a taxa de aceitação falsa, ou seja, o número de faces *spoof* classificadas erroneamente como faces genuínas dividido pelo número total de *spoof*. Especificamente, primeiro é encontrado o ponto de operação τ onde a $HTER$ é mínimo para um determinado conjunto de treino D_{treino}. Então, para o limiar τ correspondente ao ponto de operação é utilizado para calcular $HTER$ no conjunto de teste.

Nesta abordagem foi construídos um método para evitar a falsificação de dados biométricos específico por pessoa. Desta forma, o limiar τ está associado com cada indivíduo e é calculado para cada pessoa. Isso é diferente de métodos genéricos nos quais se calcula apenas um valor de $HTER$ durante o treinamento.

3.1.1 Detecção e Decomposição da Face

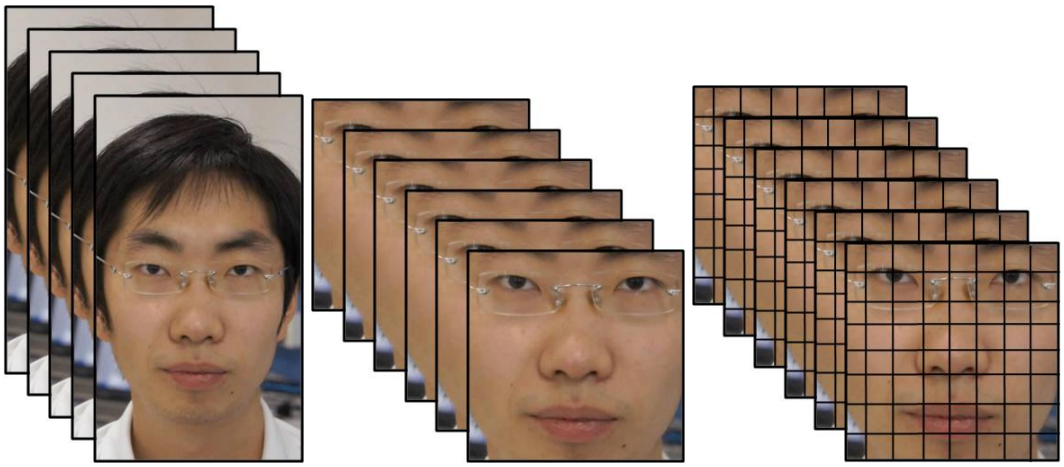
A detecção de face utilizada nesta proposta usa os algoritmos desenvolvidos em (ZHANG *et al.*, 2016). Esses autores mostram que a tarefa detecção de pontos ou alinhamento da face não é única e independente. Ao invés disso, a robustez é grandemente melhorada com informações auxiliares. Especificamente, eles otimizam a detecção de pontos juntamente com o reconhecimento heterogêneo de atributos faciais, tais quais gênero, expressão e atributos de aparência. Para resolver este problemas (ZHANG *et al.*, 2016) formularam um novo modelo de aprendizado profundo de tarefas (Ver a Seção 3.1.6.2). A detecção da face foi realizado com o algoritmo disponível pelos autores ZHANG *et al.* (2016) no endereço eletrônico (ZHANG *et al.*, 2014a). A Figura 13 ilustra o uso do algoritmo proposto em (ZHANG *et al.*, 2016) para detecção de face para três sujeitos nos conjuntos cliente e impostor da base de dados de fotografia de impostores (TAN *et al.*, 2010).

Figura 13: Exemplos do uso do algoritmo de detecção de face proposto por (ZHANG *et al.*, 2016);A primeira coluna exibe imagens de clientes e suas respectivas faces detectadas na segunda coluna. A terceira coluna exibe imagens de impostores e suas respectivas faces na quarta coluna.

Fonte: (TAN *et al.*, 2010)

Uma vez encontrada a face em um determinado quadro ela é dividida em blocos e cada um dos blocos é analisado individualmente. Esta abordagem permite encontrar a melhor correspondência entre cada bloco $I^k_{t=1}$ do quadro de referência e cada bloco J^k_t do quadro atual. A Figura 14 ilustra a divisão da face em blocos.

Figura 14: Ilustração da divisão das faces em blocos. (a) sequência dos quadros; (b) faces detectadas utilizando o programa disponível em (ZHANG *et al.*, 2014a) e (c) faces divididas em blocos.



(a) (b) (c)

Fonte: (TAN *et al.*, 2010)

3.1.2 Estimação dos Parâmetros de Movimento

Dado o quadro de referência $I_{t=1}$ e o quadro atual J_t, as respectivas faces são encontradas e divididas em blocos definindo as matrizes $I^k_{t=1}$ e J^k_t, onde k define o índice do bloco. A metodologia proposta visa encontrar os blocos correspondentes para permitir extrair características suficientes para discriminar as classes de faces genuínas e *spoof*. Para encontrar os blocos correspondentes é necessário estimar os parâmetros de movimento. Inicialmente, uma transformação afim entre os blocos é assumida e o movimento pode ser decomposto em rotação, escala e translação utilizando uma matriz de movimento $[M] = [T][R][S]$ onde as matrizes T, R e S representam a translação, rotação e escala, respectivamente (GONZALEZ; WOODS; EDDINS, 2007). Foi utilizado parcialmente o método proposto (SCHARDOSIM; SCHARCANSKI, 2017) para obter os parâmetros de translação invariante à rotação e escala entre os blocos. Posteriormente, assume-se que ainda haja alguma deformação entre os blocos, ou seja, ainda há alguma rotação, escala ou deformação não afim entre os blocos. Desta forma, sucessivas transformações do tipo TPS são aplicadas para encontrar o melhor casamento entre os blocos.

Considerando o sistema de coordenada log-polar (ρ, a), onde ρ denota o logaritmo da distância radial r de um pixel na posição (x, y) ao centroide do bloco (x_c, y_c) e a denota o ângulo radial. As coordenadas do pixel (x, y) podem ser representadas em termos das coordenadas log-polar por:

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2}, \tag{2}$$

$$a = tan^{-1}(\frac{y - y_c}{x - x_c}), \tag{3}$$

$$\rho = \log_b r, \tag{4}$$

onde, $b = \exp{(\frac{\log S_\rho}{S_\rho})}$, $x = r \cdot \cos(a)$, $y = r \cdot \sin(a)$. Uma vez que (S_x, S_y) é o tamanho do bloco no espaço cartesiano (espaço da imagem) e assumindo que a origem do sistema

de coordenada está em (x_c, y_c), o tamanho do bloco no espaço log-polar é $(S_\rho = S_x, S_a = S_y)$.

Quando a transformação log-polar é aplicada a um bloco $I_{t=1}^k$, linhas radiais do espaço cartesiano (espaço da imagem) são mapeadas para linhas horizontais no espaço log-polar (ou seja, o espaço log-polar do bloco transformado $I_{t=1}^k(\rho, a)$, onde ρ e a são os eixos log-polar horizontal e vertical, respectivamente. Escalas e rotações no espaço (x, y) transformam-se em deslocamento ao longo dos eixos (ρ, a) do espaço log-polar. Por exemplo, a versão rotacionada da imagem $I(x, y)$ mostrada na Figura 15(a) está ilustrada na Figura 15(c), onde as linhas em (ρ, a) são representadas por linhas radiais em (x, y), e rotacionando $I(x, y)$ em Figura 15(a) corresponde a translacionar $I(\rho, a)$ no espaço (ρ, a), como mostrado nas Figuras 15(b) e (d).

Figura 15: Ilustração da transformação log-polar: (a) imagem $I(x, y)$ no espaço cartesiano; (b)$J(x, y)$, versão rotacionada de $I(x, y)$; (c) $I(\rho, a)$, transformação log-polar de $I(x, y)$; e (d) $J(\rho, a)$, transformação log-polar de $J(x, y)$. $J(x, y)$ está rotacionada em $45°$ em relação a $I(x, y)$, enquanto $J(\rho, a)$ está translacionada ao longo do eixo a em 32 pixels.

(a) $I(x, y)$

(b) $J(x, y)$

(c) $I(\rho, a)$

(d) $J(\rho, a)$

Similarmente a rotações, mudanças na escala entre dois blocos podem ser identificadas mais facilmente no espaço log-polar (ρ, a) do que no espaço cartesiano (x, y). Por exemplo: se $J_t^k(x, y)$ é uma ampliação de 2 vezes de $I_{t=1}^k(x, y)$, cada pixel (x, y) em $I_{t=1}^k(x, y)$ é mapeado para $(2x, 2y)$ em $J_t^k(x, y)$. Desta forma tem-se $(x, y) \rightarrow (2x, 2x)$ e utilizando a transformação log-polar tem-se $(2x, 2y) \rightarrow (\log 2x, \log 2y) \rightarrow (\log x + \log 2, \log y + \log 2)$ e isto demonstra que um fator de escala no espaço cartesiano comporta-se da mesma forma que uma translação ao longo do eixo ρ no espaço log-polar.

Sendo assim, rotações e escalas que ocorrem no plano cartesiano comportam-se como translações no espaço log-polar. Essa propriedade permite encontrar a translação invariante à rotação e à escala entre dois blocos e é realizada através do algoritmo de casamento entre blocos no espaço log-polar, utilizando a métrica do erro quadrático médio: *Mean-Squared Error* (MSE), ver a Equação (5):

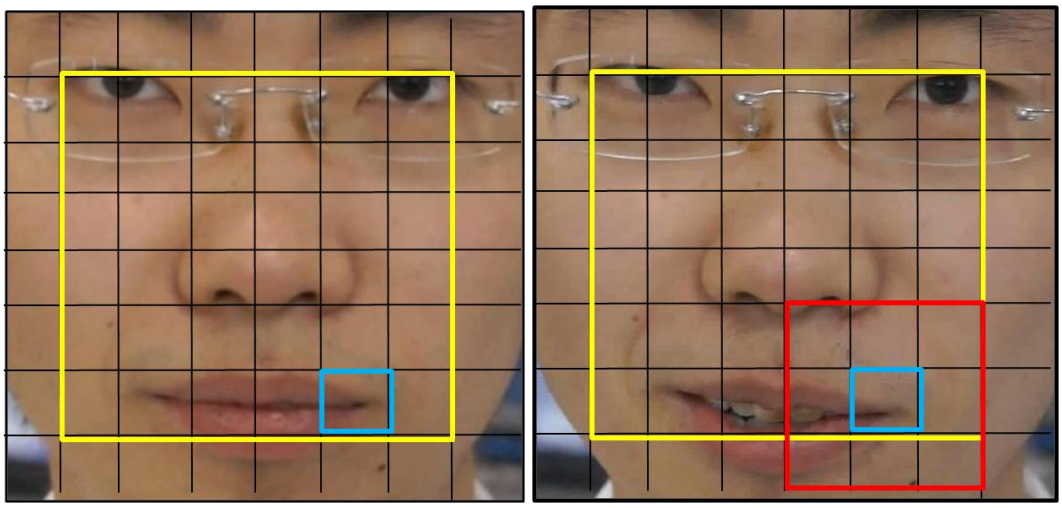
$$MSE_t^k = \frac{1}{C} \sum_{\rho=1}^{\hat{\rho}} \sum_{a=1}^{\hat{a}} \left[I_{t=1}^k(\rho, a) - J_t^k(\rho, a) \right]^2 \tag{5}$$

no qual C representa o número total de pixels no bloco $I_{t=1}^k$, $\hat{\rho}$ e $\hat{a}$ são o raio e ângulo máximos respectivamente. A metodologia utilizada foi a proposta em (SCHARDOSIM; SCHARCANSKI, 2017) e está descrita a seguir.

Inicialmente é definido uma face de referência $I_{t=1}$, sendo que durante o treinamento escolhe-se uma face neutra. Novas faces são definidas por J_t. Cada face é dividida e k blocos. De forma geral, o algoritmo de busca esparsa otimizada computa o MSE entre cada bloco do quadro de referência $I_{t=1}^k$ e a vizinhança ao redor do bloco J_t^k no espaço log-polar. Com isso é possível determinar a translação invariante à rotação e a escala entre os blocos $I_{t=1}^k(x, y)$ e $J_t^k(x, y)$ no espaço cartesiano. Detalhadamente, para cada bloco $I_{t=1}^k(x, y)$ é transformado para o espaço log-polar $I_{t=1}^k(\rho, a)$. Após, uma janela deslizante do mesmo tamanho do bloco é posicionada nas coordenadas (x', y') no quadro $J_t(x, y)$ e também é transformada para o espaço log-polar $J_t^{k,x',y'}(\rho, a)$. As posições (x', y') são esparsas (ou seja, espaçadas por $x' + dx', y' + dy', dx', dy' \in \mathbb{Z}$) dentro de uma área de busca de J_t e para cada posição esparsa (x', y') o MSE é calculado utilizando a Equação (5). Nos experimentos as faces foram divididas em $n^2 = 64$ blocos, o índice de bloco varia de $k = 1, 2, 3, ..., (n-2)^2$. Com esses parâmetros cada um dos 36 blocos analisados possui uma vizinhança de 8 blocos que delimita a janela de busca. A Figura 16(a) ilustra a face de referência com o bloco $I_{t=1}^{k=35}(x, y)$ em evidencia com a borda em azul e a janela em amarelo representa os blocos que são analisados. Já Figura 16(b) mostra o bloco $J_t^{k=35}(x, y)$ em azul e a janela em vermelho representa a área de busca do casamento entre os blocos com o índice $k = 35$.

Para cada índice de bloco k, é construído a matriz p_{map} que representa os valores do MSE computados entre os blocos $I_{t=1}^k(\rho, a)$ e $J_t^{k,x',y'}(\rho, a)$. Após completa a busca esparsa, uma segunda busca é realizada (busca exaustiva) na proximidade da posição onde ocorreu o mínimo da matriz p_{map}. Deste modo, a localização da janela de busca que leva ao menor MSE com precisão em p_{map} na posição (x'', y''), e essas coordenadas fornecem os parâmetros de translação $(\Delta x = x'' - x_c, \Delta y = y'' - y_c)$ entre os blocos $I_{t=1}^k(x, y)$ e $J_t^k(x, y)$. Finalmente, com os parâmetros de translação obtidos, cada bloco $J_t^k(x, y)$ pode ser corrigido por $(-\Delta x, -\Delta y)$. Um exemplo de estimação da translação entre duas faces pode ser visualizado na Figura 17 onde a face (a) foi translacionada em $(\Delta x = 0, \Delta y = 20)$, rotacionada em $\theta = 10°$ e escalada $S = 1.15$ gerando a Figura 17(b). Utilizando a abordagem descrita, é possível encontrar os parâmetros de translação para

Figura 16: Ilustração da janela de busca para um bloco da face, (a) janela em amarelo define todos os blocos que são analisados e em azul está o bloco $I_{t=1}^{k=35}(x, y)$; (b) em azul está o bloco $J_t^{k=35}(x, y)$ e a janela vermelha define a vizinhança 8 ao bloco.



(a) (b)

cada bloco da imagem representado pelas setas na Figura 17(a).

Figura 17: Exemplo: $(\Delta x = 0, \Delta y = 20)$, $\theta = 10°$, $S = 1.15$

(a) $I(x, y)$ (b) $J(x, y)$

3.1.3 Encontrando os Parâmetros de Transformação não afim

Na seção anterior, todos os blocos J_k que compõem a face tiveram suas translações corrigidas. Porém, a face é um objeto não-rígido e além das transformações afins (translação, rotação e escala) pode ainda haver deformações não afins. Para tratar esse problema são utilizadas sucessivas transformações do tipo TPS a fim de realizar o melhor casamento entre os blocos.

A utilização de Thin-Plates como splines de interpolação teve início com (DUCHON,

1976). A concepção de TPS apresenta características que a torna aplicável em grande parte dos problemas de modelagem (MAGNA JÚNIOR, 2012), pois possibilita a decomposição do problema em uma transformação linear e uma não linear. Além disso, sua formulação matemática assegura algumas restrições importantes à superfície interpolante, a qual é suave e se estende ao infinito em todas as direções. Na sequência são apresentados alguns conceitos e a formulação matemática de TPS.

A TPS fundamenta-se na minimização da energia de curvatura de uma placa delgada de metal fixa a alguns pontos de controle. Sua formulação garante restrições de que a superfície interpolante apresente mínima energia de deformação e que seja suave. De acordo com (BOOKSTEIN, 1989), para uma placa delgada sujeita a uma curvatura suave, a energia de dobra em um ponto com coordenadas (x, y) é expressa pela Equação (6):

$$I_f = \iint_{R^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy \tag{6}$$

e a função $f(x, y)$ que minimiza a energia de curvatura é dada pela Equação (7), (BOOKSTEIN, 1989):

$$f(x, y) = a + bx + cy + \sum W_i U(r) \tag{7}$$

no qual W_i representa os coeficientes não afins e $U(r) = r^2 \log r$ é uma função de base radial que depende da distância r que representa a distância euclidiana do ponto (x, y) ao i-ésimo ponto de controle. A função f, dada pela Equação 7 é dividida em duas partes: a soma das funções $U(r)$ que podem ser mostradas que são limitadas e assintoticamente plana e uma parte afim representando o comportamento de f no infinito.

Para o caso bidimensional, é possível reescrever a Equação (6) gerando o conjunto de Equação (8):

$$\begin{aligned} I_x &= \iint_{R^2} \left(\left(\frac{\partial^2 x}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 x}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 x}{\partial y^2} \right)^2 \right) dxdy \\ I_y &= \iint_{R^2} \left(\left(\frac{\partial^2 y}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 y}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 y}{\partial y^2} \right)^2 \right) dxdy \end{aligned} \tag{8}$$

e a forma da transformação TPS para duas dimensões fica:

$$x' = a_0 + a_1 x + a_2 y + \sum_{i=1}^{N} F_i r_i^2 \log r_i \tag{9}$$

$$y' = b_0 + b_1 x + b_2 y + \sum_{i=1}^{N} G_i r_i^2 \log r_i \tag{10}$$

Reescrevendo as Equações (9) e (10) na forma matricial e com N=4 pontos de controles é obtido a Equação (11):

$$\begin{bmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ x'_3 & y'_3 \\ x'_4 & y'_4 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & r_{1,1}^2 \log r_{1,1}^2 & r_{1,2}^2 \log r_{1,2}^2 & r_{1,3}^2 \log r_{1,3}^2 & r_{1,4}^2 \log r_{1,4}^2 \\ 1 & x_2 & y_2 & r_{2,1}^2 \log r_{2,1}^2 & r_{2,2}^2 \log r_{2,2}^2 & r_{2,3}^2 \log r_{2,3}^2 & r_{2,4}^2 \log r_{2,4}^2 \\ 1 & x_3 & y_3 & r_{3,1}^2 \log r_{3,1}^2 & r_{3,2}^2 \log r_{3,2}^2 & r_{3,3}^2 \log r_{3,3}^2 & r_{3,4}^2 \log r_{3,4}^2 \\ 1 & x_4 & y_4 & r_{4,1}^2 \log r_{4,1}^2 & r_{4,2}^2 \log r_{4,2}^2 & r_{4,3}^2 \log r_{4,3}^2 & r_{4,4}^2 \log r_{4,4}^2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 & x_4 \\ 0 & 0 & 0 & y_1 & y_2 & y_3 & y_4 \end{bmatrix} \begin{bmatrix} a_0 & b_0 \\ a_1 & b_1 \\ a_2 & b_2 \\ F_1 & G_1 \\ F_2 & G_2 \\ F_3 & G_3 \\ F_4 & G_4 \end{bmatrix} \tag{11}$$

Para encontrar a correspondência através da transformação TPS é necessário utilizar pontos de controles ou pontos correspondentes. Desta forma a TPS realiza um mapeamento entre dois conjunto de pontos: $P \rightarrow Q$, ou seja $(x_1, x_2, ..., x_N; y_1, y_2, ..., y_N) \rightarrow (x'_1, x'_2, ..., x'_N; y'_1, y'_2, ..., y'_N)$. Com um conjunto de N pontos correspondentes é possível determinar os $2(N+3)$ coeficientes do mapeamento TPS. Ainda é possível escrever a Equação (11) de uma forma mais compacta conforme a Equação (12),

$$\begin{bmatrix} Q \\ 0 \end{bmatrix} = \begin{bmatrix} P & K \\ 0 & P^T \end{bmatrix} \begin{bmatrix} A \\ W \end{bmatrix} \tag{12}$$

onde

$$K = \begin{bmatrix} 0 & r_{1,2}^2 \log r_{1,2}^2 & r_{1,3}^2 \log r_{1,3}^2 & r_{1,4}^2 \log r_{1,4}^2 \\ r_{2,1}^2 \log r_{2,1}^2 & 0 & r_{2,3}^2 \log r_{2,3}^2 & r_{2,4}^2 \log r_{2,4}^2 \\ r_{3,1}^2 \log r_{3,1}^2 & r_{3,2}^2 \log r_{3,2}^2 & 0 & r_{3,4}^2 \log r_{3,4}^2 \\ r_{4,1}^2 \log r_{4,1}^2 & r_{4,2}^2 \log r_{4,2}^2 & r_{4,3}^2 \log r_{4,3}^2 & 0 \end{bmatrix}_{N \times N} \tag{13}$$

K é a matriz que leva em conta a energia potencial entre os conjuntos de pontos.

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \\ 1 & x_4 & y_4 \end{bmatrix}_{N \times 3} \tag{14}$$

P é a matriz contendo os pontos do espaço inicial.

$$Q = \begin{bmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ x'_3 & y'_3 \\ x'_4 & y'_4 \end{bmatrix}_{N \times 2} \tag{15}$$

Q é a matriz contendo os pontos do espaço final.

$$A = \begin{bmatrix} a1 & b1 \\ a2 & b2 \\ a3 & b3 \end{bmatrix}_{3 \times 2} \tag{16}$$

A é a matriz contendo os coeficientes da parte afim da transformação.

$$W = \begin{bmatrix} F1 & G1 \\ F2 & G2 \\ F3 & G3 \\ F4 & G4 \end{bmatrix}_{N \times 2} \tag{17}$$

W é a matriz contendo os coeficientes da parte não-afim da transformação. Esse problema recai em dois casos:

Caso1: Os conjuntos de pontos (x, y) e (x', y'), ou seja antes e depois da transformação são conhecidos. O sistema linear expresso pela Equação (11) pode ser reescrito na forma da Equação (18) e o sistema linear pode ser resolvido para encontrar os coeficientes da transformação TPS.

$$\begin{bmatrix} A \\ W \end{bmatrix} = \begin{bmatrix} P & K \\ 0 & P^T \end{bmatrix}^{-1} \begin{bmatrix} Q \\ 0 \end{bmatrix} \tag{18}$$

Caso2: O conjunto de pontos (x, y) e os coeficientes são conhecidos. Neste caso, resolve-se o sistema linear expresso pela Equação (12) e o conjunto de pontos (x', y') são encontrados.

Uma vez conhecidos as coordenadas nos dois espaços e os coeficientes é possível calcular a energia de deformação que é proporcional a I_f. Essa energia de deformação quando utilizada com os conjuntos de pontos da face em dois instantes distintos reflete na energia de deformação da face e é expressa pela Equação (19). Essa energia é utilizada como uma característica na terceira abordagem de PAD na Seção 3.3 e recai no caso 1 onde são conhecidos ambos os conjuntos de pontos das faces.

$$I_F \sim WKW^T \tag{19}$$

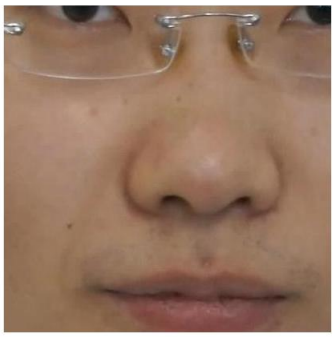
Retomando, nesta primeira abordagem de PAD, os parâmetros de translação invariante à rotação e escala entre os blocos são recuperados e assume-se que ainda haja alguma deformação entre os blocos. Dessa forma, sucessivas transformações do tipo TPS são aplicadas para encontrar o melhor casamento entre os blocos. No casamento entre blocos é utilizado um método de pesquisa *simplex* (LAGARIAS *et al.*, 1998) onde a estimativa inicial do casamento entre os blocos é o bloco $I^k_{t=1}$ e o bloco J^k_t corrigido pela translação através do método descrito na Seção 3.1.2. Nessa busca, utilizam-se N pontos de controles igualmente espaçados e sucessivas transformações TPS são realizadas variando $2(N + 3)$ coeficientes de mapeamento TPS. Esse procedimento recai no caso 2 abordado anteriormente onde o conjunto de pontos (x, y) e os coeficientes são conhecidos. Para encontrar os melhores coeficientes que ajustam um par de blocos é utilizada a métrica MSE. A Figura 18 ilustra o resultado após a deformação de cada bloco da face a fim de casar com a face de referência. Assim as Figuras 18 (c), (e) e (g) assemelham-se com a Figura 18 (a).

Figura 18: Ilustração do método de casamento de blocos, (a) face de referência $I_{t=1}$, (b) uma face no instante $J_{t=5}$, (c) $J_{t=5}$ corrigido, (d) uma face no instante $J_{t=23}$, (e) $J_{t=23}$ corrigido, (f) uma face no instante $J_{t=192}$ e (g) $J_{t=192}$ corrigido.

(a)

(b)

(c)



(d)

(e)

(f)

(g)

3.1.4 Características Extraídas

Resumindo o método proposto até o momento, para cada quadro de um vídeo, a face do indivíduo é encontrada, define-se uma face de referência e para todas as outras faces deste mesmo vídeo são casadas com a face de referência. Essa tarefa é realizada utilizando a transformação log-polar e o mapeamento TPS aplicado em cada bloco que constitui a face.

Uma vez que as faces foram corrigidas, características de texturas são extraídas pelo extrator de feição proposto Log-Radial Diff (LRD). O extrator proposto LRD é uma adaptação do extrator de características ordenado e projetado aleatoriamente. Como reportado pelos autores LIU *et al.* (2011), ao contrário de extratores de características especializados, ajustados para uma base de dados particular, projeção aleatória (RP) (DASGUPTA; GUPTA, 2003; ACHLIOPTAS, 2003) refere-se a uma técnica de projeção de pontos em um espaço de alta dimensionalidade para um subespaço de baixa dimensionalidade escolhido de forma aleatória. Essa técnica tem sido utilizada para otimização combinatorial, recuperação de informação, reconhecimento de face (WRIGHT *et al.*, 2009) e aprendizado de máquina. Características aleatórias representam um maneira computacionalmente simples e eficiente de preservar estrutura da textura sem introduzir distorções significantes.

A capacidade de preservação informação e redução de dimensionalidade é firmemente demonstrada pela teoria de *compressed sensing* (CS), a qual estabelece para sinais esparsos e compressíveis, um pequeno número de medidas lineares não-adaptativas na forma de projeções aleatórias podem capturar a maioria das informações relevantes em um sinal. Além disso, RP provê uma solução viável para o lema de Johnson-Lindenstrauss (DASGUPTA; GUPTA, 2003; ACHLIOPTAS, 2003) que estabelece para um conjunto de pontos em um espaço euclidiano de alta dimensionalidade pode ser mapeado para um espaço de dimensão logarítmica no número de pontos com as distâncias entre pontos aproximadamente preservada.

O extrator proposto LRD é aplicado a cada bloco da face J_t^k corrigido e consiste em transformar um bloco J_t^k corrigido para o espaço log-polar, calcular as diferenças entre as colunas, ordenar os valores de forma ascendente, concatenar estes resultados e projetar com uma base aleatória. O Algorítmo 1 descreve o extrator LRD:

Além das características de texturas extraídas pelo LRD são utilizados as translações ao longo do eixo x e y de cada bloco da face. Desta forma a matriz característica é a concatenação das informações de texturas e da informação de movimento

3.1.5 Escolha da Projeção

A projeção aleatória difere de outros tipos de projeções que se ajustam para uma base de dados particular. Esta técnica projeta pontos de um espaço de alta dimensionalidade para um subespaço de baixa dimensionalidade escolhido de forma aleatória (ACHLIOPTAS, 2003; DASGUPTA; GUPTA, 2003). Essa técnica tem sido utilizada para recuperação de informação, reconhecimento de face (WRIGHT *et al.*, 2009) e aprendizado de máquina.

A matriz de projeção aleatória $R(i,j)$ com elementos r_{ij} utilizada nesta abordagem é a seguinte:

$$r_{ij} = \sqrt{3} \begin{cases} +1 & \text{com probabilidade } 1/6 \\ 0 & \text{com probabilidade } 2/3 \\ -1 & \text{com probabilidade } 1/6 \end{cases} \tag{20}$$

Outro tipo de projeção foi testado. A análise de componentes principais(PCA) é utili-

Algoritmo 1: Método de extração de característica Log-Radial Diff (LRD).

Entrada: Um bloco da face corrigido J_t^k com tamanho $M \times M$
Saída: Feição LRD y_t^k

Computa P, a transformação log-polar de J_t^k;
Seja $P(i,j) = [P^R(i,j), P^G(i,j), P^B(i,j)]$ um pixel no espaço de cor RGB que compõem um bloco;
para $j \leftarrow 2$ ***até*** M **faça**
    para $j \leftarrow 1$ ***até*** M **faça**
        $R(i) \leftarrow P^R(i,j) - P^R(i,j-1)$;
        $G(i) \leftarrow P^G(i,j) - P^G(i,j-1)$;
        $B(i) \leftarrow P^B(i,j) - P^B(i,j-1)$;
    $R^S \leftarrow ordena\{R\}$;
    $G^S \leftarrow ordena\{G\}$;
    $B^S \leftarrow ordena\{B\}$;
    $x \leftarrow [x, R^S, G^S, B^S]^T$
Computa Φ, a matriz de projeção aleatória;
Calcula a feição LRD:
$y_k = \Phi x$;
retorna y_t^k.

zada em redução de dimensionalidade, compressão de dados, extração de feições e visualização de dados. É uma projeção ortogonal dos dados para um espaço linear de baixa dimensionalidade e representa os dados em uma dimensão reduzida maximizando a variância dos dados projetados.

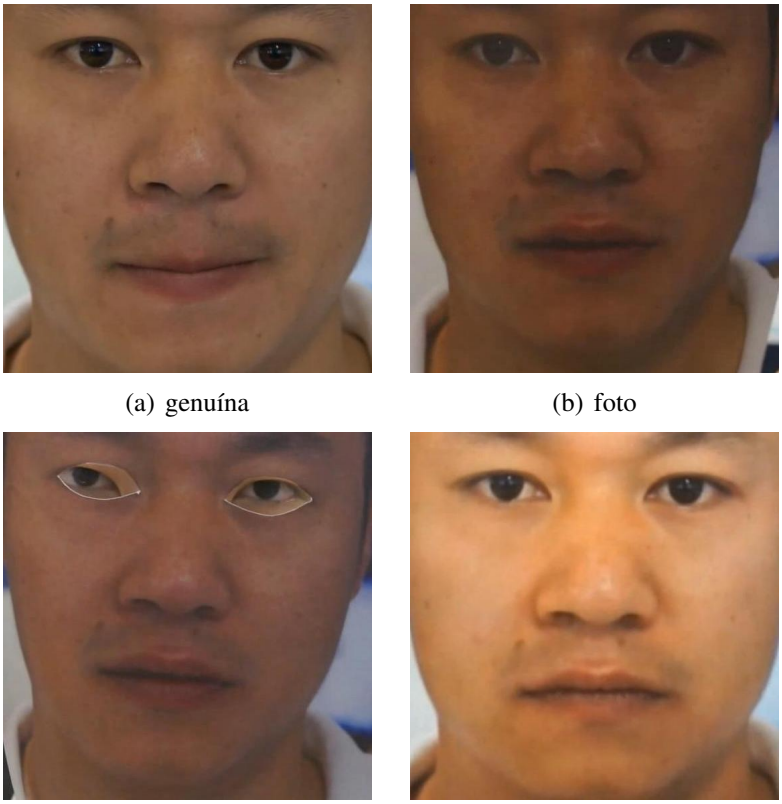
Além da projeção aleatória e a representação dos dados em suas componentes principais, foi estudado a análise de discriminante linear (LDA). É um método popular para extração de feições que preserva a separabilidade de classe. As funções de projeção de LDA são obtidas de forma a maximizar a covariância entre classes e minimizara covariância dentro da classe.

Os autores (CAI; HE; HAN, 2008) propõem um novo algoritmo para análise de discriminantes. Os dados são projetados utilizando PCA, evitando problemas de singularidade. Com isso, é possível selecionar a percentagem da soma dos maiores autovalores no passo da PCA. Em sequência, LDA é aplicada para encontrar um espaço que melhor separa as classes.

Nesta abordagem é utilizado um classificador supervisionado Máquinas de Vetores de Suporte (SVMs) *Kernel* de função de base radial (RBF) $K(x, x') = e^{-\gamma\|x-x'\|^2}$. Onde as informações de textura e movimento compõem a matriz característica. Dessa forma o classificador SVM-RBF é treinado utilizando as informações dos blocos ao longo do conjunto treino. Nesse método 1 é utilizado um classificador específico por usuário e a Figura 19 ilustra 4 quadros, um de cada classe de um único usuário da base de dados CASIA (ZHANG *et al.*, 2012). Ao longo de cada um do vídeos desse usuário, as faces são extraídas, divididas em blocos e casadas. A partir dos blocos corrigidos são extraídas as feições utilizadas pelo classificador. Essas feições são os parâmetros de translação dos blocos e o vetor LRD. A ilustração do classificador customizado ao usuário está na Figura 20 onde as feições extraídas foram projetadas para um espaço de dimensionalidade reduzida igual a 3. É possível verificar que existe uma separação interclasse visível e um

agrupamento intraclasse desejado principalmente nas classes de ataques representado na Figura 20 nas cores vermelho, verde e azul. Já a Figura 21 mostra as regiões de fronteira para cada uma das classes. Os resultados experimentais e discussões desta abordagem estão na Seção 4.2.

Figura 19: Exemplo das 4 classes da base CASIA



(a) genuína (b) foto

(c) cortes (d) vídeo

Fonte: (ZHANG *et al.*, 2012)

3.1.6 Observações

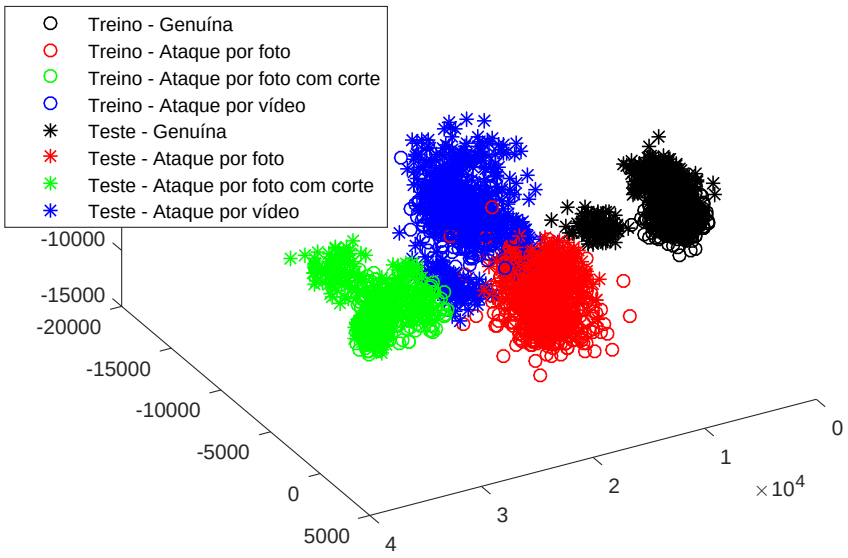
3.1.6.1 Vantagens e Desvantagens do Classificador SVMs

Segundo os autores LORENA; CARVALHO (2007), as SVMs constituem uma técnica de aprendizado que vem recebendo crescente atenção da comunidade de Aprendizado de Máquina. Os resultados da aplicação dessa técnica são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNAs). Exemplos de aplicações de sucesso podem ser encontrados em diversos domínios, como na categorização de textos [19], na análise de imagens e em Bioinformática.

As SVMs são embasadas pela teoria de aprendizado estatístico, desenvolvida por Vapnik(VAPNIK, 1995). Essa teoria estabelece uma série de princípios que devem ser seguidos na obtenção de classificadores com boa generalização, definida como a sua capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado ocorreu.

Os autores (LORENA; CARVALHO, 2007) citam algumas vangens no uso desse classificador. As SVMs são robustas diante de dados de grande dimensão, sobre os quais

Figura 20: Projeção dos pontos em baixa dimensionalidade



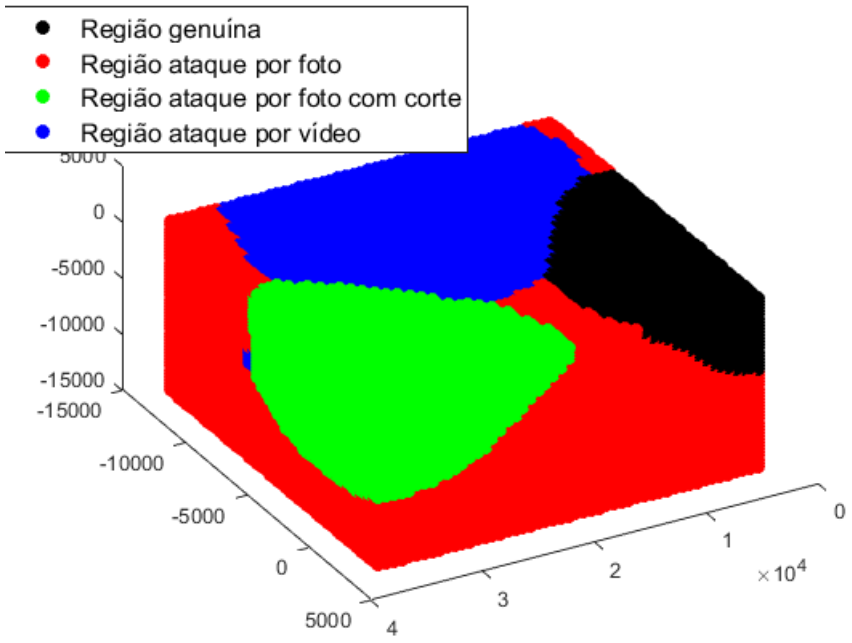
outras técnicas de aprendizado comumente obtêm classificadores super ou sub ajustados. Outra característica atrativa é a convexidade do problema de otimização formulado em seu treinamento, que implica na existência de um único mínimo global. Essa é uma vantagem das SVMs sobre, por exemplo, as Redes Neurais Artificiais (RNAs) Perceptron Multicamadas (Multilayer Perceptron) (HAYKIN, 1998), em que há mínimos locais na função objetivo minimizada. Além disso, o uso de funções Kernel na não-linearização das SVMs torna o algoritmo eficiente, pois permite a construção de simples hiperplanos em um espaço de alta dimensão de forma tratável do ponto de vista computacional. Isso é enfatizado pelos autores AURIA; MORO (2008) que ao introduzir o *kernel*, as SVMs ganham flexibilidade na escolha da forma do limiar que separa separa duas classes, que não precisa ser linear e nem precisa ter a mesma forma funcional para todos os dados, já que sua função é não-paramétrica e opera localmente. Uma vez que o *kernel* contém implicitamente uma transformação não-linear, nenhuma suposição é necessária sobre a forma funcional da transformação que torna os dados linearmente separáveis. A transformação ocorre implicitamente em uma base teórica robusta e o julgamento de perícia humano não é necessário de antemão.

Entre as principais limitações das SVMs encontram-se a sua sensibilidade a escolhas de valores de parâmetros e a dificuldade de interpretação do modelo gerado por essa técnica, problemas que têm sido abordados em diversos trabalhos recentes, como (IMBAULT; LEBART, 2004) e (CASTRO *et al.*, 2007), respectivamente.

3.1.6.2 Pontos Faciais Utilizando Deep Learning

A detecção de pontos faciais é tradicionalmente abordada como um problema único e independente ZHANG *et al.* (2016). Abordagens populares incluem técnicas de ajuste de modelos (ZHU; RAMANAN, 2012) e métodos baseados em regressão (BURGOS-ARTIZZU; PERONA; DOLLáR, 2013). Por exemplo, SUN; WANG; TANG (2013) propõem para detectar pontos de referência faciais por meio de regressão grosseira-a-fina

Figura 21: Regiões do classificador SVM-RBF



usando uma cascata de redes neurais convolucionais profundas. Este método mostra uma precisão superior em comparação com métodos anteriores como por exemplo (CAO *et al.*, 2012) e sistemas comerciais existentes. No entanto, o método requer uma complexa arquitetura em cascata do modelo profundo.

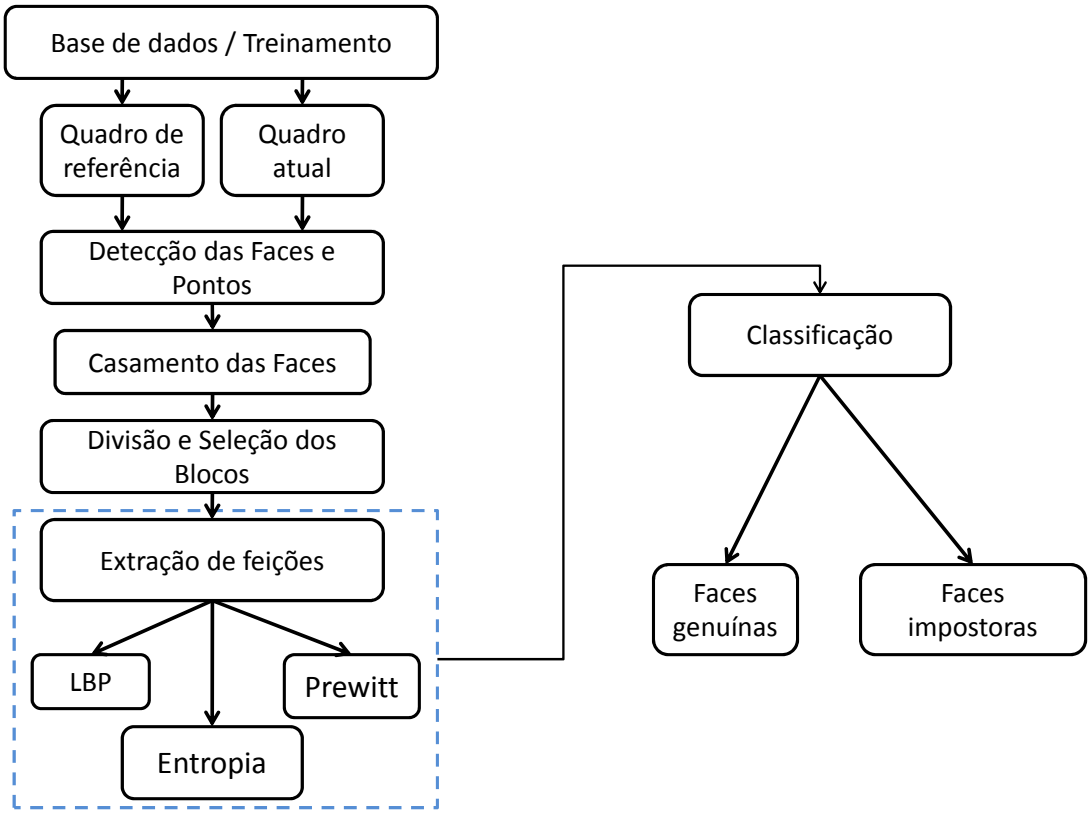
Os autores ZHANG *et al.* (2014b) acreditam que a detecção facial não é um problema autônomo, mas sua estimação pode ser influenciada por vários fatores heterogêneos e sutilmente correlacionados. Por exemplo, quando uma criança está sorrindo, sua boca é amplamente aberta. A descoberta e a exploração eficazes de tal atributo facial intrinsecamente correlacionado ajudariam a detectar mais precisamente os cantos da boca. Além disso, a distância interocular é menor nas faces em grandes ângulos de rotação. Tais informações podem ser aproveitadas como fonte adicional de informação para restringir o espaço da solução dos pontos de referência na face. Dado o rico conjunto de tarefas plausíveis relacionadas, tratar da detecção dos pontos característicos faciais uma tarefa isolada é contra produtivo.

Para este fim, os autores ZHANG *et al.* (2014b) propuseram uma Rede Convolutiva Profunda com Restrições-de-Tarefas (TCDCN) para otimizar conjuntamente a detecção de pontos faciais com um conjunto de tarefas. Especificamente, formularam uma função de perda para a restrição-de-tarefa que permite que os erros das tarefas relacionadas sejam retropropagados conjuntamente para melhorar a generalização da detecção dos pontos. Para acomodar tarefas relacionadas com diferente dificuldades de aprendizado e taxas de convergência, foi elaborado um critério de parada antecipada para facilitar a convergência da aprendizagem.

3.2 Método 2

Neta abordagem é adotado o problema de detecção de ataque de falsificação da face de um usuário como um problema de classificação binária (SOLDERA *et al.*, 2017), de

Figura 22: Diagrama do método 2.



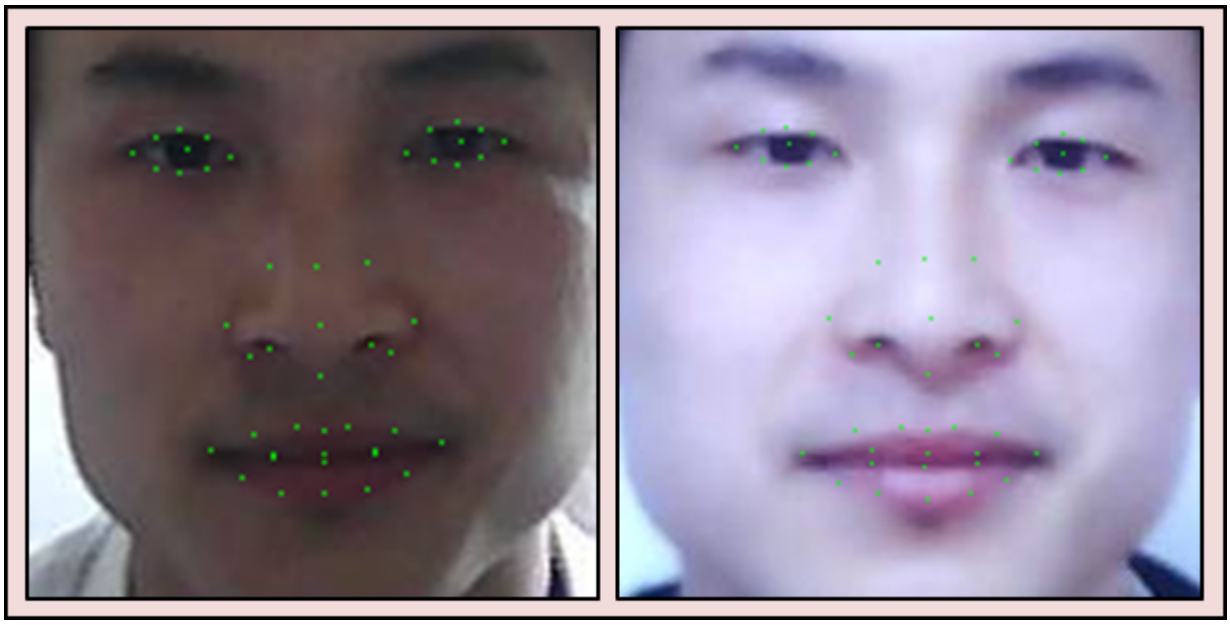
modo que as estatísticas de todo o conjunto de imagens, consistindo de ambos os rostos humanos vivos e fotografias das faces, podem ser totalmente explorados. Inicialmente, cada face é alinhada usando um transformação não linear aplicada aos pontos de referencia da face detectada. Posteriormente, cada rosto é dividido em blocos e características locais são extraídas de cada bloco representando sua cor e textura local. Usando uma função de perda e máquina de vetores de suporte, *Support Vector Machine* (SVMs) como um classificador supervisionado, o melhor conjunto de blocos para comparar imagens faciais são selecionados durante a fase de treinamento e utilizados na fase de testes. O diagrama representado na Figura 22 ilustra as principais etapas desse método. A seguir, a abordagem é apresentada com mais detalhes.

3.2.1 Detecção e Extração dos Pontos Faciais

Para cada quadro, a face é extraída e L pontos de referência são localizado na face usando a abordagem proposta em (MILBORROW; NICOLLS, 2014). Estes autores apresentaram uma técnica para localizar pontos, i.e.*landmarks* em imagens de rostos humanos. Substituíram os perfis de gradiente 1D do Modelo de Forma Ativo, Active Shape Model (ASM) clássico por uma forma simplificada dos descritores da transformação de características invariante a escalas, *Scale-Invariant Feature Transform* (SIFT), e usam Multivariate Adaptive Regression Splines (MARS) para o casamento dos descritores capaz de detectar 68 pontos faciais. Nesta tese foram utilizados os algoritmos disponibilizados pelos autores MILBORROW; NICOLLS (2014) na página (MILBORROW; NICOLLS, 2013). Nesta tese é utilizado um subconjunto de $L = 47$ pontos de referência sendo utili-

zados como pontos fiduciais. Estes pontos fiduciais contém os olhos, nariz e boca, como mostrado na Figura 23, que ilustram os pontos faciais em uma imagem genuína e uma imagem impostora.

Figura 23: Ilustração dos $L = 47$ pontos fiduciais localizados em uma face do genuína (imagem à esquerda) e uma face impostora (imagem à direita). Os pontos representados são um subconjunto dos 68 pontos encontrados através do método (MILBORROW; NICOLLS, 2014)



Fonte: (TAN *et al.*, 2010)

3.2.2 Casamento das Faces

Os pontos da face são usados para realizar o casamento das faces. Ou seja, cada face é alinhada em conformidade com uma face de referência através de uma transformação TPS, que pode ser dividida em uma parte afim e uma parte não-afim. Esta transformação que leva os pontos do espaço deformado P para o espaço Q conforme consta na Seção 3.1.3. Nessa abordagem ambos os conjuntos de pontos (x, y) e (x', y') são conhecidos recaindo no caso 1 abordado anteriormente. Assim o sistema linear expresso pela Equação (12) pode ser resolvido para encontrar os coeficientes da transformação TPS. Uma vez encontrados os coeficientes, aplica-se a transformação para a face inteira de forma que o quadro atual case com o quadro de referência.

3.2.3 Divisão da Face e Características Extraídas

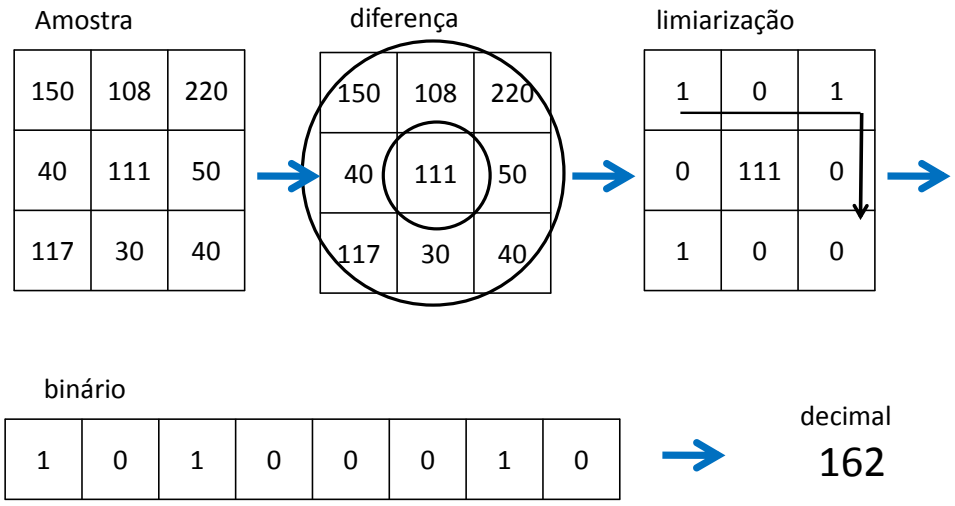
Dadas as imagens faciais alinhadas através dos pontos de fiduciais e a transformação não linear mencionada acima, cada face é dividida em 9 blocos. Para representar as texturas locais de cada bloco, três características texturais diferentes foram avaliadas: Padrão Binário Local (OJALA; PIETIKAINEN; MAENPAA, 2002), *Local Binary Pattern* (LBP), Entropia e as texturas de bordas capturadas usando o operador de Prewitt.

O extrator de textura LBP é capaz de representar um a textura de uma forma muito compacta e eficiente. Esse método é resumido a seguir. Dada uma imagem, o LBP divide essa imagem em células. Cada célula possui um pixel central e uma vizinhança. Para

o exemplo da Figure 24 esse pixel central (111) possui uma vizinhança 8. A próxima etapa consiste em diferença entre o pixel central e a vizinhança seguido pela limiarização do resultado. Quando os pixels da vizinhança são maiores que o pixel central, a posição receba 1, noutro caso 0. Terminada a limiarização, esses valores são concatenados no sentido horário formando um número binário de 8-bit. Esse número é convertido para base decimal e guardado. Esse procedimento é realizado para todas as células da imagem e em seguida é calculado um histograma desses valores. Há diversas variações do extrator de textura LBP, para uma vizinhança 8 e raio 1 a Equação (21) rege esse extrator.

$$LBP_{P=8,R=1} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \tag{21}$$

Figura 24: Método LBP (OJALA; PIETIKAINEN; MAENPAA, 2002)



Já a entropia é uma medida estatística de aleatoriedade que pode ser usada para caracterizar a textura de uma imagem e é expressa pela Equação (22):

$$h = -\sum_{k=0}^{255} p_k(r_k) \log_2 p_r(r_k) \tag{22}$$

onde a o histograma normalizado é definido por:

$$p_r(r_k) = \frac{n_k}{n} \tag{23}$$

sendo n o número total de pixels e $r_k \in [0, 255]$ os tons de cinza.

Enquanto o operador de Prewitt é um detector de borda que utiliza os seguintes *kernels* para representar bordas horizontais e verticais:

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} * J_t^k, G_y = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * J_t^k \tag{24}$$

onde $*$ representa a operação de convolução e J_t^k é um bloco da face. A característica que é utilizada como feição para PAD é a magnitude G do operador Prewitt definida pela Equação (25):

$$G = \sqrt{G_x^2 + G_x^2} \tag{25}$$

O objetivo é extrair as características de textura para cada bloco nos conjuntos de treinamento (ou seja, o conjunto de treinamento das faces genuínas e o conjunto de treinamento das faces de ataque). Durante a fase de treinamento, os blocos são selecionados os melhores blocos que separam as classes de faces genuínas/impostoras. Para encontrar os melhores blocos que representam as classes seriam necessário realizar todas as permutação dos 9 blocos $(9! = 362880)$, mas isto é inviável. Desta forma é realizado uma amostragem aleatória de 1000 permutações apenas. A partir dos blocos selecionados, as características são extraídas e concatenadas em uma matriz. Esta matriz é usada para treinar o algoritmo SVMs juntamente com as etiquetas genuína/impostora. Durante a fase de treinamento (ou seja, usando apenas os conjuntos de treinamento), o classificador SVMs seleciona o melhor subconjunto dos índices de localização dos blocos k, onde k é um subconjunto de permutações aleatórias do conjunto de todos os blocos, através da seguinte função de perda:

$$E = \frac{1}{2} \sum_{n=1}^{|k|} \{y\,(x_n, k) - t_n\}^2 + \frac{\lambda}{2} |k|^2 \tag{26}$$

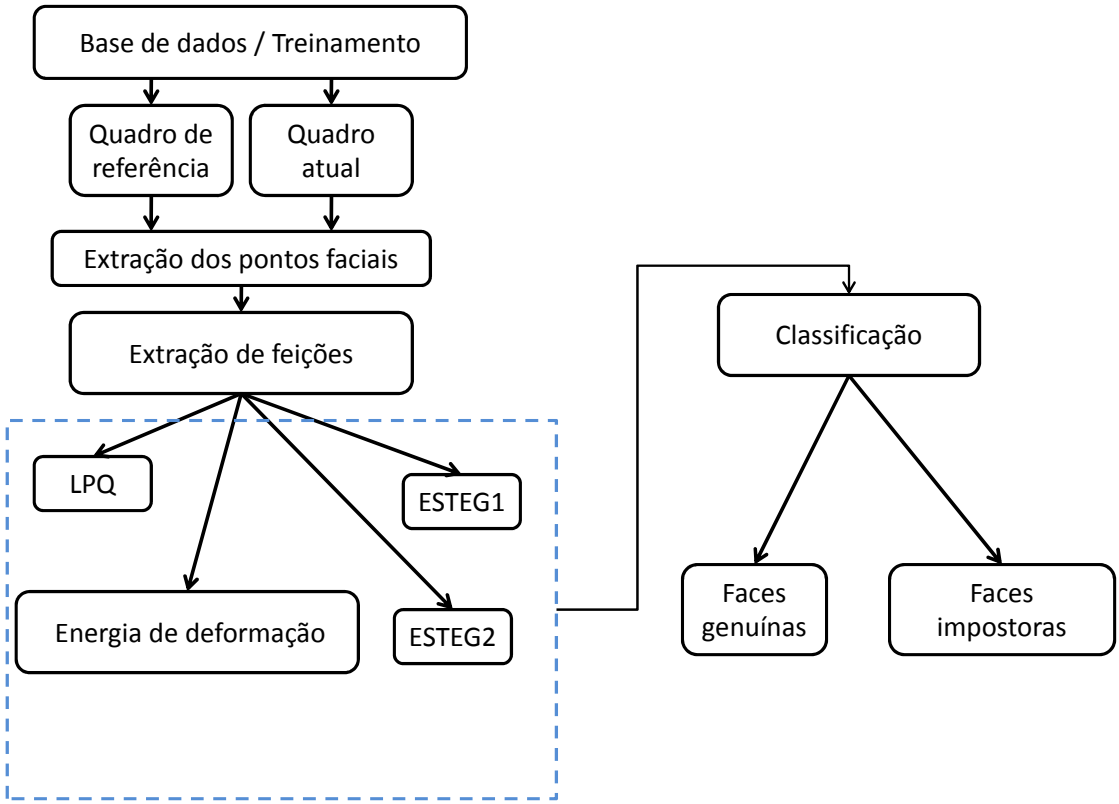
onde $y\,(x_n, k)$ é a etiqueta de saída do classificador SVMs para uma dada amostra de textura x_n, t_n é a etiqueta do conjunto verdade, a constante λ representa um termo de ponderação e $|k|$ é a cardinalidade do subconjunto k. Os índices dos blocos correspondentes que minimizam a função de perda na Eq. 26 na fase de treino são os mesmos utilizados na fase de teste, onde cada novo quadro é dividido em blocos. Usando estas localizações dos blocos (índices) encontradas durante o treinamento, novas amostras são extraídas e utilizadas como entrada no classificador SVMs treinado. O objetivo de utilizar essa função de perda consiste em encontrar o menor número de blocos das faces capaz de discernir as classes genuínas e impostoras.

Este esquema de detecção de falsificação foi testado na base de dados da Universidade de Nanquim de Aeronáutica e Astronáutica (NUAA), que contém um banco de dados de impostor de fotografia disponível publicamente obtido por uma webcam genérica. Os resultados desta abordagem estão na Seção 4.3.

3.3 Método 3

A terceira abordagem deste trabalho visa desenvolver um método de contra-ataques de apresentação capaz de integrar as diversas qualidades ressaltadas na Tabela 2. Isso pode ser atingindo agrupando diferentes técnicas à metodologia proposta. A solução para este problema trata-se de extrair características esteganográficas, informações de deformação da face e texturas da face. Projetar estas características para um espaço discriminante e alimentar um classificador binário cuja resposta será uma etiqueta indicando se a face é genuína ou impostora. Resumindo o método, dado um vídeo da base de dados, define-se um quadro de referência $I_{t=1}$, escolhe-se uma face neutra e os quadros seguintes são definidos por J_t. São extraídos 68 pontos característicos da face (olhos, boca, nariz, sobrancelha e perímetro) utilizando o método (KAZEMI; SULLIVAN, 2014). Uma vez que a face

Figura 25: Diagrama do método 3.



é delimitada utilizando estes pontos, diversas características são extraídas. Para informação de movimentos é calculado a energia de deformação através da transformação *Thin Plate Spline* (TPS) entra cada quadro J_t e o quadro de $I_{t=1}$. Já informações de texturas são extraídos utilizando um descritor de Quantização de Fase Local (*Local Phase Quantization*) (LPQ) (OJANSIVU; HEIKKILÄ, 2008) e métodos de esteganografia (CHEN; SHI, 2008; PEVNY; BAS; FRIDRICH, 2010). As características de movimento e textura formam um vetor característico e podem ser projetadas em uma dimensão reduzida utilizando as técnicas de PCA e/ou LDA. Após isto são classificadas utilizando um classificador de redes neurais artificiais (Ver Seção 3.3.4). Utilizando essa abordagem são construídos os modelos de classe para as faces genuínas e ataques.

3.3.1 Características Derivadas da análise esteganográfica

Esteganografia é uma técnica de segurança da informação que tem a função de ocultar dados em um objeto de cobertura. Para diversos tipos de aplicações, sobretudo as militares, comunicações seguras são de suma importância. Uma vez que o objeto de cobertura oculte a informação, ele passa a ser denominado estego-objeto.

Esteganografia de imagem é para esconder mensagens secretas em imagens para comunicação encoberta, enquanto esteganálise de imagem visa detectar a presença de dados ocultos em uma determinada imagem (CHEN; SHI, 2008). Pesquisadores têm feito esforços para desenvolver esteganografia e esquemas de esteganálise para imagens JPEG (*Joint Photographic Experts Group*) desde 1990. Fundada em 1991, livre e eficiente, o padrão de imagem JPEG tem sido dominante para compressão com perdas de ima-

gem. Atualmente, as imagens JPEG são muito populares na internet. Portanto, a esteganografia de imagem JPEG e esteganálise são de significado prático, sem dúvida. Nos últimos anos, surgiram muitas ferramentas de esteganografia para imagem JPEG, como OutGuess (PROVOS, 2001), F5 (WESTFELD, 2001), e o modelo baseado esteganografia (MB) (SALLEE, 2005), foram publicados e tornaram-se disponível publicamente.

Nesta tese, utiliza-se a esteganálise para encontrar padrões gerados durante ataque de apresentação. Pois durante um ataque a sistemas biométricos faciais a imagem sempre é amostrada um segunda vez. Isto ocorre tanto em uma mídia impressa quanto em mídia digital. São utilizados dois métodos de esteganálise. O primeiro utiliza correlações intrabloco e interbloco entre os coeficientes da imagem JPEG (CHEN; SHI, 2008). O segundo método, utiliza uma forma calibrada dos coeficientes DCT(Transformada Cosseno Discreta) da imagem JPEG (KODOVSKY; FRIDRICH, 2009)

3.3.2 Local Phase Quantization Feature

Segundo os autores (OJANSIVU; HEIKKILÄ, 2008) o modelo discreto invariante espacialmente para o borramento (blurring) pode ser expresso pelo função:

$$g(x, y) = (f * h)(x, y)) \tag{27}$$

Onde $f(x, y)$ é a imagem original, $g(x, y)$ é a imagem observada e $h(x, y)$ é o espectro de frequência do desfoque, em inglês *point spread function* (PSF). No domínio de Fourier isto corresponde:

$$G(\mathbf{u}) = F(\mathbf{u}) \cdot h(\mathbf{u}) \tag{28}$$

onde $G(\mathbf{u})$, $F(\mathbf{u})$ and $h(\mathbf{u})$ são as tranformadas de Fourier Discreta(DFT) da imagem borrada $g(x, y)$, da imagem original $f(x, y)$ e da PSF $h(x, y)$, respectivamente. No domínio da frequência é possível separar a magnitude e a fase na Eq. 28 da seguinte forma:

$$\begin{aligned} |G(\mathbf{u})| &= |F(\mathbf{u})| \cdot |h(\mathbf{u})| \\ \angle G(\mathbf{u}) &= \angle F(\mathbf{u}) \cdot \angle h(\mathbf{u}) \end{aligned} \tag{29}$$

Assumindo que o borramento PSF $h(x, y)$ simetricamente central, tal que $h(x, y) = h(-x, -y)$, assim a sua transformada de Fourier é sempre real e consequentemente a sua fase é uma função de dois valores, dado por:

$$\angle h(u, v) = \begin{cases} 0, & \text{se } h(u, v) \geq 0 \\ \pi, & \text{se } h(u, v) < 0 \end{cases} \tag{30}$$

consequentemente

$$\angle G(u, v) = \angle F(u, v) \text{ para todo } h(u, v) \geq 0 \tag{31}$$

Desta forma a fase da imagem observada $\angle G(u, v)$ é invariante a borramento centralmente simétrico nas frequências onde $h(u, v)$ é positivo. Em caso ideais de movimento e borramentos devido ao desfocamento, a seção transversal de $h(x)$ é retangular (BANHAM; KATSAGGELOS, 1997).Isto resulta em um espectro $h(u, v)$ cuja secção transversal é uma função $sinc$ contendo também valores negativos. Os valores de $h(u, v)$ são sempre positivos antes do primeiro cruzamento por zero na frequência $\approx$ (comprimento do borrão) / (frequência de amostragem) que satisfaz (BANHAM; KATSAGGELOS, 1997).

Para o caso do PSF gaussiano, que modela o borrão de turbulência atmosférica (BANHAM; KATSAGGELOS, 1997), $h(u, v)$ é também gaussiana com apenas valores positivos que sempre satisfazem a condição (BANHAM; KATSAGGELOS, 1997).

Na prática, a invariância ao borramento não pode ser completamente alcançada por causa do tamanho finito das imagens observadas. A convolução da imagem ideal com o borrão PSF se estende para além das fronteiras da imagem observada, desta forma parte da informação é perdida. Quando a extensão do desfoque é grande o suficiente em comparação com o tamanho da imagem, este efeito de borda torna-se perceptível.

3.3.2.1 Transformada de Fourier de curto termo

O método de quantização de fase local (LPQ) é fundamentado na invariância de desfoque pela propriedade do espectro da fase de Fourier descrito pela 31. A LPQ utiliza a informação de fase local extraída com 2-D DFT ou, mais precisamente, transformada de Fourier de curto termo (STFT) computada sobre uma vizinhança $\mathcal{N}_x$ retangular $M \times M$ em cada posição de pixel x da imagem $f(x)$ definida por:

$$F(\mathbf{u}, \mathbf{x}) = \sum_{y \in \mathcal{N}_x} f(\mathbf{x} - \mathbf{y}) e^{-j2\pi \mathbf{u}^T \mathbf{y}} = \mathbf{w}_\mathbf{u}^T \mathbf{f}_\mathbf{x} \tag{32}$$

onde $\mathbf{w_u}$ é o vetor base da DFT na frequência $\mathbf{u}$, e $\mathbf{f_x}$ é outro vetor contendo todas as M^2 amostras de imagens de $\mathcal{N}_x$. Como pode ser notado a partir de 32, uma maneira eficiente de implementar o SIFT é utilizar as convoluções 2-D $f(x)e^{-j2\pi \mathbf{u}^T \mathbf{x}}$ para todos os $\mathbf{u}$. Uma vez que as funções de bases são separáveis, o cálculo pode ser executado usando convoluções 1-D para as linhas e colunas sucessivamente. No LPQ apenas quatro coeficientes complexos são considerados, correspondendo a frequências 2-D $\mathbf{u1} = [a, 0]^T$, $\mathbf{u2} = [0, a]^T$, $\mathbf{u3} = [a, a]^T$ e $\mathbf{u4} = [a, -a]^T$ onde a é uma frequência escalar abaixo do primeiro cruzamento por zero de condição 31. Sejam:

$$\mathbf{F}_\mathbf{x}^c = \left[F\left(\mathbf{u}_1, x\right), F\left(\mathbf{u}_2, x\right), F\left(\mathbf{u}_3, x\right), F\left(\mathbf{u}_4, x\right)\right], \tag{33}$$

e

$$\mathbf{F}_\mathbf{x} = \left[Re\left\{\mathbf{F}_\mathbf{x}^c\right\}, Re\left\{\mathbf{F}_\mathbf{x}^c\right\}\right]^T \tag{34}$$

onde $Re\{.\}$ e $Im\{.\}$ retornam partes reais e imaginárias de um número complexo, respectivamente. A matriz de transformação $8 \times M^2$ correspondente é:

$$\mathbf{W} = \left[Re\left\{\mathbf{w_{u_1}}, \mathbf{w_{u_2}}, \mathbf{w_{u_3}}, \mathbf{w_{u_4}}\right\}, Im\left\{\mathbf{w_{u_1}}, \mathbf{w_{u_2}}, \mathbf{w_{u_3}}, \mathbf{w_{u_4}}\right\}\right]^T, \tag{35}$$

então

$$\mathbf{F_x} = \mathbf{W}\mathbf{f_x} \tag{36}$$

3.3.2.2 Análise Estatística dos Coeficientes

Assumindo que a função de imagem $f(x)$ é um resultado de um processo Markov de primeira ordem, onde o coeficiente de correlação entre os valores de pixel adjacentes é ρ, e a variância de cada amostra é σ^2. Sem perda de generalidade, é possível supor que $\sigma^2 = 1$. Como resultado, a covariância entre as posições $\mathbf{x}_i$ e $\mathbf{x}_j$ torna-se

$$\sigma_{ij} = \rho^{||\mathbf{x}_i - \mathbf{x}_j||} \tag{37}$$

onde $||.||$ denota a norma L_2, e a matriz de covariância para todas as M amostras em $\mathcal{N}_x$ podem ser expressadas por

$$\mathbf{C} = \begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & 1 & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & 1 \end{bmatrix} \tag{38}$$

Assim, a matriz de covariância do vetor de coeficiente transformado F_x pode ser obtida a partir de

$$\mathbf{D} = \mathbf{WCW}^T \tag{39}$$

e pode-se notar que $\mathbf{D}$ não é uma matriz diagonal para $\rho > 0$, significando que os coeficientes são correlacionados.

3.3.2.3 Descorrelação e Quantização

Antes da quantização, os coeficientes são descorrelacionados, pois é possível demonstrar que a informação é preservada ao máximo na quantização escalar se as amostras a serem quantizadas forem estatisticamente independentes. Assumindo a distribuição gaussiana, independência pode ser alcançada usando uma transformação *whitening*

$$\mathbf{G_x} = \mathbf{V}^T\mathbf{F_x} \tag{40}$$

onde $\mathbf{V}$ é uma matriz ortonormal derivada da decomposição do valor singular (SVD) da matriz D que é

$$\mathbf{D} = \mathbf{U\Sigma V}^T \tag{41}$$

note que $\mathbf{V}$ pode ser resolvido para um valor fixo de ρ.

Note que $\mathbf{G_x}$ é computado para todas as posições das imagens, ou seja, $\mathbf{x} \in \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$, e os vetores resultantes são quantizados utilizando um quantizador escalar simples

$$q = \begin{cases} 1, & \text{se } g_j \geq 0 \\ 0, & \text{se caso contrário} \end{cases} \tag{42}$$

onde g_j é a j-ésima componente $\mathbf{G_x}$. Os coeficientes quantizados são representados como inteiros entre $0 - 255$ utilizando codificação binária

$$b = \sum_{j=1}^{8} q_j 2^{j-1} \tag{43}$$

Finalmente, um histograma desses valores inteiros de todas as posições da imagem é composto e usado como um vetor de característica de 256 dimensões na classificação.

Os inteiros resultantes b são invariantes ao desfoque simétrico central evidenciando que a janela $\mathcal{N}_x$ é infinitamente grande e o espectro de frequência do desfoque PSF é positivo nos locais das amostras $\mathbf{u_1} - \mathbf{u_4}$. A segunda condição é facilmente atendida se a é suficientemente pequeno. No entanto, a primeira condição não pode ser satisfeita na prática, e, portanto, completa invariância não é alcançada, mas como mostrado nos experimentos, mesmo em uma vizinhança relativamente pequena é suficiente para uma robustez razoável ao desfoque.

Descorrelação e quantização não têm nenhum efeito sobre a propriedade da invariância do desfoque. Na transformação de *whitening*, os vetores de coeficiente estão sujeitos a uma rotação 8-dimensional que causa apenas uma mudança de fase uniforme para todos os vetores. Na quantização, o espaço de 8 dimensões é dividido em 256 hipercubos, e a atribuição de um vetor a um desses hipercubos depende apenas da informação de fase.

3.3.3 Características de Deformação

Nesse método 3, utiliza-se a energia de deformação das faces durante um vídeo ou uma sequência de quadros. Prova-se através dos experimentos relatados a seguir que a informação da energia de deformação é uma característica importante para a PAD. Ataques através de fotografia são mais óbvios de imaginar que a deformação ao longo dos quadros será muito diferente das deformações que ocorrem na face durante um acesso a um sistema biométrico facial. Essas deformações também ocorrem em ataques por vídeos, pois ao reproduzir um vídeo em um tela, a face é apenas uma representação de um objeto 3D em um plano 2D. Como discutido anteriormente na Seção 2.1, isso ocorre pois sempre há movimentos faciais sejam voluntários ou involuntários causando assim variações na energia de deformação entre as faces para diferentes instantes de tempo.

São extraídos 68 pontos característicos da face utilizando o método (KAZEMI; SULLIVAN, 2014). Esses pontos são extraídos para cada face e são definidos como pontos fiduciais. O cálculo de energia de deformação é realizado sempre entre a face de referência $I_{t=1}$ e cada uma das faces da sequência de quadros J_t. Dessa forma são conhecidos ambos os conjuntos de pontos e de acordo com a Seção 3.1.3 essa abordagem recai no caso 1. Após calcular os coeficientes da transformação TPS utilizando a Equação (18) é possível calcular a energia de deformação entre os pares de face com a Equação (19).

3.3.4 Classificador de Redes Neurais Artificiais

Segundo os autores ABIODUN *et al.* (2018), nos últimos anos, as redes neurais artificiais (RNAs) tornaram-se populares e úteis modelo para classificação, agrupamento, reconhecimento de padrões e predição em muitas disciplinas. As RNAs são um tipo de modelo para aprendizado de máquina e relativamente competitiva para regressão convencional e modelos estatísticos. Atualmente, a inteligência artificial (aprendizado de máquina, rede neural, aprendizagem profunda, robótica), segurança da informação, *big data*, computação em núvem, internet, e a ciência forense são pontos de interesse e temas interessantes de tecnologia da informação e comunicação. As aplicações das RNAs podem ser avaliadas com respeito a fatores de análise de dados, como precisão, velocidade de processamento, latência, desempenho, tolerância à falha, volume, escalabilidade e convergência. O grande potencial das RNAs é o processamento de alta velocidade fornecido em uma implementação paralela maciça e isto aumentou a necessidade de pesquisas neste domínio. As RNAs podem ser desenvolvidas e utilizadas para reconhecimento de imagem, processamento de linguagem natural e assim por diante. Hoje em dia, As RNAs são usadas principalmente para a aproximação universal de funções em paradigmas numéricos devido às suas excelentes propriedades de auto-aprendizagem, adaptabilidade, tolerância a falhas, não-linearidade e avanço no mapeamento de entrada e saída.

Esses fatores de análise de dados dão mais razão para que as RNAs sejam eficazes, eficientes e bem-sucedidas em fornecer um alto nível de capacidade no tratamento de problemas complexos e não complexos em muitas esferas da vida. As RNAs são capazes de lidar com problemas na agricultura, ciência, ciência médica, educação, finanças, gestão, segurança, engenharia, negociação de *commodities* e arte. Incluindo problemas na fabricação, transporte, segurança de computadores, bancos, seguros, gerenciamento de propriedades, marketing, energia e os desafios que não podem ser resolvidos pela capacidade computacional dos procedimentos tradicionais e da matemática convencional. Apesar destes extensas aplicações das RNAs, há uma necessidade crescente de abordar o problema da adoção de uma abordagem sistemática na fase de desenvolvimento das RNAs

para melhorar seu desempenho. Por exemplo, uma abordagem para avaliar os principais fatores e tópicos em uma escolha de conjuntos de dados (tamanho, volume, pequeno, grande e outros), a precisão dos dados, instrumento de dados, padronização de dados, tipo de entradas de dados, divisão de dados e pré-processamento de dados, validações, processamento e técnica de saída.

Nessa tese, particularmente no método 3 é utilizado um classificado de redes neurais artificiais do tipo *Scaled Conjugate Gradient*(SCG) para classificar as feições extraídas das faces visando detectar ataques de apresentação a sistemas biométricos faciais. Segundo o autor MøLLER (1993) que desenvolveu o algoritmo de aprendizado supervisionado SCG tem o desempenho comparado com o do algoritmo padrão *back propagation* (BP). O algoritmo SCG é totalmente automatizado, não há parâmetros críticos dependentes de usuário, e evita uma busca demorada de linha, o que outras abordagens utilizam em cada iteração para determinar o tamanho do passo apropriado. Experimentos mostraram que SCG é consideravelmente mais rápido BP e outras abordagens. Foi utilizado o classificador RNA disponível no Matlab (MATLAB NEURAL TOOLBOX 11.0, 2017b) utilizando a função de treinamento SCG com 10 camadas ocultas e utilizado a entropia cruzada como função de desempenho.

4 RESULTADOS EXPERIMENTAIS E DISCUSSÕES

4.1 Análise de Desempenho

A Tabela 6 apresenta a matriz de confusão para o classificador binário SVMs com os rótulos conhecidos para os blocos/faces

Segundo os autores (ANJOS; MARCEL, 2011), um sistema de detecção de fraudes está sujeito a dois tipos de erros, ou o acesso real é rejeitado (falsa rejeição) ou um ataque é aceito (falsa aceitação). Para medir o desempenho de um sistema de detecção de fraudes, é utilizada a *half Total Error Rate*, Taxa da metade do erro total, que combina a *False Acceptance Rate*(FAR), Taxa de falsa aceitação e a *False Rejection Rate*(FRR), Taxa de falsa rejeição que está definido no Capítulo 3, Equação (1) e reescrita aqui $HTER(\tau, D) = \frac{FRR+FAR}{2}$. Onde D denota o conjunto de dados usado. Já que tanto o FAR quanto o FRR dependem do limiar τ, eles estão fortemente relacionados um com o outro: aumentando a FAR reduzirá o FRR e vice-versa. Por esta razão, os resultados são geralmente apresentados utilizando curvas *Receiver Operating Characteristic*(ROC) ou *Detection-Error Tradeoff* (DET) (MARTIN *et al.*, 1997), que basicamente representa a FRR(abcissa) contra FAR(ordenada) para valores diferentes do limiar τ. Outra medida amplamente usada para resumir o desempenho de um sistema é a *Equal Error Rate*(EER), Taxa de Erro Igual, definida como o ponto ao longo da curva ROC ou DET onde o FAR é igual ao FRR.

Em relação aos conjuntos de treino, desenvolvimento e teste, é recomentado que as amostras de treinamento e desenvolvimento sejam usadas para o aprendizado dos classificadores (ANJOS; MARCEL, 2011). Um exemplo trivial é usar o conjunto de treinamento para treinando o próprio classificador e os dados de desenvolvimento para estimar quando parar de treinar. Uma segunda possibilidade, que pode generalizar menos, é fundir ambos os conjuntos de treinamento e desenvolvimento, usando o conjunto mesclado como dados de treinamento e formular um critério de parada. Finalmente, o conjunto de teste deve ser usado exclusivamente para relatar taxas de erro e curvas de desempenho, i.e. curvas ROC e DET. Se um único número é desejado, um limite τ deve ser escolhido no conjunto de desenvolvimento e o HTER relatado deve ser dados do conjunto de teste. Para uniformizar os relatórios, é recomendado a escolha do limiar τ da EER no conjunto de desenvolvimento. Para bases de dados onde não há conjunto de desenvolvimento, este limiar τ deve ser encontrado no conjunto de treino.

Na fase de treinamento, para cada indivíduo, a matriz característica é extraída e uma seleção automática dos blocos é realiza de forma a minimizar a taxa HTER. Dessa forma o Classificador SVMs é treinado especificamente por pessoa e para um determinado conjunto de teste. Durante a fase de testes, as novas faces são avaliadas apenas no conjunto de blocos que foram selecionados previamente. Em outras palavras, apenas nos blocos

		Classificador SVMs	
		Genuína	Impostora
Conjunto verdade	Genuína	Verdadeiro positivo	Falso negativo
	Impostora	Falso positivo	Verdadeiro negativo

Tabela 6: Matriz de Confusão para a classificação das faces

selecionados têm as suas feições extraídas. Uma nova matriz característica é construída contendo as feições do LRD e as informações de movimento. Uma vez que a matriz característica é construída, é submetida ao classificador SVMs treinado.

4.2 Resultados Experimentais do Método 1

A primeira abordagem desta tese visa desenvolver um método contra ataques de apresentação específico por pessoa. Os teste realizados foram com a base de dados CASIA (ZHANG *et al.*, 2012). Durante a fase de treinamento, para cada indivíduo, a matriz característica é extraída e uma seleção automática dos blocos é realizada de forma a minimizar a taxa HTER. Dessa forma o Classificador SVMs é treinado especificamente por pessoa e para um determinado conjunto de teste. Durante a fase de testes, as novas faces são avaliadas apenas no conjunto de blocos que foram selecionados previamente. Em outras palavras, apenas nos blocos selecionados têm as suas feições extraídas. Uma nova matriz característica é construída contendo as feições do LRD e as informações de movimento. Após a construção da matriz característica, ela é submetida ao classificador SVM já treinado.

A primeira análise foi realizada para investigar qual a melhor dimensão da projeção aleatória. A Figura 26 ilustração da dimensão de projeção em função da taxa HTER. As bordas da caixa azul delimitam o $25°$ e $75°$ percentis e os bigodes estendem até os valores extremos. Realizando uma análise de incerteza com 50 repetições para cada indivíduo com intervalo de confiança de 95% é possível determinar a melhor dimensão da projeção aleatória. Estes resultados estão na Tabela 7. A Figura 27 mostra o comportamento melhor para baixas dimensões onde as taxas HTER alcançam valores menores apesar do alto desvio padrão.

Figura 26: Ilustração da dimensão de projeção em função da taxa HTER no conjunto teste

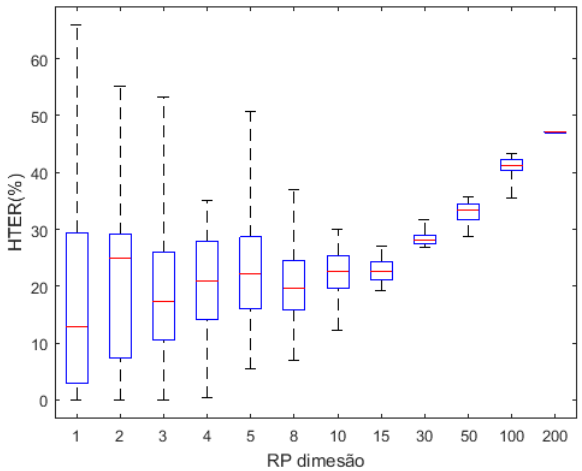
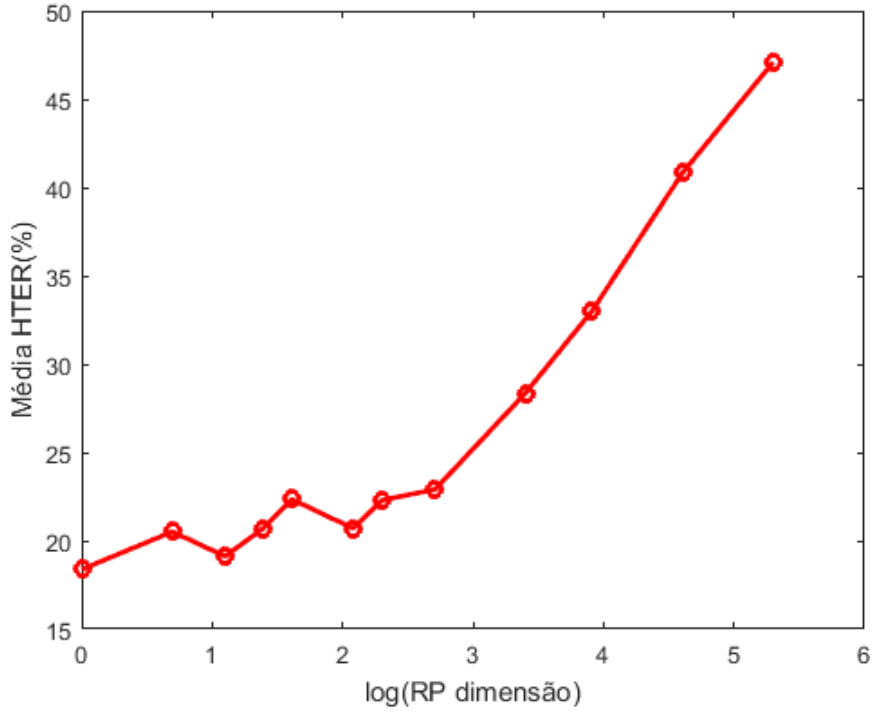
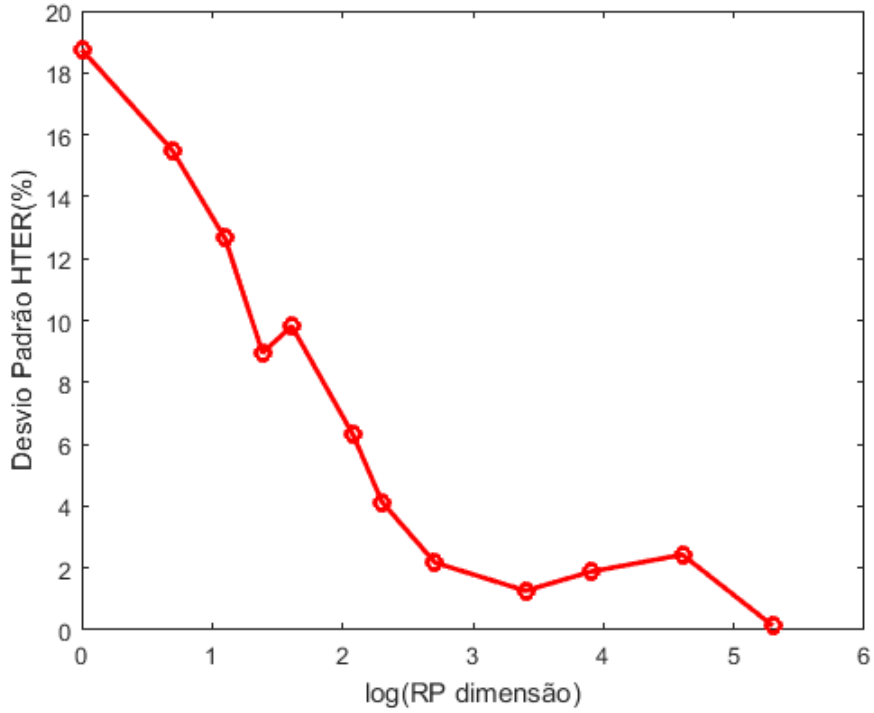


Figura 27: Média e desvio padrão da dimensão de projeção em função da taxa HTER



(a) Média



(b) Desvio padrão

Tabela 7: Intervalo de Confiança para HTER

RP dimensão	Intervalo de Confiança de HTER(%)
1	18.39 ± 5.20
2	20.52 ± 5.48
3	19.13 ± 3.51
4	20.69 ± 3.06
5	22.35 ± 3.46
8	20.71± 2.26
10	22.31± 1.15
15	22.91± 0.78
30	28.37± 0.56
50	33.02± 0.78
100	40.84± 1.32
200	47.09± 0.15

A próxima análise refere-se aos diferentes sujeitos da base de dados CASIA em função da taxa HTER para a dimensão da projeção aleatória ajustada para 3. Os resultados estão ilustrados nas Figuras 28 e 29.

Figura 28: Sujeitos em função da taxa HTER

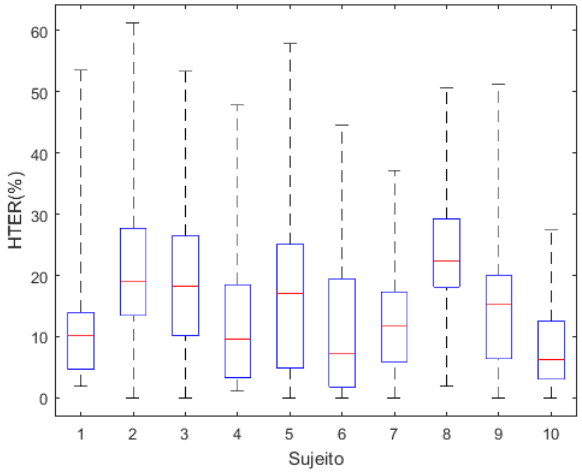
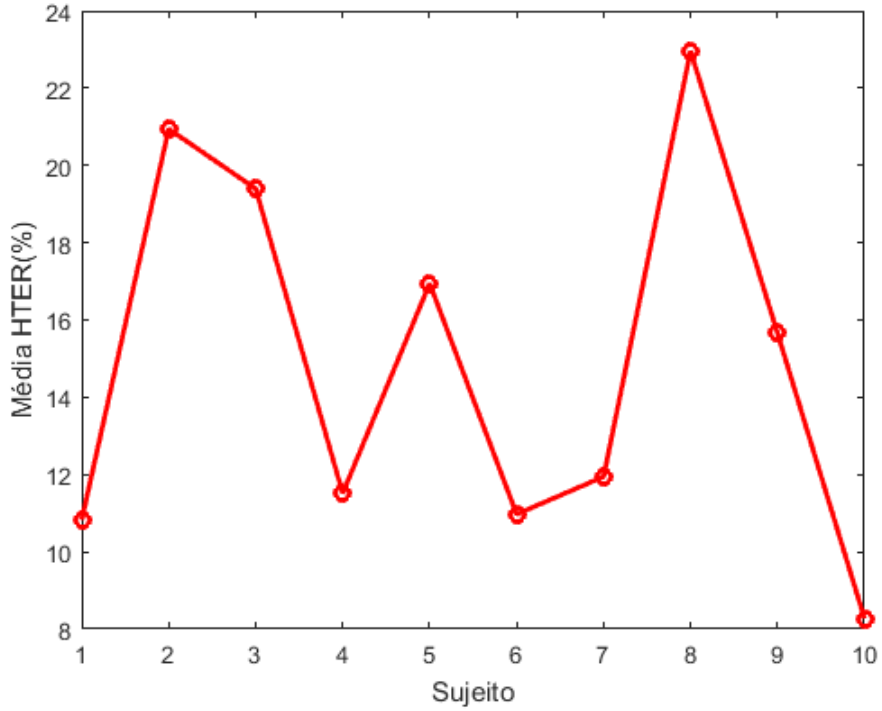
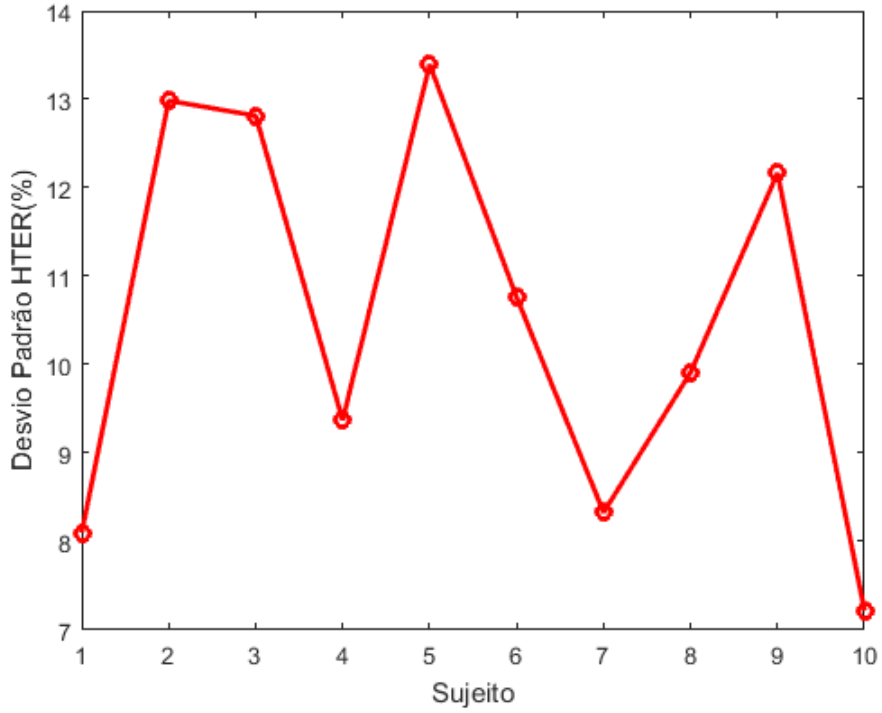


Figura 29: Média e desvio padrão. Sujeitos em função da taxa HTER



(a) Média



(b) Desvio padrão

Realizando uma análise de incerteza com 50 repetições e um intervalo de confiança de 95% é possível determinar aa taxa de erro HTER em função dos sujeitos. Estes resultados estão na Tabela 8

Tabela 8: Intervalo de Confiança dos Sujeitos em função da taxa HTER

Sujeito	Intervalo de Confiança de HTER(%)
1	10.83 ± 2.25
2	20.98 ± 3.60
3	19.40 ± 2.55
4	11.52 ± 2.60
5	16.95 ± 3.71
6	10.98 ± 2.98
7	11.95 ± 2.31
8	22.96 ± 2.75
9	15.66 ± 3.37
10	8.28 ± 2.00
Todos	14.95 ± 1.02

A seguir, os resultados para investigar qual a melhor taxa do número de componentes da PCA durante a projeção PCA-LDA (CAI; HE; HAN, 2008). A Figura 30 e 31 mostram que o melhor compromisso está na faixa da taxa $PCA = 40\%$ e $PCA = 70\%$

Figura 30: HTER em função da taxa PCA

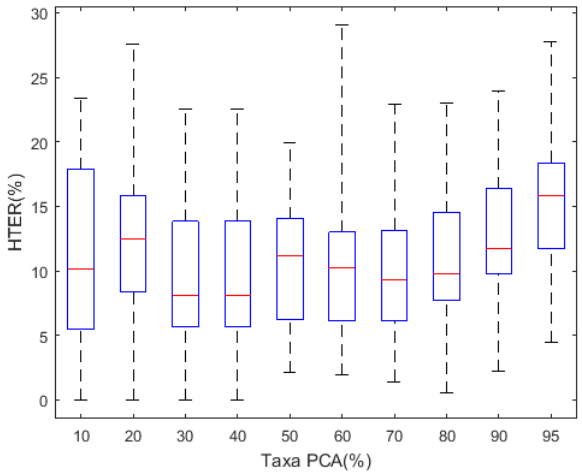
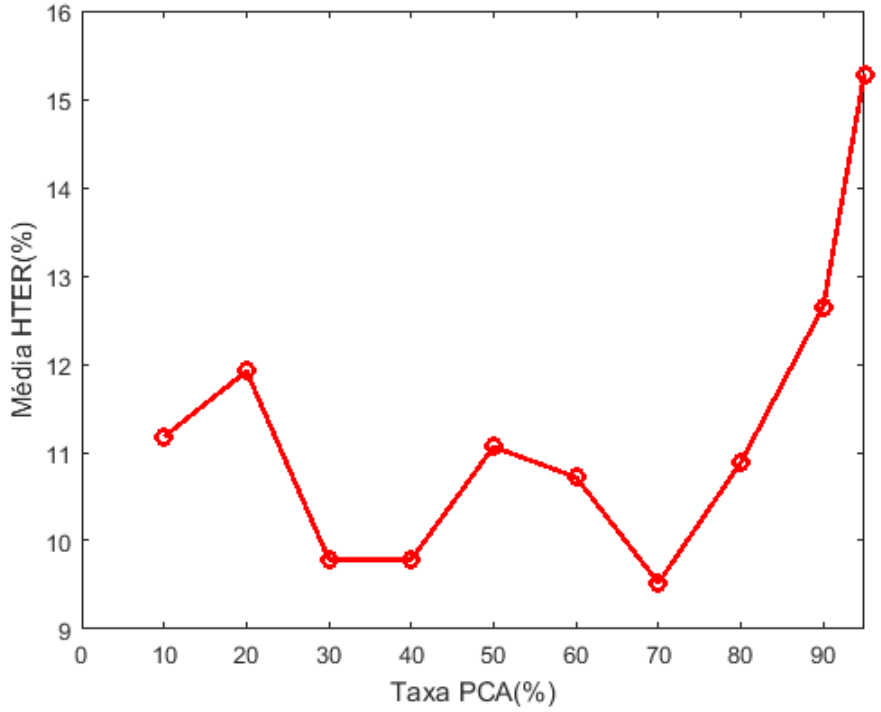
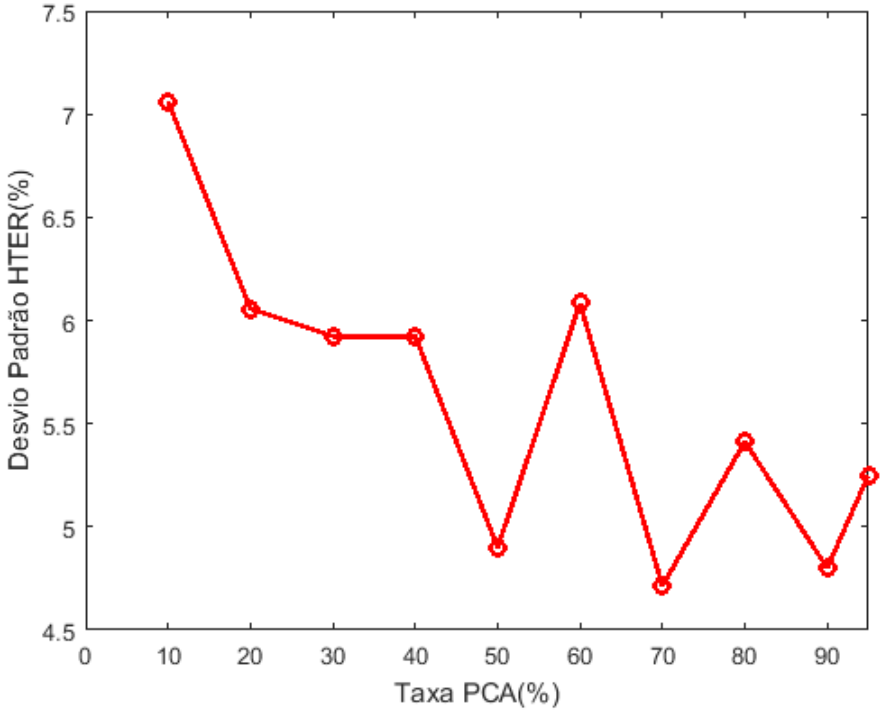


Figura 31: Media e desvio padrão. HTER em função da taxa PCA



(a) Média



(b) Desvio padrão

Ajustando a taxa $PCA = 70\%$ os resultados ilustrados pela Figura 32 com as médias e desvios na Figura 33

Figura 32: Sujeitos em função de HTER com taxa PCA 70%

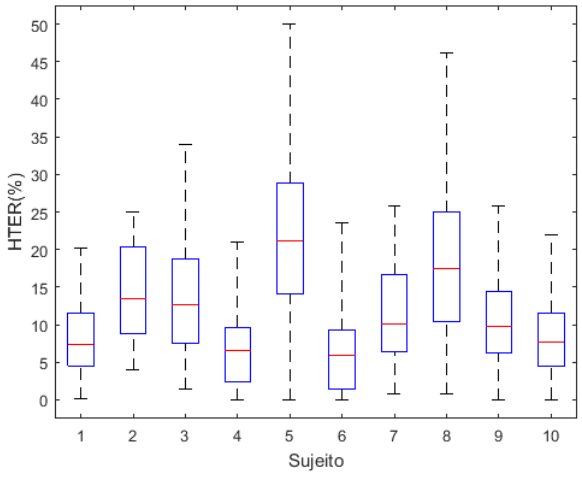
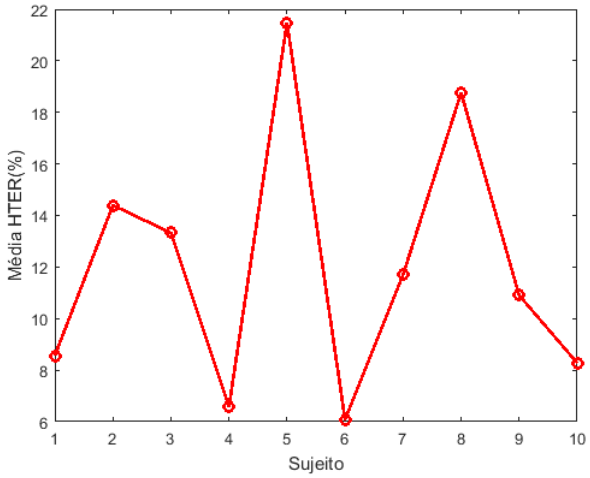
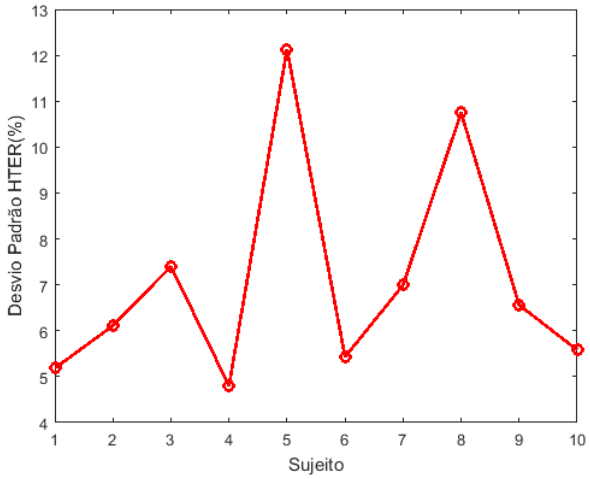


Figura 33: Media e desvio padrão. Sujeitos em função de HTER com taxa PCA 70%



(a) Média



(b) Desvio padrão

E a Tabela 9 resume os resultado dos sujeitos em função do intervalo de confiança HTER com a taxa PCA 70%

Tabela 9: Sujeitos vs HTER com taxa PCA 70% (50 repetições, $\alpha = 0.05$)

Sujeito	Intervalo de Confiança de HTER(%)
1	8.56 ±1.43
2	14.40±1.69
3	13.32± 2.05
4	6.59± 1.33
5	21.48± 3.36
6	6.08± 1.14
7	11.70± 1.94
8	18.76± 2.98
9	10.90± 1.82
10	8.28± 1.55
Todos	11.60± 0.74

Os mesmos experimentos foram realizados com taxa PCA 40%, ver Figs 34 e 35. Assim como a Tabela 10 mostra o intervalo de confiança de HTER para a taxa $PCA = 40\%$.

Figura 34: Sujeitos em função de HTER com taxa PCA 40%

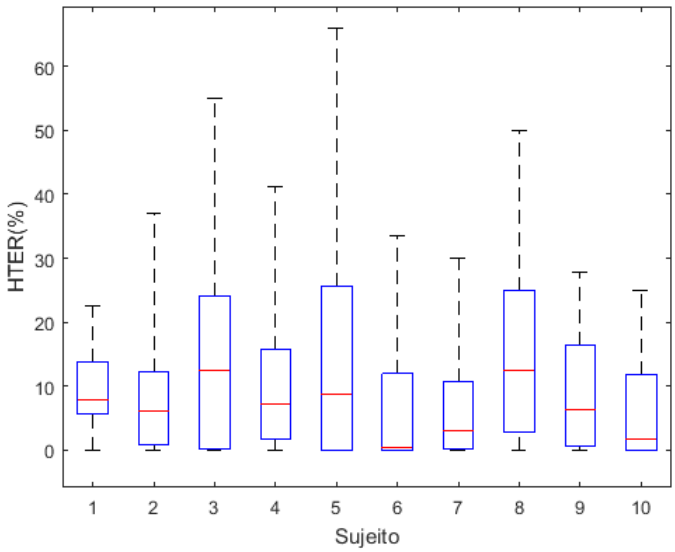
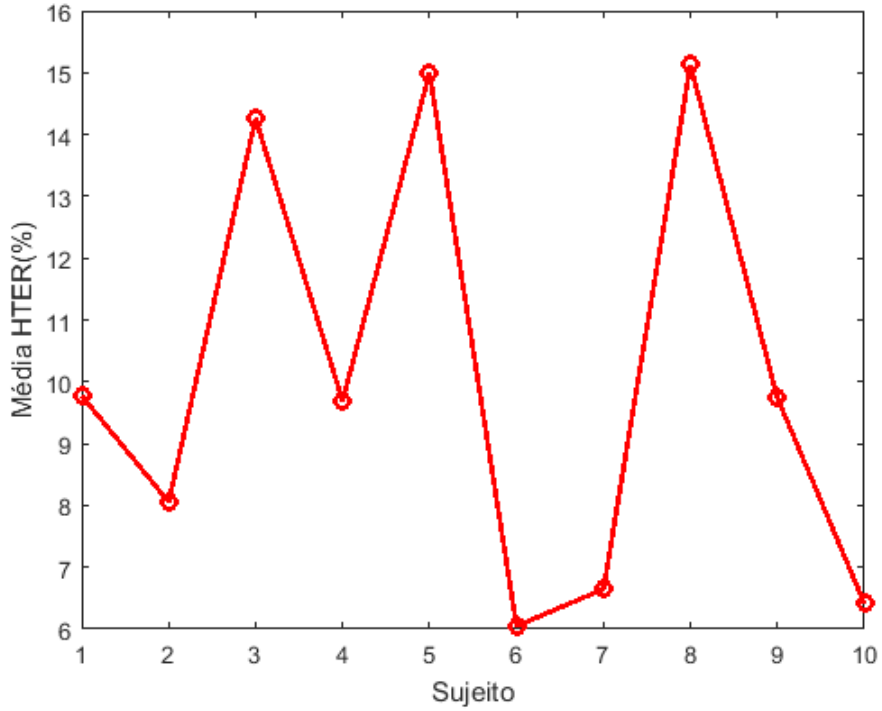
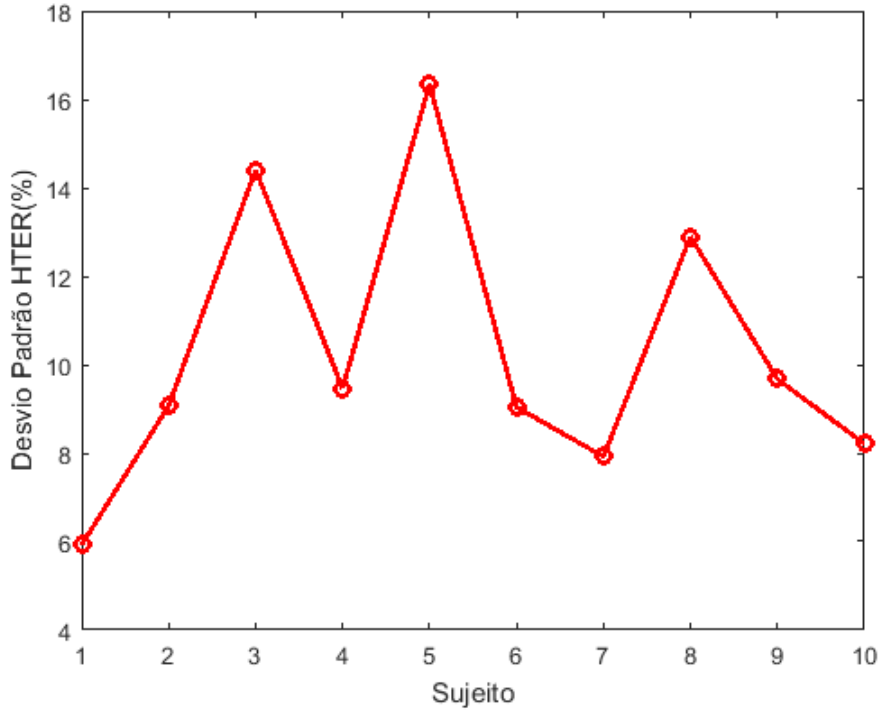


Figura 35: Media e desvio padrão. Sujeitos em função de HTER com taxa PCA 40%



(a) Média



(b) Desvio padrão

Tabela 10: Sujeitos vs HTER com taxa PCA 40% (50 repetições, $\alpha = 0.05$)

Sujeito	Intervalo de Confiança de HTER($\%$)
1	9.78 ± 1.64
2	8.06 ± 2.51
3	14.26 ± 3.99
4	9.69 ± 2.62
5	15.00 ± 4.53
6	6.05 ± 2.67
7	6.65 ± 2.20
8	15.13 ± 3.57
9	9.74 ± 4.35
10	6.41 ± 2.28
Todos	10.15 ± 1.03

Na abordagem pelo método 1 os melhores resultados foram obtidos com a projeção PCA-LDA $HTER = 10.15 \pm 1.03$ com a taxa $PCA = 40\%$ para o conjunto teste na base de dados CASIA (ZHANG *et al.*, 2012).

4.3 Resultados Experimentais do Método 2

Os resultados foram obtidos em termos de Taxa Positiva Verdadeira (TPR), Taxa Negativa Verdadeira (TNR) e Taxa de Detecção Correta (CDR), onde TPR, TNR e CDR tendem a 100% para uma perfeita discriminação entre faces genuínas e faces impostoras. Enquanto o $CDR = (TP + TN)/(TP + FP + FN + TN)$ fornece uma estimativa da robustez do método. A Tabela 11 resume os resultados em termos de TPR, TNR e CDR para o conjunto de treinamento e teste do banco de dados NUAA. É possível observar que na medida que a constante λ está diminuindo, o número de blocos selecionadas aumentam. O melhor resultado ocorre quando $\lambda = 0.0001$ com a seleção automática dos blocos k = 2, 8 e 9. O classificador SVMs atinge $TPR = 97,7\%$, $TNR = 95,7\%$ e $CDR = 96,7\%$ para o conjunto de treinamento e $TPR = 99,4\%$, $TNR = 52,6\%$ e $CDR = 69,8\%$ para o conjunto de teste.

Tabela 11: Resultados Experimentais para Detecção de ataques de apresentação Base NUAA

	Treino		Teste		
λ	TPR	TNR	TPR	TNR	Blocos selecionados
10	0.745	0.887	0.726	0.686	k=8
1	0.745	0.887	0.726	0.686	k=8
0.5	0.745	0.887	0.726	0.686	k=8
0.2	0.745	0.887	0.726	0.686	k=8
0.1	0.745	0.887	0.726	0.686	k=8
0.01	0.948	0.942	0.986	0.512	k=2, 8
0.001	0.948	0.942	0.986	0.512	k=2, 8
0.0001	0.977	0.957	0.994	0.526	k=2, 8, 9
0.00001	0.997	0.967	0.550	0.511	k=2,4,5,6,8,9
0.000001	0.993	0.974	0.533	0.535	k=1,2,3,5,6,7,8,9
0.0000001	0.993	0.974	0.533	0.535	k=1,2,3,5,6,7,8,9

4.4 Resultados Experimentais do Método 3

Nesta terceira abordagem para a detecção de ataques de apresentação foram utilizadas 4 características responsáveis por discernir faces genuínas das impostoras. Estas técnicas foram apresentadas na Seção 3.3 e são elas: o descritor LPQ (OJANSIVU; HEIKKILÄ, 2008), a energia de deformação via TPS (BOOKSTEIN, 1989), e os 2 métodos de estaganálise (CHEN; SHI, 2008; KODOVSKY; FRIDRICH, 2009).

4.4.1 Análise de Variância

Para justificar a utilização das 4 características utilizadas na metodologia desenvolvida foram realizados um projeto de experimento fatorial completo. O projeto experimental possui 3 fatores controláveis. O primeiro é a base de dados escolhida NUAA (TAN *et al.*, 2010) ou CASIA (ZHANG *et al.*, 2012). O segundo fator controlável são as características extraídas sendo elas vindas do descritor LPQ, da energia de deformação das faces e dos dois conjuntos de coeficientes da esteganálise. Já o terceiro fator controlável é o tipo de normalização.

O fator controlável $BASE$ possui 2 níveis (NUAA ou CASIA). Já o fator controlável $FEATURE$ possui 15 níveis (são as 4 feições e todas as suas combinações). Por último, o fator controlável $NORM$ que possui 5 níveis (nenhuma normalização, Max, Min-Max, Z-score, Sub-mean).

Em relação à normalização dos dados, foram avaliados 5 diferentes tipos de normalizações aplicadas às feições extraídas, são elas:

Nenhuma normalização:

$$(z_i^k)_N = z_i^k \tag{44}$$

Normalização max:

$$(z_i^k)_N = \frac{z_i^k}{z_{max}^k} \tag{45}$$

Normalização min-max:

$$(z_i^k)_N = \frac{z_i^k - z_{min}^k}{z_{max}^k - z_{min}^k} \tag{46}$$

Normalização Z-score:

$$(z_i^k)_N = \frac{z_i^k - \mu_{z_i^k}}{\sigma_{z_i^k}} \tag{47}$$

Normalização sub-mean:

$$(z_i^k)_N = z_i^k - \mu_{z_i^k} \tag{48}$$

Esse é um experimento de fatorial completo que gerou 150 tratamentos distintos. Para cada tratamento foram realizadas 100 repetições totalizando 15000 observações para todo o experimento. A variável de resposta é a taxa $HTER$. Esse experimento foi conduzido de maneira aleatorizada e o conjunto treino foi dividido em 3 subconjuntos compreendendo 50% para treino, 25% para validação e 25% para teste. Isso foi realizado em cada observação durante o treinamento do classificador RNA.

Em cada observação, os níveis dos fatores controláveis são selecionados aleatoriamente de acordo com a Figura 36. Não há um padrão visível no resíduo e isso implica que a premissa de independência dos erros foi respeitada. O conjunto de treino da base de dados é subdividido aleatoriamente na proporção descrita. Em seguida uma RNA (ver Seção 3.3.4) é treinada para classificar as faces de acordo com as classes de cada quadro dos vídeo e a etiqueta do conjunto verdade. A RNA itera durante a fase de treinamento a fim de minimizar a taxa de erro $HTER(treino)$. Após o treinamento, o conjunto de teste da base de dados é confrontado com a RNA já treinada e a taxa de erro $HTER(teste)$ é obtida.

Dados este experimento fatorial completo, a análise de variância foi utilizada para testar se há ou não diferenças significativas entre os grupos e desta forma verificar a hipótese $h0$ (MONTGOMERY, 2001), ou seja, analisar se as diferenças entre os tratamentos foram significativos assumindo a distribuição Fisher-Snedecor. Neste caso, $F_{calculado}$ é comparada com os valores da distribuição Ficher-Snedecor F_{tabela}, e se $F_{calculado} > F_{tabela}$ a fonte de variação é significativa. A Tabela 12 mostra que os fatores controláveis $BASE$, $FEATURE$, $NORM$ e as todas as suas interações $BASE * FEATURE$, $BASE * NORM$, $FEATURE * NORM$ e $BASE * FEATURE * NORM$ obtiveram variações significativas rejeitando a hipótese nula $h0$ e desta forma comprovam que há diferenças significativos do método quando altera-se a base de dados. Há variações significativas nas características extraídas que justificam as suas utilizações como feições

Figura 36: Sequência de medidas em função do resíduo. Não há um padrão visível no resíduo implicando que a premissa de independência dos erros não foi violada.

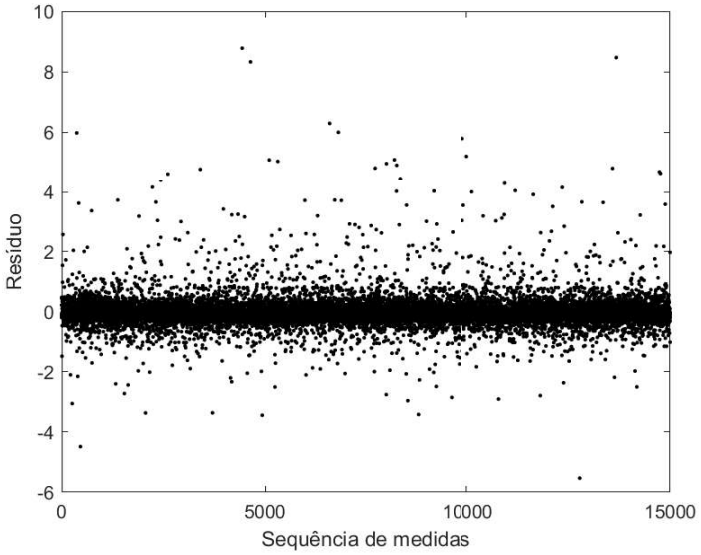


Tabela 12: Análise de Variância do Projeto de Experimento.

Fator Controlável	Soma Qua.	G.L	Média Qua.	$F_{calc.}$	$F_{tab.}$	Conclusão
$BASE$	6188.1	1	6188.1	24709.73	3.84	Significativo
$FEATURE$	1429583.1	14	102113.1	407749.47	1.69	Significativo
$NORM$	34394.4	4	8598.6	34335.26	2.37	Significativo
$BASE * FEATURE$	103420.9	14	7387.2	29497.99	1.69	Significativo
$BASE * NORM$	3823.6	4	955.9	3816.99	2.37	Significativo
$FEATURE * NORM$	49120.7	56	877.2	3502.58	1.33	Significativo
$BASE * FEATURE * NORM$	41604.1	56	742.9	2966.61	1.33	Significativo
Erro	3718.9	14850	0.3	-	-	-
TOTAL	1671853.8	14999	-	-	-	-

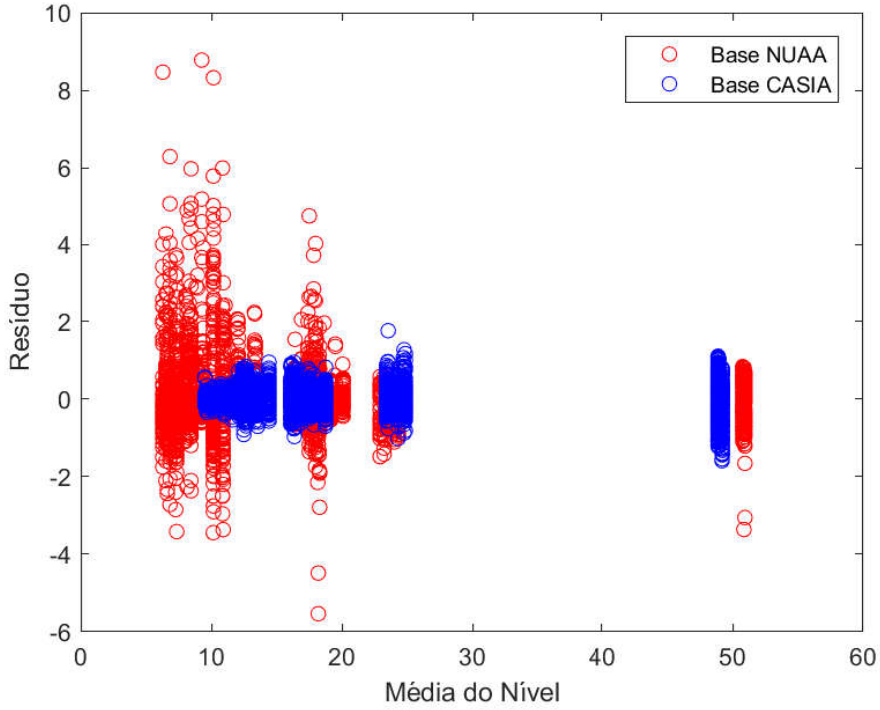
para a detecção de ataques de apresentação. Além disso, há variações significativas nos diferentes tipos de normalizações testadas.

Esses resultados podem ser visualizados através dos gráficos da médio do nível em função do resíduo. A Figura 37 ilustra a variação significativa do fator controlável $Base$ quando os 150 tratamentos são evidenciados de acordo com os níveis NUAA e CASIA. É visível que os melhores resultados foram obtidos para a base NUAA, possivelmente pois essa base contém apenas uma modalidade de ataque.

Quando os tratamentos são evidenciados de acordo com os níveis do fator controlável $FEATURE$, é possível identificar quais combinações de feições extraídas obtiveram as menores taxas de HTER. A Figura 38 ilustra esse caso com os melhores resultados para as feições 11 que representa as características LPQ, energia de deformação e esteganálise1, 13 que representa as características LPQ, esteganálise1 e esteganálise2) e 15 que representa todas as características analisadas LPQ, energia de deformação, esteganálise1 e esteganálise2.

Da mesma forma, quando os tratamentos são evidenciados de acordo com os níveis do fator controlável $NORM$, é possível identificar quais combinações de feições extraídas obtiveram as menores taxas de HTER. Nesse caso as normalizações max, min-max e sub-mean obtiveram os melhores resultados de acordo a Figura 39. Após essas análises as 4

Figura 37: Média do nível em função do resíduo para o fator controlável $Base$. Em vermelho estão os pontos referentes à base NUAA e em vermelho, CASIA.



características tem o seu uso justificados juntamente com a normalização MIN-MAX.

4.4.2 PAD Baseada em Quadros e Vídeos

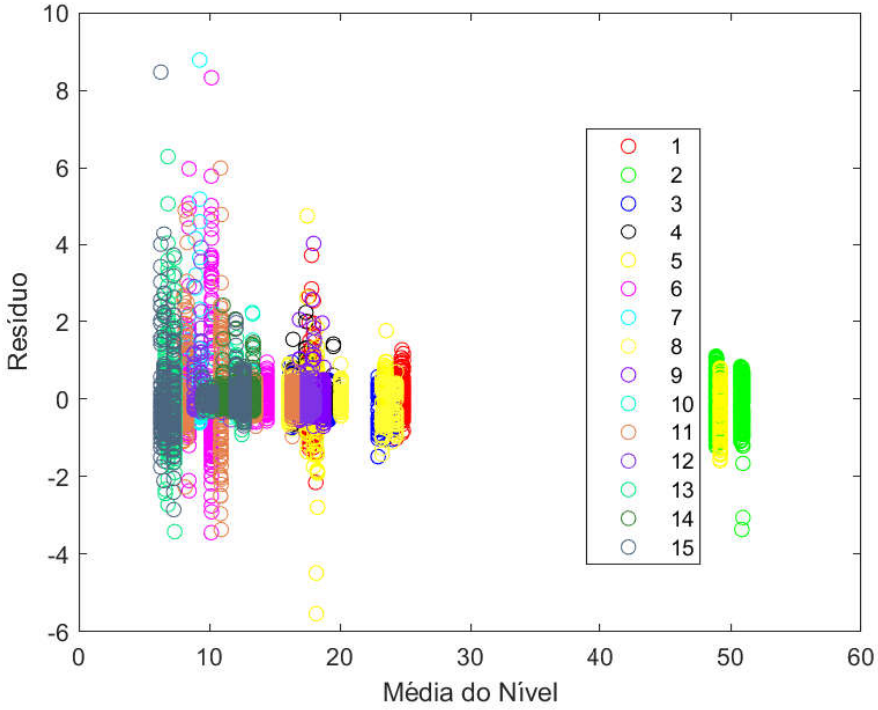
A detecção de ataques de apresentação pode ser informada a cada quadro, após alguns quandros ou ao final de um vídeo. Os autores (ANJOS; MARCEL, 2011) já relatavam melhorias no HTER na medida que as informação dos quadros são acumuladas. Nessa momento é apresentado os resultados para as bases NUAA (TAN *et al.*, 2010) e CASIA (ZHANG *et al.*, 2012) frente ao acúmulo de informação ao longo do tempo.

Melhorias na classificação das faces na medida que acumula informação dos quadros para a base NUAA (TAN *et al.*, 2010). Classificando com apenas 1 quadro alcança uma $CRR = 94.05\%$. Já utilizando a votação de 90 quadros a taxa sobe para $CRR = 98.09\%$ conforme a Figura 40. O mesmo acontece quando analisadas as taxas FRR, FAR e HTER conforme ilustrado na Figura 41 alcaçando uma taxa $HTER = 1.121\%$ quando é utilizado a votação de 90 quadros.

Esse aumento da detecção correta das classes com o acúmulo de informação também ocorre na base de dados CASIA (ZHANG *et al.*, 2012). Conforma a Figura 42 a taxa CRR sobe de $CRR = 95.06\%$ para $CRR = 96.33\%$ quando o acúmulo de informação passa de 1 quadro para 90 quadros. Uma melhoria mais expressiva acontece ao analisar a taxa HTER que cai de $HTER = 6.39\%$ para $HTER = 4.69\%$ nas mesmas condições.

A seguir são apresentados os resultados baseados em vídeo. Isso significa que foi realizada uma votação das etiquetas do classificador RNA ao longo de todos os quadros que compões um vídeo de um único sujeito. Para a base de dados NUAA (TAN *et al.*, 2010) conjunto teste faces genuínas tem-se uma taxa $FRR = 0\%$ ou seja classificou perfeitamente todas os vídeos. Esse resultado está ilustrado na Figura 44 onde embaixo de cada sequência de quadros o número de quadros detectados como genuíno | número

Figura 38: Média do nível em função do resíduo para o fator controlável $FEATURE$. Os melhores resultados são para as feições 11 (LPQ, energia de deformação e esteganálise1), 13(LPQ, esteganálise1 e esteganálise2) e 15 (LPQ, energia de deformação, esteganálise1 e esteganálise2) .



de quadros detectados como impostor. Enquanto que o conjunto teste faces impostoras tem-se uma taxa $FAR = 0\%$ ou seja classificou perfeitamente todas os vídeos quando ataques foram apresentados. Esse resultado está ilustrado na Figura 45 onde da mesma forma embaixo de cada sequência de quadros o número de quadros detectados como genuíno | número de quadros detectados como impostor. Os resultados de PAD baseados em vídeos para ambas as bases de dados estão na Tabela 1.

Tabela 13: PAD baseada em vídeo para ambas as bases

Base	FRR(%)	FAR(%)	HTER(%)
NUAA (TAN *et al.*, 2010)	0	0	0
CASIA (ZHANG *et al.*, 2012)	7.77	3.33	5.55

Figura 39: Média do nível em função do resíduo para o fator controlável $NORM$. Os melhores resultados são para as normalizações max, min-max e sub-mean.

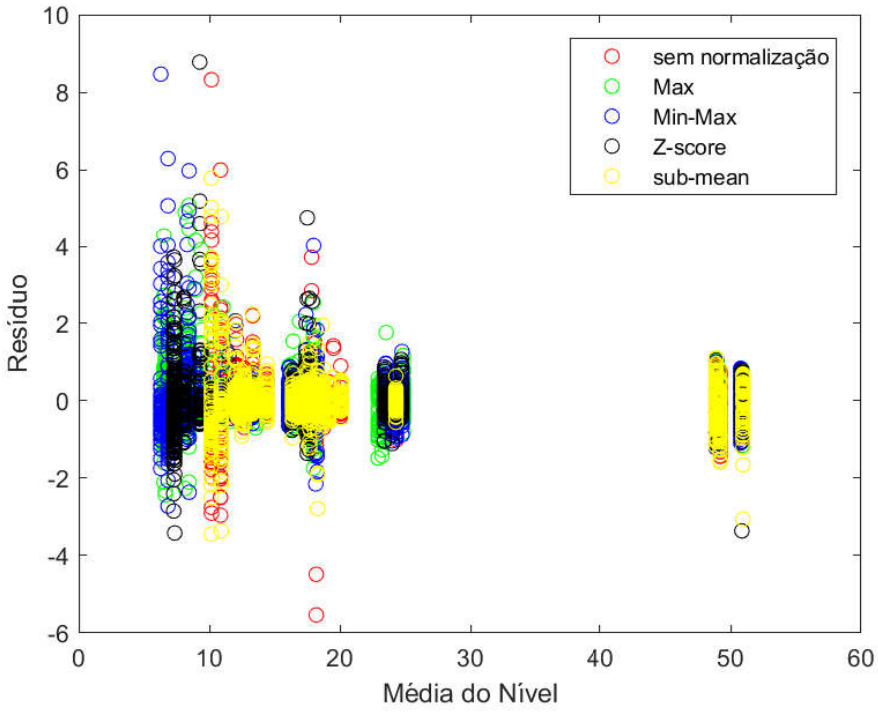


Figura 40: Acúmulo de informação em função da taxa CRR, base NUAA.

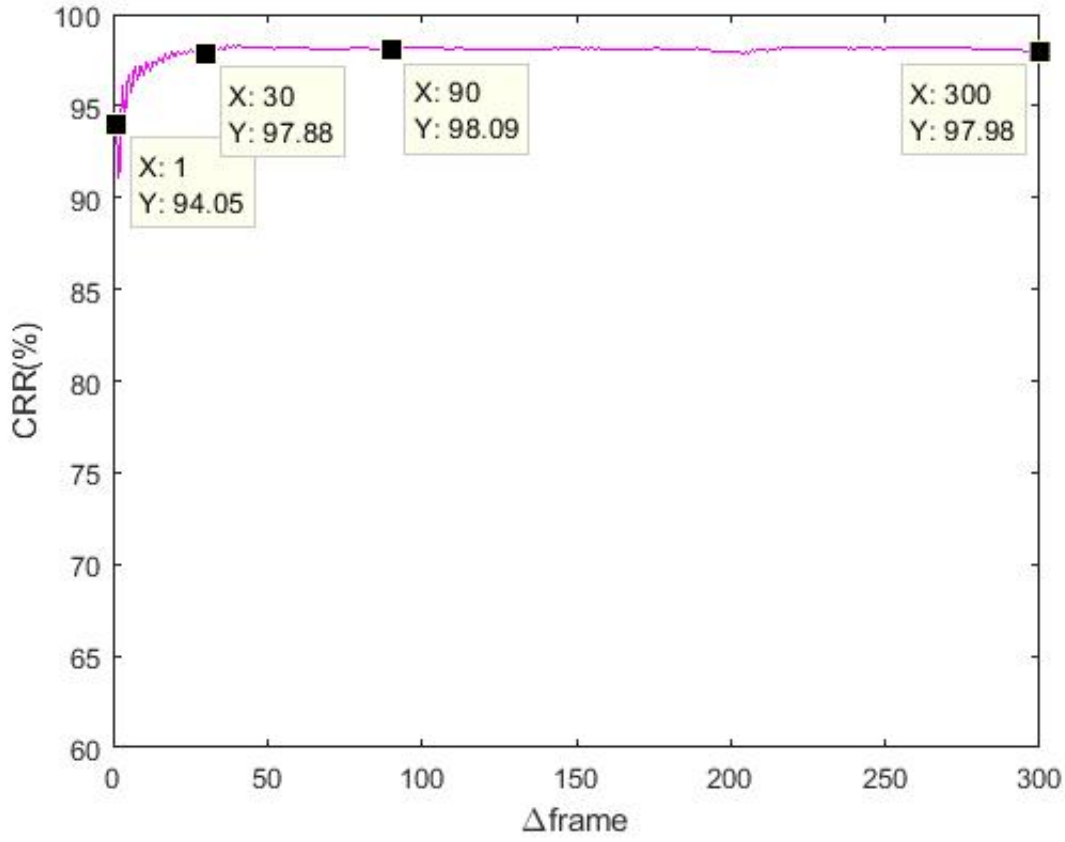


Figura 41: Acúmulo de informação em função da taxas FRR, FAR e HTER, base NUAA.

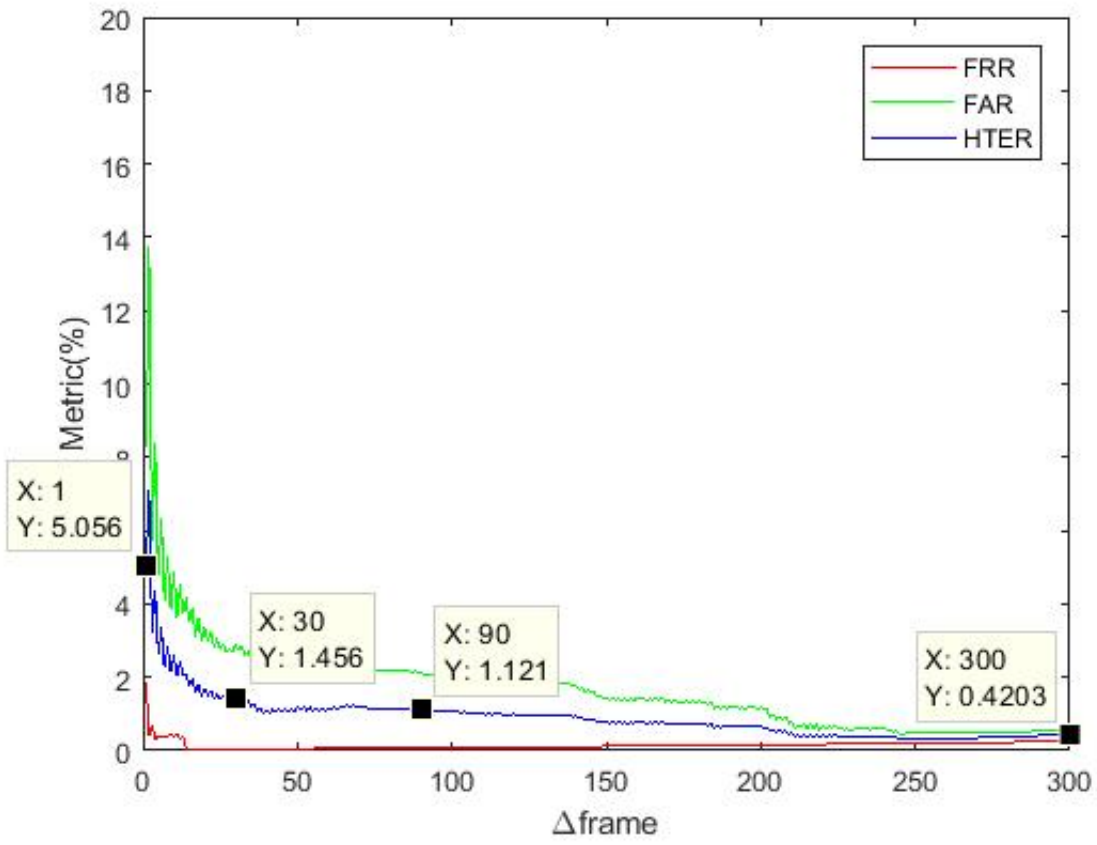


Figura 42: Acúmulo de informação em função da taxa CRR, base CASIA.

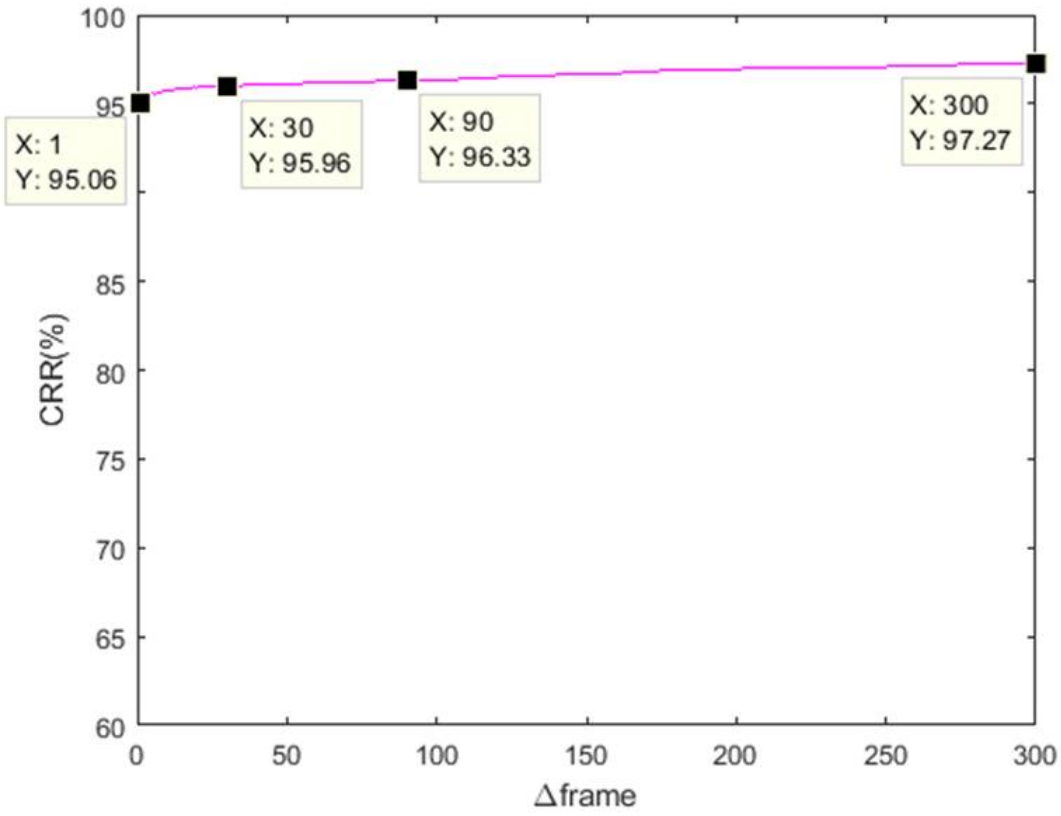


Figura 43: Acúmulo de informação em função da taxas FRR, FAR e HTER, base NUAA.

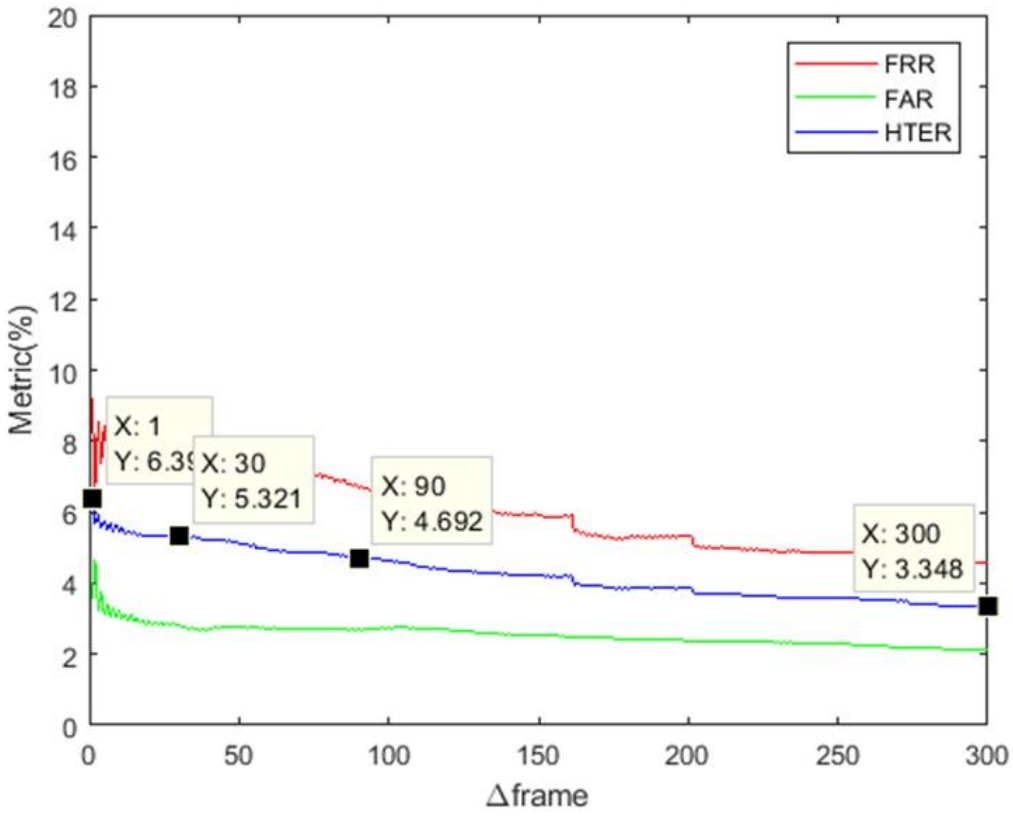


Figura 44: PAD baseada em vídeo no conjunto teste com sujeitos genuínos, base NUAA (TAN *et al.*, 2010).

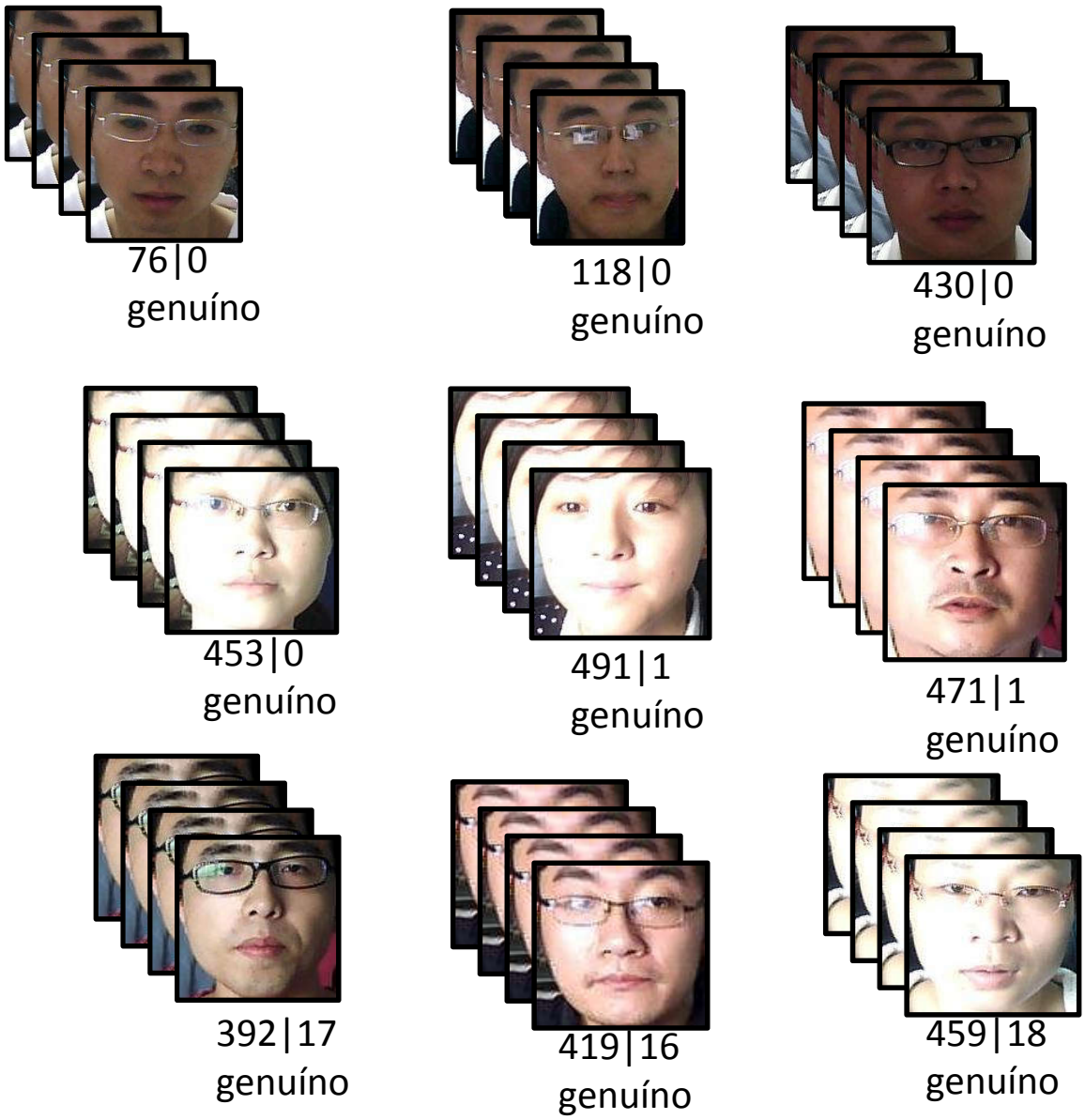
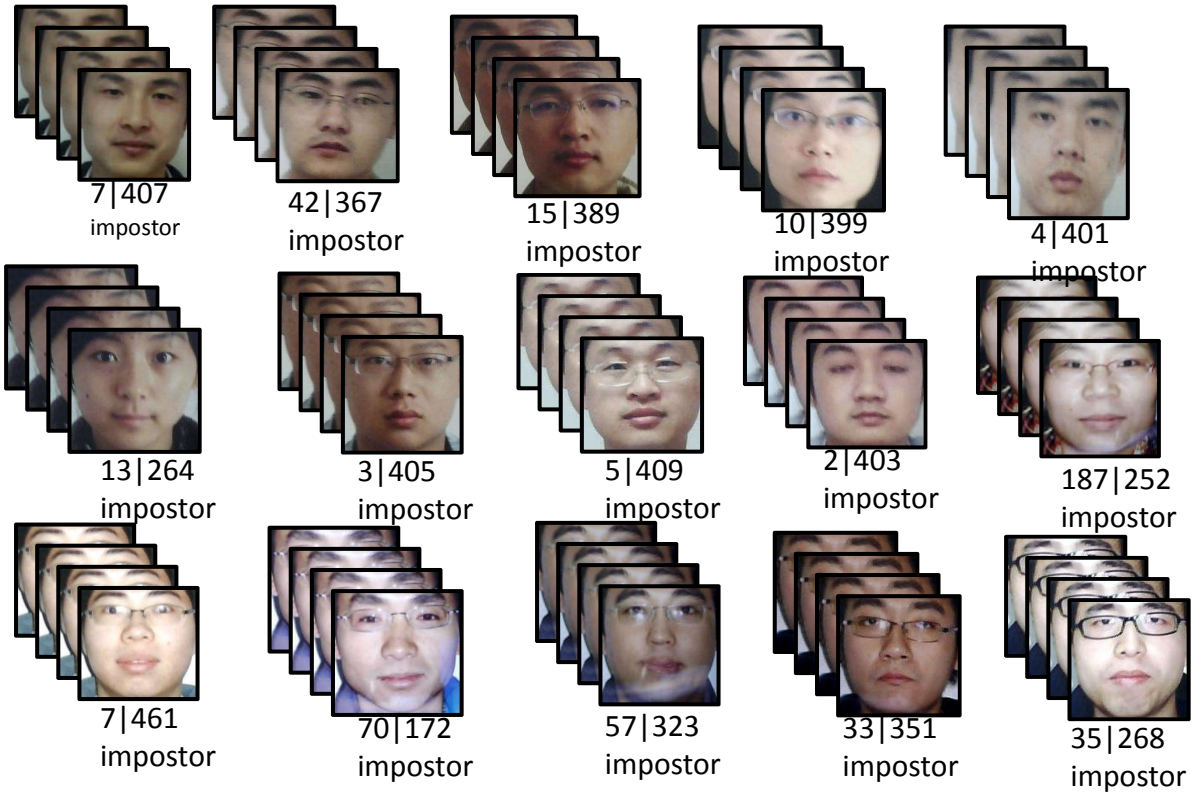


Figura 45: PAD baseada em vídeo no conjunto teste com sujeitos impostores, base NUAA (TAN *et al.*, 2010).



4.4.3 Resultados da Classificação Comparativos

A seguir os resultados experimentais comparativos com o estado da arte para a base de dados NUAA (TAN *et al.*, 2010). Os resultados de outros autores exibidos na Tabela 14 foram retirados diretamente da literatura, o símbolo $-$ significa que o resultado não foi disponibilizado. O método 3 proposto nessa tese alcança resultados competitivos com métodos contemporâneos. Inclusive taxas de CRR melhores que os métodos propostos em (AKBULUT *et al.*, 2017; KOSE; DUGELAY, 2012).

Tabela 14: Comparação com o estado da arte, base NUAA

Método	CCR(%)	FRR(%)	FAR(%)	HTER(%)
CNN (AKBULUT *et al.*, 2017)	76.31	-	-	-
LRF-ELM (AKBULUT *et al.*, 2017)	84.04	-	-	-
Uniform LBPV (KOSE; DUGELAY, 2012)	86.95	11.87	13.75	12.81
All LBPV (KOSE; DUGELAY, 2012)	88.03	9.16	13.61	11.38
Proposto	**94.29**	**1.96**	**7.89**	**4.93**
Mult. LBP(MAATTA; HADID; PIETIKAINEN, 2011)	98.00	4.4	0.6	2.5
Qual. image (LUAN *et al.*, 2017)	98.80	-	-	-

Da mesma forma, os resultados experimentais comparativos com o estado da arte para a base de dados CASIA (ZHANG *et al.*, 2012) são exibidos na Tabela 15. O método 3 proposta nessa tese alcança resultados competitivos com métodos contemporâneos. Inclusive taxas de HTER melhores que os métodos propostos em (ZHANG *et al.*, 2012; FREITAS PEREIRA *et al.*, 2013; WEN; HAN; JAIN, 2015).

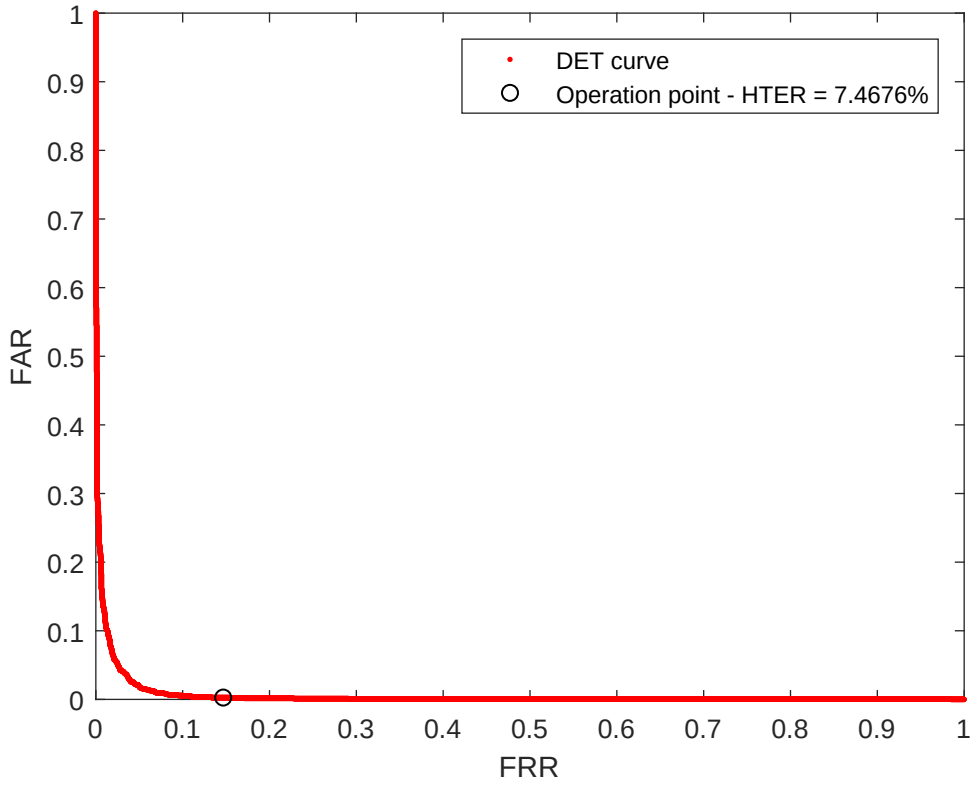
Tabela 15: Comparação com o estado da arte, base CASIA

Método	HTER(%)
DoG *baseline* (ZHANG *et al.*, 2012)	27.55
LBP-TOP u2 (FREITAS PEREIRA *et al.*, 2013)	9.05
DoG+LBP+SVMs *baseline* (WEN; HAN; JAIN, 2015)	7.85
IDA+SVMs (WEN; HAN; JAIN, 2015)	6.70
Proposto	**6.51**
LBP+SVMs *baseline* (WEN; HAN; JAIN, 2015)	3.40
Dicionário (MANJANI *et al.*, 2017)	1.3

4.4.4 Curvas características

As curvas *detection error tradeoff* (DET) é uma plotagem gráfica das taxas de erro para sistemas de classificação binária, plotando a taxa de rejeição falsa em função da taxa de aceitação falsa (FAR) e servem para destacar as diferenças importantes das região operacional. A taxa de erro igual, *Equal Error Rate* (EER) é encontrada para condição $FRR = FAR$. Nessa tese, para o conjunto de treinamento da base NUAA (TAN *et al.*, 2010) o $EER = 0\%$ alcançando o melhor valor possível. Já no conjunto de teste que está representado pela Figura, 46 alcança um $EER = 7.46\%$. Quando ajustado o classificador de acordo a curva DET, os resultados por quadro para a base NUAA são $FRR_{quadro} = 1.04\%$, $FAR_{quadro} = 6.87\%$ e $HTER_{quadro} = 3.95\%$. Enquanto os resultados por vídeo para a base NUAA são $FRR_{pessoa} = 0\%$, $FAR_{pessoa} = 0\%$ e $HTER_{pessoa} = 0\%$.

Figura 46: Curva DET base de dados NUAA, conjunto de treino.



A Figura 47 ilustra a curva DET para o conjunto de treinamento da base CASIA. A EER é encontrada para condição $FRR = FAR$ com o valor $EER = 0.80\%$. Já no conjunto de teste está ilustrado pela Figura 48 com $EER = 9.42\%$. Quando ajustado o classificador de acordo a curva DET, os resultados por quadro para a base CASIA (ZHANG *et al.*, 2012) são $FRR_{quadro} = 14.75\%$, $FAR_{quadro} = 5.52\%$ e $HTER_{quadro} = 10.14\%$. Já os resultados por pessoa para a base CASIA $FRR_{pessoa} = 11.11\%$, $FAR_{pessoa} = 3.70\%$ e $HTER_{pessoa} = 7.40\%$.

Figura 47: Curva DET base de dados CASIA, conjunto de treino.

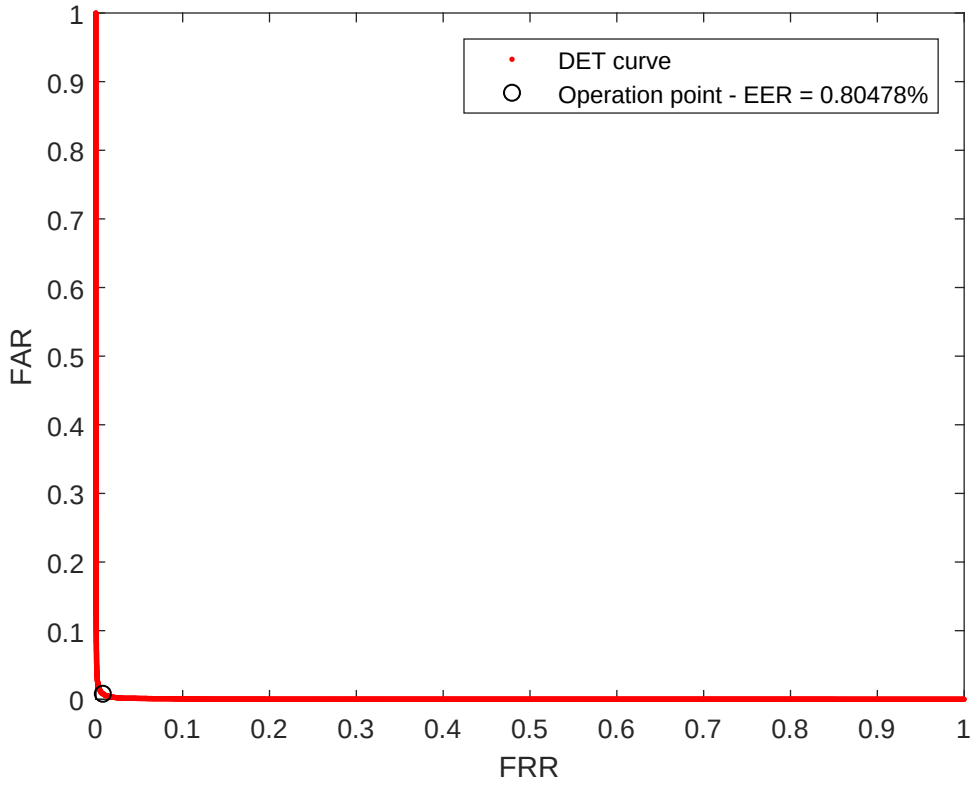
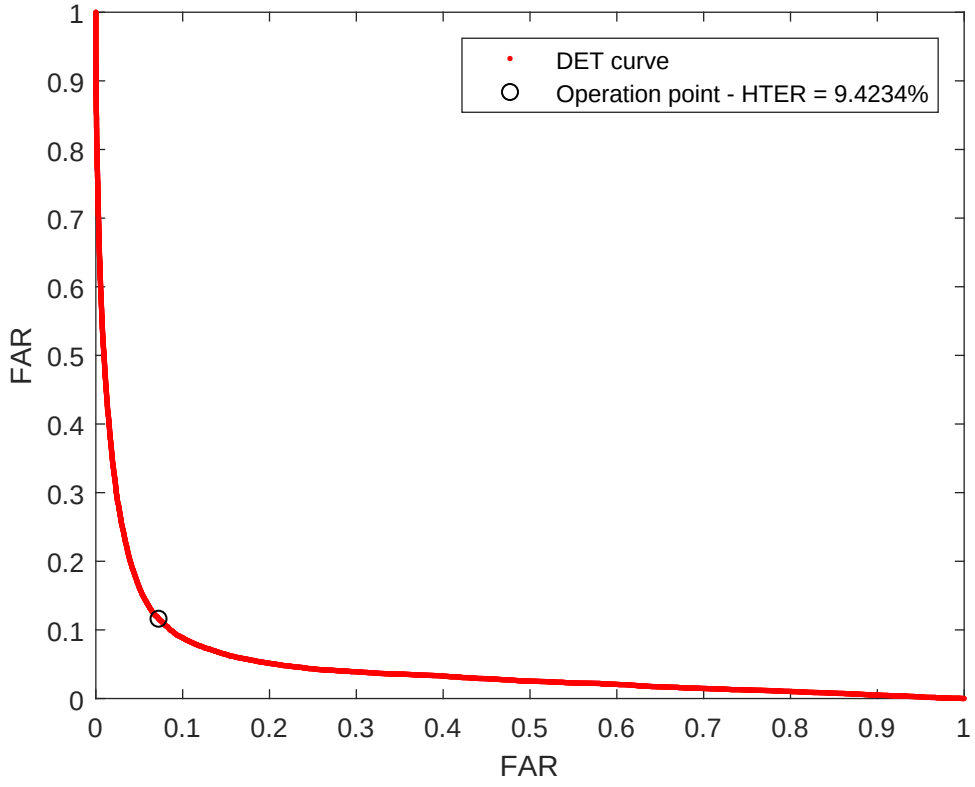


Figura 48: Curva DET base de dados CASIA, conjunto de teste.



5 CONCLUSÕES

Sendo assim, foram propostas 3 novas metodologias capazes de discernir ataques faciais a sistemas biométricos no contexto de ataque de apresentação o que torna possível detectar vivacidade e verificar fraudes a sistemas biométricos quando apresenta-se uma imagem ou um vídeo. Essas abordagens desenvolvidas concretizam os objetivos propostos. Nos métodos desenvolvidos foram utilizadas diversas características capazes separar as classes de faces genuínas das impostoras. Desta forma as perguntas de pesquisas estão respondidas e comprovadas via experimentos. Os dois primeiros métodos acabaram fazendo parte de uma evolução da abordagem PAD de forma que as diferentes abordagens foram testadas e as melhores feições e técnicas foram preservadas e culminaram no método 3.

Em relação as hipóteses levantadas, a metodologia proposta na primeira abordagem atinge uma taxa $HTER = 10.15 \pm 1.03$ com a taxa $PCA = 40\%$ para detectar ataques de apresentação. Este resultado foi alcançado com movimento dos blocos da face juntamente com extrator de textura LRD. Apesar desses resultados, o extrator LRD foi substituído pelo descritores LPQ e coeficientes da esteganálise devido ao resultados insatisfatórios. A segunda abordagem ao problema utiliza LBP, Entropia e as texturas de bordas em faces alinhadas como características extraídas dos blocos das faces. Esta abordagem obteve resultados insatisfatórios alcançando uma taxa de detecção correta de apenas $CDR = 69,8\%$ para a base NUAA. Já a terceira abordagem utiliza o descritor LPQ, a energia de deformação das faces mais duas técnicas de esteganálise extrair informações sobre as imagens apresentadas. Os dados são projetados utilizando PCA-LDA e treinados e testados com uma RNA. Com esta abordagem foi possível atingir resultados competitivos com o estado da arte comprovados experimentalmente. Sendo que os melhores resultados obtidos foram com a projeção PCA-LDA, descritor LPQ, energia de deformação e esteganografia usando um classificador de redes neurais artificiais alcançando uma taxa $HTER = 4.93\%$ e $HTER = 6.51\%$ para as bases de dados NUAA e CASIA respectivamente.

Em relação a trabalhos futuros, é possível desenvolver estratégias novas e inovadoras para o processamento e análise de dados multimodais de biometria. Assim explorar as novas formas de ataque a sistemas biométricos (WILD *et al.*, 2016) que desafiam os detectores de falsificação de última geração, sugerindo abordagens sistemáticas adicionais para combater tentativas de falsificação. Por exemplo, a análise multimodal que incorpora a fusão de dados provenientes da face e da impressão digital podem aumentar a robustez frente a ataques de apresentação.

REFERÊNCIAS

ABIODUN, O. I. *et al.* State-of-the-art in artificial neural network applications: a survey. **Heliyon**, Oxford, v.4, n.11, p.1 – 41, 2018.

ACHLIOPTAS, D. Database-friendly random projections: johnson-lindenstrauss with binary coins. **Journal of Computer and System Sciences**, Oxford, v.66, n.4, p.671 – 687, 2003. Special Issue.

AKBULUT, Y. *et al.* Deep learning based face liveness detection in videos. In: INTERNATIONAL ARTIFICIAL INTELLIGENCE AND DATA PROCESSING SYMPOSIUM (IDAP), 2, 2017, Malatya. **Proceedings...** New York: IEEE, 2017. p.1–4.

AKHTAR, Z. *et al.* Robustness of multi-modal biometric systems under realistic spoof attacks against all traits. In: WORKSHOP ON BIOMETRIC MEASUREMENTS AND SYSTEMS FOR SECURITY AND MEDICAL APPLICATIONS (BIOMS), 2, 2011, Milan. **Proceedings...** New York: IEEE, 2011. p.1–6.

ALBU, R. D. Face anti-spoofing based on Radon transform. In: INTERNATIONAL CONFERENCE ON ENGINEERING OF MODERN ELECTRIC SYSTEMS (EMES), 13, 2015, Oradea. **Proceedings...** New York: IEEE, 2015. p.1–4.

ANJOS, A.; CHAKKA, M. M.; MARCEL, S. Motion-based counter-measures to photo attacks in face recognition. **IET Biometrics**, New York, v.3, n.3, p.147–158, Sept 2014.

ANJOS, A.; MARCEL, S. Counter-measures to photo attacks in face recognition: a public database and a baseline. In: INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS (IJCB), 1, 2011, Washington, DC. **Proceedings...** New York: IEEE, 2011. p.1–7.

AURIA, L.; MORO, R. A. **Support Vector Machines (SVM) as a Technique for Solvency Analysis**. Berlin: DIW Berlin, German Institute for Economic Research, 2008. (Discussion Papers of DIW Berlin 811).

BANHAM, M. R.; KATSAGGELOS, A. K. Digital image restoration. **IEEE Signal Processing Magazine**, New York, v.14, n.2, p.24–41, March 1997.

BAO, W. *et al.* A liveness detection method for face recognition based on optical flow field. In: INTERNATIONAL CONFERENCE ON IMAGE ANALYSIS AND SIGNAL PROCESSING, 16, 2009, Taizhou. **Proceedings...** New York: IEEE, 2009. p.233–236.

BENLAMOUDI, A. *et al.* Face spoofing detection using local binary patterns and Fisher Score. In: INTERNATIONAL CONFERENCE ON CONTROL, ENGINEERING INFORMATION TECHNOLOGY (CEIT), 3, 2015, Tlemcen. **Proceedings...** New York: IEEE, 2015. p.1–5.

BHARADWAJ, S. *et al.* Computationally Efficient Face Spoofing Detection with Motion Magnification. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 29, 2013, Portland. **Proceedings...** New York: IEEE, 2013. p.105–110.

BIGGIO, B. *et al.* Adversarial Biometric Recognition : a review on biometric system security from the adversarial machine-learning perspective. **IEEE Signal Processing Magazine**, New York, v.32, n.5, p.31–41, Sept 2015.

BOOKSTEIN, F. L. Principal warps: thin-plate splines and the decomposition of deformations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v.11, n.6, p.567–585, Jun 1989.

BOULKENAFET, Z.; KOMULAINEN, J.; HADID, A. Face Spoofing Detection Using Colour Texture Analysis. **IEEE Transactions on Information Forensics and Security**, New York, v.11, n.8, p.1818–1830, Aug 2016.

BOULKENAFET, Z. *et al.* OULU-NPU: a mobile face presentation attack database with real-world variations. In: INTERNATIONAL CONFERENCE ON AUTOMATIC FACE GESTURE RECOGNITION (FG 2017), 12, 2017, Washington, DC. **Proceedings...** New York: IEEE, 2017. p.612–618.

BURGOS-ARTIZZU, X. P.; PERONA, P.; DOLLáR, P. Robust Face Landmark Estimation under Occlusion. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 14, 2013, Sydney. **Proceedings...** New York: IEEE, 2013. p.1513–1520.

CAI, D.; HE, X.; HAN, J. SRDA: an efficient algorithm for large-scale discriminant analysis. **IEEE Transactions on Knowledge and Data Engineering**, New York, v.20, n.1, p.1–12, Jan 2008.

CAO, X. *et al.* Face alignment by Explicit Shape Regression. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 28, 2012, Providence. **Proceedings...** New York: IEEE, 2012. p.2887–2894.

CASTRO, J. *et al.* Extraction of fuzzy rules from support vector machines. **Fuzzy Sets and Systems**, Oxford, v.158, n.18, p.2057 – 2077, 2007.

CHAKKA, M. M. *et al.* Competition on counter measures to 2-D facial spoofing attacks. In: INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS (IJCB), 1, 2011, Washington. **Proceedings...** New York: IEEE, 2011. p.1–6.

CHEN, C.; SHI, Y. Q. JPEG image steganalysis utilizing both intrablock and interblock correlations. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, 22, 2008, Seattle. **Proceedings...** New York: IEEE, 2008. p.3029–3032.

CHINGOVSKA, I.; ANJOS, A.; MARCEL, S. On the effectiveness of local binary patterns in face anti-spoofing. In: INTERNATIONAL CONFERENCE OF

BIOMETRICS SPECIAL INTEREST GROUP (BIOSIG), 11, 2012, Darmstadt. **Proceedings...** New York: IEEE, 2012. p.1–7.

COSTA-PAZO, A. *et al.* The Replay-Mobile Face Presentation-Attack Database. In: INTERNATIONAL CONFERENCE OF THE BIOMETRICS SPECIAL INTEREST GROUP (BIOSIG), 15, 2016, Darmstadt. **Proceedings...** New York: IEEE, 2016. p.1–7.

CRISTIANINI, N.; SHAWE-TAYLOR, J. **An Introduction to Support Vector Machines**: and other kernel-based learning methods. New York: Cambridge University Press, 2000.

DASGUPTA, S.; GUPTA, A. An Elementary Proof of a Theorem of Johnson and Lindenstrauss. **Random Struct. Algorithms**, New York, v.22, n.1, p.60–65, Jan. 2003.

DUCHON, J. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. **ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique**, [S.l.], v.10, n.R3, p.5–12, 1976.

FREITAS PEREIRA, T. de *et al.* Can face anti-spoofing countermeasures work in a real world scenario? In: INTERNATIONAL CONFERENCE ON BIOMETRICS (ICB), 6, 2013, Madrid. **Proceedings...** New York: IEEE, 2013. p.1–8.

FREITAS PEREIRA, T. de *et al.* LBP-TOP Based Countermeasure against Face Spoofing Attacks. In: ASIAN CONFERENCE ON COMPUTER VISION WORKSHOPS, 7, 2013, Daejeon. **Proceedings...** Berlin: Springer, 2013. p.121–132.

Front Cover. **IEEE Signal Processing Magazine**, New York, v.32, n.5, p.C1–C1, Sep. 2015.

GALBALLY, J.; MARCEL, S.; FIERREZ, J. Image Quality Assessment for Fake Biometric Detection: application to iris, fingerprint, and face recognition. **IEEE Transactions on Image Processing**, New York, v.23, n.2, p.710–724, Feb 2014.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. **Digital Image Processing Using MATLAB**. Upper Saddle River: Prentice Hall Press, 2007.

HADID, A. *et al.* Biometrics Systems Under Spoofing Attack: an evaluation methodology and lessons learned. **IEEE Signal Processing Magazine**, New York, v.32, n.5, p.20–30, Sept 2015.

HAYKIN, S. **Neural Networks**: a comprehensive foundation. 2.ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

HE, X.; LU, Y.; SHI, P. A Fake Iris Detection Method Based on FFT and Quality Assessment. In: CHINESE CONFERENCE ON PATTERN RECOGNITION, 1, 2008, Beijing. **Proceedings...** New York: IEEE, 2008. p.1–4.

IMBAULT, F.; LEBART, K. A stochastic optimization approach for parameter tuning of support vector machines. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 17, 2004, Cambridge. **Proceedings...** New York: IEEE, 2004. v.4, p.597–600 Vol.4.

JAIN, A. K.; NANDAKUMAR, K.; NAGAR, A. Biometric Template Security. **EURASIP J. Adv. Signal Process**, London, p.113–130, jan 2008.

KAZEMI, V.; SULLIVAN, J. One millisecond face alignment with an ensemble of regression trees. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29, 2014, Columbus. **Proceedings...** New York: IEEE, 2014. p.1867–1874.

KODOVSKY, J.; FRIDRICH, J. Calibration Revisited. In: ACM WORKSHOP ON MULTIMEDIA AND SECURITY, 11, 2009, Princeton. **Proceedings...** New York: ACM, 2009. p.63–74.

KOLLREIDER, K.; FRONTHALER, H.; BIGUN, J. Evaluating liveness by face images and the structure tensor. In: WORKSHOP ON AUTOMATIC IDENTIFICATION ADVANCED TECHNOLOGIES (AUTOID'05), 4, 2005, Buffalo. **Proceedings...** New York: IEEE, 2005. p.75–80.

KOLLREIDER, K.; FRONTHALER, H.; BIGUN, J. Verifying liveness by multiple experts in face biometrics. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 21, 2008, Anchorage. **Proceedings...** New York: IEEE, 2008. p.1–6.

KOLLREIDER, K.; FRONTHALER, H.; BIGUN, J. Non-intrusive liveness detection by face images. **Image and Vision Computing**, Oxford, v.27, n.3, p.233 – 244, 2009. Special Issue on Multimodal Biometrics.

KOLLREIDER, K. *et al.* Real-Time Face Detection and Motion Analysis With Application in Liveness Assessment. **IEEE Transactions on Information Forensics and Security**, New York, v.2, n.3, p.548–558, Sept 2007.

KOSE, N.; DUGELAY, J. L. Classification of captured and recaptured images to detect photograph spoofing. In: INTERNATIONAL CONFERENCE ON INFORMATICS, ELECTRONICS VISION (ICIEV), 1, 2012, Dhaka. **Proceedings...** New York: IEEE, 2012. p.1027–1032.

LAGARIAS, J. C. *et al.* Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. **SIAM J. on Optimization**, Philadelphia, v.9, n.1, p.112–147, May 1998.

LI, J. *et al.* Live face detection based on the analysis of Fourier spectra. **Proc.SPIE**, Bellingham, v.5404, p.5404–5412, 2004.

LI, Y.; TAN, X. An Anti-Photo Spoof Method in Face Recognition Based on the Analysis of Fourier Spectra with Sparse Logistic Regression. In: CHINESE CONFERENCE ON PATTERN RECOGNITION, 2, 2009, Nanjing. **Proceedings...** New York: IEEE, 2009. p.1–5.

LIU, L. *et al.* Sorted Random Projections for robust texture classification. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 13, 2011, Barcelona. **Proceedings...** New York: IEEE, 2011. p.391–398.

LORENA, A. C.; CARVALHO, A. C. P. L. F. de. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, Porto Alegre, v.14, n.2, p.43 – 67, 2007.

LUAN, X. *et al.* Face liveness detection with recaptured feature extraction. In: INTERNATIONAL CONFERENCE ON SECURITY, PATTERN ANALYSIS, AND CYBERNETICS (SPAC), 1, 2017, Shenzhen. **Proceedings...** New York: IEEE, 2017. p.429–432.

MAATTA, J.; HADID, A.; PIETIKAINEN, M. Face spoofing detection from single images using micro-texture analysis. In: INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS (IJCB), 1, 2011, Washington. **Proceedings...** New York: IEEE, 2011. p.1–7.

MAGNA JÚNIOR, J. P. **O uso de Thin-Plate Splines na transformação de coordenadas com modelagem de distorções entre realizações de referenciais geodésicos**. 2012. 117 f. Tese (doutorado) — Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, 2012.

MANJANI, I. *et al.* Detecting Silicone Mask-Based Presentation Attack via Deep Dictionary Learning. **Transactions on Information Forensics and Security**, New York, v.12, n.7, p.1713–1723, July 2017.

MARSICO, M. D. *et al.* Moving face spoofing detection via 3D projective invariants. In: IAPR INTERNATIONAL CONFERENCE ON BIOMETRICS (ICB), 5, 2012, New Delhi. **Proceedings...** New York:IEEE, 2012. p.73–78.

MARTIN, A. F. *et al.* The DET curve in assessment of detection task performance. In: EUROSPEECH, 2, 1997, Rhodes. **Proceedings...** Baixas: ISCA, 1997. p.22–25.

MATLAB Neural Toolbox 11.0. The MathWorks, Natick, MA, USA.

MILBORROW, S.; NICOLLS, F. **STASM Active Shape Models with SIFT Descriptors and MARS**. Disponível em: <http://www.milbo.users.sonic.net/stasm/>. Acesso em: fevereiro 2019.

MILBORROW, S.; NICOLLS, F. Active shape models with SIFT descriptors and MARS. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION THEORY AND APPLICATIONS (VISAPP), 9, 2014, Lisbon. **Proceedings...** New York:IEEE, 2014. v.2, p.380–387.

MøLLER, M. F. A scaled conjugate gradient algorithm for fast supervised learning. **Neural Networks**, Oxford, v.6, n.4, p.525 – 533, 1993.

MONTGOMERY, D. C. **Design and analysis of experiments**. 5th.ed. New York: John Wiley & Sons, 2001.

OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **Transactions on Pattern Analysis and Machine Intelligence**, New York, v.24, n.7, p.971–987, July 2002.

OJANSIVU, V.; HEIKKILÄ, J. Blur Insensitive Texture Classification Using Local Phase Quantization. In: INTERNATIONAL CONFERENCE ON IMAGE AND SIGNAL PROCESSING, 1, 2008, Cherbourg. **Proceedings...** Berlin:Springer, 2008. p.236–243.

PAN, G. *et al.* Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 11, 2007, Rio de Janeiro. **Proceedings...** New York: IEEE, 2007. p.1–8.

PAN, G. *et al.* Monocular camera-based face liveness detection by combining eyeblink and scene context. **Telecommunication Systems**, Berlin, v.47, n.3, p.215–225, Aug 2011.

PEVNY, T.; BAS, P.; FRIDRICH, J. Steganalysis by Subtractive Pixel Adjacency Matrix. **Transactions on Information Forensics and Security**, New York, v.5, n.2, p.215–224, June 2010.

PRODANOV, C. C.; FREITAS, E. **Metodologia do Trabalho Científico**: métodos e técnicas da pesquisa e do trabalho acadêmico - 2ª edição. Novo Hamburgo: Editora Feevale, 2013.

PROVOS, N. Defending Against Statistical Steganalysis. In: CONFERENCE ON USENIX SECURITY SYMPOSIUM, 10, 2001, Washington, D.C. **Proceedings...** Berkeley: Usenix Association, 2001.

RODRIGUES, R. N.; LING, L. L.; GOVINDARAJU, V. Robustness of multimodal biometric fusion methods against spoof attacks. **Journal of Visual Languages and Computing**, Oxford, v.20, n.3, p.169 – 179, 2009.

SALLEE, P. Model-based methods for steganography and steganalysis. **International Journal of Image and Graphics**, Hackensack, v.05, n.01, p.167–189, 2005.

SCHARDOSIM, L. R.; SCHARCANSKI, J. Motion Detection and Compensation of Affine Deformations in Infrared Retinal Videos. **IEEE Transactions on Instrumentation and Measurement**, New York, v.66, n.1, p.33–44, Jan 2017.

SOLDERA, J. *et al.* Facial biometrics and applications. **IEEE Instrumentation Measurement Magazine**, New York, v.20, n.2, p.4–30, April 2017.

SUN, L. *et al.* Blinking-Based Live Face Detection Using Conditional Random Fields. In: ADVANCES IN BIOMETRICS, 2, 2007, Seoul. **Proceedings...** Berlin:Springer, 2007. p.252–260.

SUN, Y.; WANG, X.; TANG, X. Deep Convolutional Network Cascade for Facial Point Detection. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29, 2013, Portland. **Proceedings...** New York: IEEE, 2013. p.3476–3483.

TAN, X. *et al.* Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 11, 2010, Heraklion. **Proceedings...** Berlin: Springer, 2010. p.504–517.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. Berlin, Heidelberg: Springer-Verlag, 1995.

WANG, Y. *et al.* Robust face anti-spoofing with depth information. **Journal of Visual Communication and Image Representation**, Oxford, v.49, p.332 – 337, 2017.

WEN, D.; HAN, H.; JAIN, A. K. Face Spoof Detection With Image Distortion Analysis. **IEEE Transactions on Information Forensics and Security**, New York, v.10, n.4, p.746–761, April 2015.

WESTFELD, A. F5—A Steganographic Algorithm. In: INTERNATIONAL WORKSHOP ON INFORMATION HIDING, 4, 2001, Prague. **Proceedings...** Berlin: Springer, 2001. p.289–302.

WILD, P. *et al.* Robust multimodal face and fingerprint fusion in the presence of spoofing attacks. **Pattern Recognition**, Oxford, v.50, p.17 – 25, 2016.

WRIGHT, J. *et al.* Robust Face Recognition via Sparse Representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v.31, n.2, p.210–227, Feb 2009.

YANG, J. *et al.* Face liveness detection with component dependent descriptor. In: INTERNATIONAL CONFERENCE ON BIOMETRICS (ICB), 6, 2013, Madrid. **Proceedings...** New York:IEEE, 2013. p.1–6.

YANG, J. *et al.* Person-Specific Face Antispoofing With Subject Domain Adaptation. **IEEE Transactions on Information Forensics and Security**, New York, v.10, n.4, p.797–809, April 2015.

ZHANG, Z. *et al.* Face liveness detection by learning multispectral reflectance distributions. In: INTERNATIONAL CONFERENCE ON AUTOMATIC FACE GESTURE RECOGNITION AND WORKSHOPS, 9, 2011, Santa Barbara. **Proceedings...** New York:IEEE, 2011. p.436–441.

ZHANG, Z. *et al.* A face antispoofing database with diverse attacks. In: INTERNATIONAL CONFERENCE ON BIOMETRICS (ICB), 5, 2012, New Delhi. **Proceedings...** New York:IEEE, 2012. p.26–31.

ZHANG, Z. *et al.* **TCDCN face alignment tool**. Disponível em: <http://mmlab.ie.cuhk.edu.hk/projects/TCDCN.html>. Acesso em: fevereiro 2019.

ZHANG, Z. *et al.* Learning and Transferring Multi-task Deep Representation for Face Alignment. **CoRR**, [S.l.], v.abs/1408.3967, 2014.

ZHANG, Z. *et al.* Learning Deep Representation for Face Alignment with Auxiliary Attributes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v.38, n.5, p.918–930, May 2016.

ZHU, X.; RAMANAN, D. Face detection, pose estimation, and landmark localization in the wild. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 28, 2012, Providence. **Proceedings...** New York: IEEE, 2012. p.2879–2886.