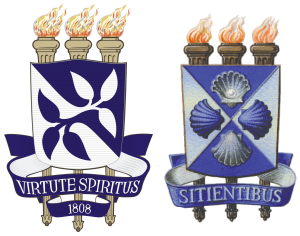

Universidade Federal da Bahia
Universidade Estadual de Feira de Santana

DISSERTAÇÃO DE MESTRADO

Um estudo sistemático sobre detecção de impostor facial

Luiz Otávio de Oliveira Souza Júnior

Mestrado Multiinstitucional em Ciência da Computação
MMCC

Salvador - BA

2016

LUIZ OTÁVIO DE OLIVEIRA SOUZA JÚNIOR

UM ESTUDO SISTEMÁTICO SOBRE DETECÇÃO DE IMPOSTOR FACIAL

Dissertação apresentada ao Mestrado em Ciência da Computação da Universidade Federal da Bahia e Universidade Estadual de Feira de Santana, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Luciano Rebouças de Oliveira

Salvador - BA

2016

S729 Souza Júnior, Luiz Otávio de Oliveira.
Um estudo sistemático sobre detecção de impostor facial / Luiz Otávio de Oliveira Souza Júnior. – Salvador, 2016.
96 f. : il. color.

Orientador: Prof. Dr. Luciano Rebouças de Oliveira.

Dissertação (Mestrado) – Universidade Federal da Bahia. Instituto de Matemática, 2016.

1. Visão por computador. 2. Sistemas de reconhecimento de padrões. 3. Biometria. 4. Processamento de imagens. I. Oliveira, Luciano Rebouças de. II. Universidade Federal da Bahia. Instituto de Matemática. III. Título.

CDD: 006.37

LUIZ OTÁVIO DE OLIVEIRA SOUZA JUNIOR

UM ESTUDO SISTEMÁTICO SOBRE DETECÇÃO DE IMPOSTOR FACIAL

Esta Dissertação foi julgada adequada à obtenção do título de Mestre em Ciência da Computação e aprovada em sua forma final pelo Programa Multi-institucional de Pós-Graduação em Ciência da Computação da UFBA-UEFS.

Salvador, 07 de janeiro de 2016.

Prof. Dr. Luciano Rebouças de Oliveira
Examinador Interno

Prof. Dr. Ricardo Araújo Rios
Examinador Interno

Prof. Dr. Eduardo Manuel de Freitas Jorge
Examinador Externo

Dedico esta dissertação à minha família, que me apoiou nessa jornada, aos colegas do Ivisionlab com os quais passei muitas horas do mestrado e aos professores que me orientaram durante o curso.

AGRADECIMENTOS

Agradeço ao meu orientador professor Dr. Luciano Rebouças de Oliveira pelo apoio na pesquisa, pelo comprometimento com o projeto que envolveu a pesquisa e a elaboração de um artigo em um *journal* internacional relacionado com esta dissertação, bem como pelas várias horas de trabalho contínuo e desgastante, porém com resultados relevantes para a análise e submissão do artigo.

Ao professor Dr. Maurício Pamplona Segundo pelo apoio na construção do artigo durante o projeto de pesquisa.

À minha família, que sempre me serviram de inspiração, meus pais, Luiz Otávio de Oliveira Souza (*in memoriam*) e Anaíta Monteiro Souza, à minha irmã, Luana Monteiro Souza, e à minha namorada, Patrícia Júlia Brito Pereira, que sempre estiveram do meu lado durante toda essa jornada. Obrigado pela confiança, sabedoria e perseverança.

Aos meus colegas de turma, com os quais cursei as disciplinas do mestrado e pelos ensinamentos de cada um nas salas de aula, contribuindo para a experiência tanto educacionais, profissionais, como grandes valores de amizade.

Aos amigos do laboratório de Visão Computacional e aos colegas de outras Universidades, que estiveram presente na minha vida durante a realização deste mestrado, oferecendo incentivo e apreciação. Obrigado pelo apoio nessa caminhada.

Agradeço a todas as pessoas que, de uma forma ou de outra, contribuíram para a conclusão desta pesquisa.

"A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo."

—ALBERT EINSTEIN

RESUMO

A face é uma das características humanas mais exploradas em sistemas automáticos de identificação, podendo ser usada tanto em controle de acesso em sistemas de segurança, quanto para identificação de suspeitos, apenas para citar alguns exemplos. Principalmente, em sistemas biométricos de controle de acesso, a identificação facial de um sujeito pode ser burlada a partir de fotos, vídeos ou máscaras, permitindo que impostores tenham acesso a tais sistemas. Estes ataques ocorrem por meio de captura de uma imagem de face genuína, gravação de movimentos dos olhos ou boca de um usuário com direito de acesso ou mesmo máscaras confeccionadas especialmente para imitar um usuário genuíno do sistema. Por conta disso, métodos de detecção de impostor facial passam a ser imprescindíveis para auxiliar sistemas de reconhecimento facial, a fim de distinguir uma imagem falsa (imagem da imagem ou imagem de um vídeo) de uma face real (imagem de um usuário). No presente trabalho, é proposto um estudo sistemático sobre os sistemas de detecção de impostor facial publicados na literatura da área de Reconhecimento de Padrões em Imagens, nos últimos 8 anos. O estudo inicia com uma taxonomia dos trabalhos mais relevantes e uma evolução temporal destes; em seguida, analisa comparativamente os resultados dos trabalhos da literatura sobre as bases de dados mais utilizadas para avaliações comparativas, buscando esclarecer a relevância dos resultados encontrados sob o ponto de vista das métricas utilizadas; por fim, após analisar as características dos métodos criados no passado, propõe perspectivas futuras no que se refere a sistemas mais robustos e que possam ser avaliados a partir de bases de dados mais complexas, bem como métricas menos enviesadas para avaliação dos resultados. O estudo discute algumas questões abertas sobre o tema, visando contribuir para a construção de sistemas aplicáveis no mundo real.

Palavras-chave: Ataques, Impostores, Sistemas de reconhecimento facial, Detecção de impostor facial.

ABSTRACT

Face is one of the most exploited human characteristics in automatic identification systems, being used both in access control security systems, or for identifying suspects, just to cite a few examples. Mainly in biometric access control systems, facial identification of a subject may be circumvented from photos, videos or masks, allowing for impostors to access such systems. These attacks occur by capturing an image of genuine face, recording eye or mouth movements of a user granted to access, or even masks made especially to mimic a genuine user of the system. Because of that, face spoofing detection methods become essential to support face recognition systems in order to distinguish a false image (image of an image or image of a video) from a real one (image of a user). In the present study, we propose a systematic study of face imitation detection systems reported in the literature of Images Pattern Recognition field in the last eight years. The study begins with a taxonomy of the most important work and a temporal progress of these; then a comparative analysis of the results of the studies in the literature over the most commonly used databases for benchmarking, seeking to clarify the significance of the findings from the point of view of the used metrics; finally, after analyzing the characteristics of the methods developed in the past, we propose future perspectives regarding more robust systems, evaluated over more complex databases, as well as less biased metrics to assess the results. The study discusses some open questions on the topic, to contribute to the construction of systems applicable in the real world.

Keywords: Attacks, Impostors, Face recognition system, Face spoofing detection.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

ABREVIAÇÕES

1D-DHMM	**1D-D**iscrete **H**idden **M**arkov **M**odel
1D-FFT	**1D-F**ast **F**ourier **T**ransform
2D-DFT	**2D-D**iscrete **F**ourier **T**ransform
2D-FFT	**2D-F**ast **F**ourier **T**ransform
3DMAD	**3D** **M**ask **A**ttack **D**atabase
ACC	**ACC**uracy
AUC	**A**rea **U**nder **C**urve
CF	**C**olor **F**requency
CLM	**C**onstrained **L**ocal **M**odels
CNN	**C**onvolutional **N**eural **N**etwork
CRF	**C**onditional **R**andom **F**ields
DCT	**D**iscrete **C**osine **T**ransform
DMD	**D**ynamic **M**ode **D**ecomposition
DoG	**D**ifference of **G**aussians
EER	**E**qual **E**rror **R**ate
EGM	**E**lastic **G**raph **M**atching
EPSC	**E**xpected **P**erformance and **S**oofability **C**urves
FAR	**F**alse **A**cceptance **R**ate
FRR	**F**alse **R**ejection **R**ate

GEGMG	**G**eneralized **E**lastic **G**raph **M**atching
GLCM	**G**ray **L**evel **C**o-occurrence **M**atrices
GMM	**G**aussian **M**ixture **M**odels
HMM	**H**idden **M**arkov **M**odels
HOF	**H**istograms of **M**agnitudes of **O**ptical **F**low
HOG	**H**istograms of **O**riented **G**radient
HOOF	**H**istograms of **O**riented **O**ptical **F**low
HSC	**H**istograms of **S**hearlet **C**oefficients
HTER	**H**alf **T**otal **E**rror **R**ate
IDA	**I**mage **D**istortion **A**nalysis
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers
IQM	**I**mage **Q**uality **M**easures
LBP	**L**ocal **B**inary **P**attern
LBP-TOP	**L**ocal **B**inary **P**attern from **T**hree **O**rthogonal **P**lanes
LBPV	**L**ocal **B**inary **P**attern **V**ariance
LDA	**L**inear **D**iscrimant **A**nalysis
LGS	**L**ocal **G**raph **S**tructure
LLR	**L**inear **L**ogistic **R**egression
LPQ	**L**ocal **P**hase **Q**uantization
LR	**L**ogistic **R**egression
MEGM	**M**orphological **E**lastic **G**raph **M**atching
MLP	**M**ulti**L**ayer **P**erceptron

NFA	**N**umber **F**alse **A**cceptance
NFR	**N**umber **F**alse **R**ejection
OFL	**O**ptical **F**low of **L**ines
P2D-HMM	**P**seudo **2D**-**H**idden **M**arkov **M**odel
PCA	**P**rincipal **C**omponent **Analysis**
PDBNN	**P**robabilistic **D**ecision **B**ased **N**eural **N**etwork
PLS	**P**artial **L**east **S**quares
RASL	**R**obust **A**lignment **S**parse and **L**ow rank decomposition
RFI	**R**econhecimento **F**acial em **I**magens
RGB	**R**ed **G**reen **B**lue
RNA	**R**edes **N**eurais **A**rtificiais
ROC	**R**eceiver **O**perating **C**haracteristics
SLR	**S**parse **L**ogistic **R**egression
SLRBLR	**S**parse **L**ow **R**ank **B**ilinear **L**ogistic **R**egression
SOM	**S**elf **O**rganized **M**ap
SVM	**S**upport **V**ector **M**achines
TKL	**T**ransformada de **K**arhunen-**L**oeve

Capítulo 1

INTRODUÇÃO

Com a introdução de novos tipos de sistemas de segurança baseados em senhas alfanuméricas e gráficas para controle de acesso, é observável, constantemente, a ocorrência de violações e fraudes, seja por perda da senha secreta, seja por quebra desta senha (GURAV et al., 2014), (LI et al., 2014), (UDDIN et al., 2014), (ANWAR; IMRAN, 2015). Pode-se definir tais sistemas como não-biométricos. Por outro lado, sistemas que utilizam características biométricas dos seres humanos passam a ser pervasivos, atualmente. Tais sistemas são denominados de biométricos, e, ao considerar características pessoais e únicas, espera-se inibir as vulnerabilidades inerentes a sistemas não-biométricos (MEADOWCROFT, 2008).

Dentre as várias características humanas que podem ser utilizadas para acesso a sistemas biométricos, a face vem sendo crescentemente explorada à medida que métodos de reconhecimento facial a partir de imagens (RFI) tornam-se mais robustos. Ainda que aparentemente mais confiáveis, sistemas do tipo RFI ainda podem sofrer violações devido a invasões de impostores os quais utilizam imagens, vídeos ou máscaras que imitam usuários genuínos do sistema (TAN et al., 2010), (ZHANG et al., 2012), (ERDOGMUS; MARCEL, 2013). Em vista disso, a tarefa de detecção de impostores é gradativamente mais explorada na comunidade científica e indústria com o objetivo de distinguir uma face de um impostor (imagem da imagem ou imagem de um vídeo ou máscara) de uma face genuína (imagem de um usuário do sistema).

Sistemas baseados em RFI podem ser aplicados a uma ampla variedade de situações (DUC; MINH, 2009), (JAIN; LI, 2005), (IBRAHIM; ZIN, 2011), desde sistemas que requerem baixa segurança (por exemplo, mídias sociais e *smartphone*) até aplicações de alta segurança (como, controle de fronteira e vigilância de vídeo), conforme ilustrado na Fig. 1.1. Um sistema RFI tem geralmente os seguintes módulos: (i) localização da face; (ii) normalização da imagem; (iii) extração das características e (iv) associação das características entre uma imagem de face a ser verificada e uma imagem de face registrada em uma base de dados base de dados (mais detalhes no Capítulo 2). Atualmente, é essencial que sistemas baseados em RFI contenham também um módulo para detecção de impostor facial (mais detalhes no Capítulo 3), que usualmente segue o seguinte *pipeline* de

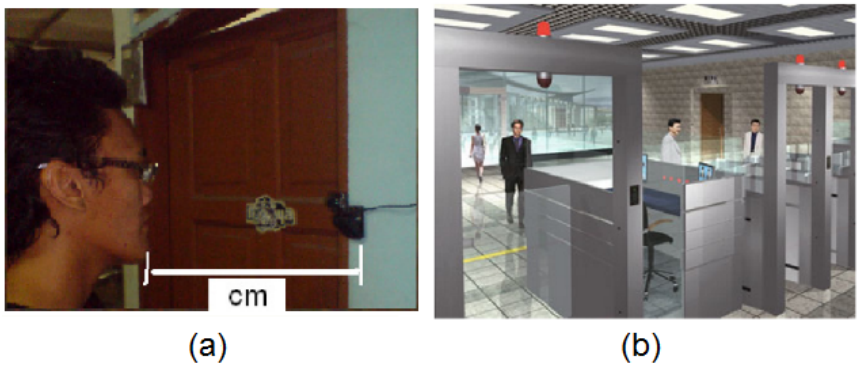

Figura 1.1 Exemplo de dois cenário utilizando sistema biométrico facial: (a) sistema de controle de acesso em um ambiente restrito, com a distância da face a ser verificada e do sensor de captura entre 40 e 60 centímetros e (b) controle de fronteira. Imagens retiradas de (IBRAHIM; ZIN, 2011) e (JAIN; LI, 2005), respectivamente.

informações: (a) reconhecimento facial; (b) descritores e (c) classificadores para detectar se a imagem-alvo é autêntica ou não, conforme ilustrado na Fig.1.2. Ao seguir este *pipeline*, diversas abordagens de detecção de impostor facial foram propostas nos últimos oito anos. O objetivo do presente estudo é, portanto, realizar uma revisão e análise dos trabalhos de detecção de falsificação de faces mais relevantes na literatura, no sentido de compreender o progresso desse campo de pesquisa, apontar a evolução cronológica dos métodos e técnicas propostas até então, indicar tendências dos métodos já publicados e, por fim, discutir perspectivas futuras para a melhoria dos sistemas de detecção de impostores faciais.

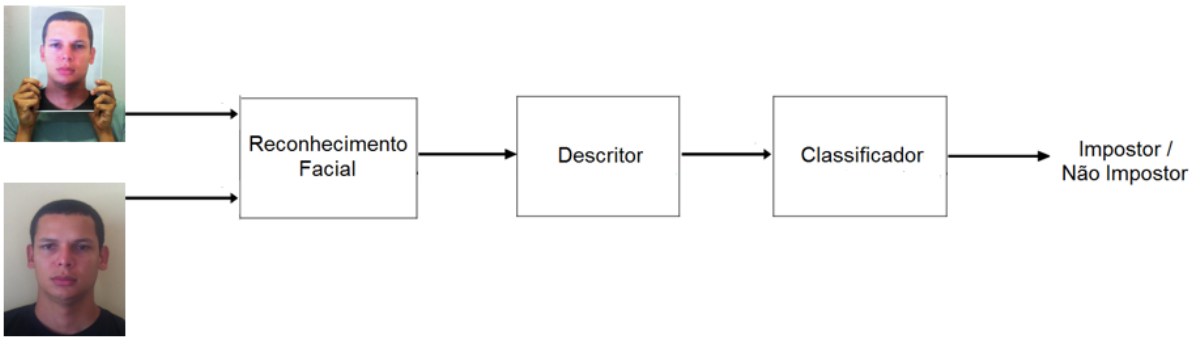

Figura 1.2 *Pipeline* geral de uma detecção de impostor facial. Seguindo as setas da esquerda para a direita: a imagem de face real e falsa passa pelo sistema de reconhecimento facial, a fim de obter a correspondência dessas imagens de entrada com outras imagens de face previamente cadastradas na base de dados; em seguida, as características dessas imagens são extraídas por meios de descritores e analisadas através de classificadores; após a classificação, o resultado de cada imagem de face processada pode ser uma das duas opções: impostor ou não impostor.

1.1 MOTIVAÇÃO

A Tabela 1.1 apresenta um sumário dos estudos sistemáticos existentes na área de detecção de impostores faciais por imagem. Foram levantadas seis propriedades relevantes para caracterizar os estudos, incluindo o nosso que se encontra na última linha da tabela. Só foram encontrados três trabalhos de sistematização sobre o tema (CHAKRABORTY; DAS, 2014), (GALBALLY; MARCEL; FIERREZ, 2014), (PARVEEN et al., 2015).

Chakraborty e Das (2014) categorizam os trabalhos de acordo com a abordagem para detectar impostores. As abordagens são divididas em conformidade com a análise feita sobre o tipo de ataque: baseada em frequência e textura, baseada em focagem variável, baseada em movimento dos olhos, baseada em fluxo ótico, baseada em piscar dos olhos, baseada em descritores dependente de componentes, baseada em forma 3D, baseada em classificação binária, baseada em dicas da cena, baseada em movimento dos lábios, baseada em contexto e, por fim, combinação de técnicas. Após a taxinomização dos métodos, os autores discutem as vantagens de cada técnica em relação aos indicadores de vivacidade (*liveness*). O trabalho em (CHAKRABORTY; DAS, 2014) destaca isoladamente os resultados dos métodos em cada abordagem e não apresenta um estudo aprofundado sobre as bases de dados para avaliação dos métodos, nem uma discussão pormenorizada sobre tendências e questões abertas sobre a área de detecção de impostores.

Galbally et al. (2014) propuseram uma pesquisa contendo uma evolução cronológica de falsificações biométricas multimodais. A cronologia é baseada nas características físicas de impressão digital, face e íris. Os autores categorizaram os trabalhos obtidos em três grupos de técnicas de acordo com a etapa no *pipeline*: em nível de sensor, em nível de características e em nível de fusão entre métodos utilizados sobre diferentes características biométricas. Os trabalhos são também agrupadas pelos tipos de ataque e base de dados utilizados. Uma discussão dos resultados dos métodos obtidos em seis bases de dados públicas disponíveis também é apresentada de modo a caracterizar qualitativamente cada uma das bases e métodos analisados. A análise dos resultados dos métodos existentes foi

Tabela 1.1 Avaliação dos estudos mais relevantes na literatura sobre detecção de impostor facial

Referência	**Taxonomia dos métodos**	**Descrição dos ataques**	**Evolução cronológica**	**Análise das bases de dados**	**Comparação entre os métodos**	**Tendências e perspectivas**
(CHAKRABORTY; DAS, 2014)	√					
(GALBALLY; MARCEL; FIERREZ, 2014)		√	√	√	√	√
(PARVEEN et al., 2015)	√	√		√	√	√
Nosso estudo	√	√	√	√	√	√

realizada somente com base na taxa de erro. Ao final, os autores mencionam um resumo e uma discussão em relação as lições, fatos e desafios dos métodos de detecção multimodal de impostores, utilizando impressão digital e face, bem como métricas e cenários de falsificação biométrica. Não há nenhuma análise sobre perspectivas futuras de métodos de detecção de impostores usando face. Pode-se notar, portanto, que o trabalho não se refere puramente à detecção de impostores por face.

No estudo sobre métodos de detecção de impostores faciais, Parveen et al. (2015) apresenta uma arquitetura geral, composta de sensor, pré-processamento/extração das características e classificação, que funciona como base para a taxinomização dos métodos de detecção de impostores faciais. Parveen et al. analisam os métodos (características + classificação) levantados no estudo a partir dos indicadores de vivacidade, sendo determinados quatro tipos: análise de movimento, detecção de sinais de vida, análise de textura e sensor térmico. Cinco bases são levantadas para comparar os resultados dos métodos existentes. A análise dos resultados das abordagens existentes foi realizada com base nas taxas de erro, tais como: *half total error rate* (HTER) e *equal error rate* (EER) e, inclusive das taxas de precisão *area under curve* (AUC) e *accuracy* (ACC). Por fim, os autores apresentam uma classificação dos sistemas de detecção de falsificação de acordo com três níveis de custo de sistemas: baixo, médio e alto. Esses níveis estão relacionados aos três indicadores de vivacidade, tais como: textura, sinais de vida + movimento e sinais de vida + dispositivo de sensor adicional. Por fim, são apontadas as vantagens e desvantagens dos sistemas levantados, como a complexidade de implementação, se o usuário colabora ou não para a detecção, e a reprodução dos tipos de ataques. O trabalho em (PARVEEN et al., 2015) não apresenta uma taxonomia e descrição aprofundada dos trabalhos coletados e não aborda uma evolução cronológica em sistemas de detecção de impostor facial.

Em nosso estudo, os métodos são categorizados a partir do pipeline geral de métodos para detecção de objetos. Essa abordagem deve favorecer não só uma taxonomia mais concisa, mas também a pesquisas em outras áreas que podem facilmente transferir conhecimento dos trabalhos aqui apresentados para detecção de impostores. Para uma discussão abrangente, optou-se também por descrever os tipos de ataques, buscando uma conexão entre cada tipo e as categorias dos métodos (descritores e classificadores). Diferentemente do trabalho em (CHAKRABORTY; DAS, 2014), a presente revisão de literatura, apresenta uma visão geral da cronologia e evolução temporal dos métodos, buscando determinar possíveis soluções mais robustas. As bases de dados são mostradas e avaliadas especificamente para biometria facial, ao contrário do estudo encontrado em (PARVEEN et al., 2015) que discute várias características físicas do indivíduo. Uma comparação entre os métodos foi realizada para todas as sete bases de dados usualmente utilizadas pelos pesquisadores da área, e uma discussão analítica foi feita, considerando o viés de cada métrica e os resultados (muitas vezes perfeitos) obtidos pelos métodos, em contraposição às análises de resultados dos trabalhos em (GALBALLY; MARCEL; FIERREZ, 2014) e (PARVEEN et al., 2015), onde os resultados são analisados isoladamente. Nenhum dos três estudos levantados discute tendências sobre os métodos propostos, nem perspectivas futuras para construção e avaliação dos métodos de forma mais robusta, assim como é realizado em nosso estudo.

O trabalho foi motivado, portanto, à construção de uma análise abrangente, comparativa e crítica sobre o tema, visando principalmente completar uma lacuna de discussões abertas para a transferência de tecnologia entre os métodos propostos na academia e sua aplicação na indústria.

1.2 OBJETIVOS

Apesar da evolução em pesquisas sobre os sistemas de reconhecimento facial e ataques de falsificações de faces, muito há o que se aperfeiçoar com relação a métodos para impedir o acesso de intrusos em tais sistemas. O objetivo deste trabalho é, portanto, conceber um estudo sistemático sobre os métodos existentes de detecção de impostor facial a fim de definir o cenário atual e propor melhorias futuras.

Dentre os objetivos específicos da pesquisa, cabe mencionar as seguintes metas:

- Abordar as técnicas aplicadas na tarefa de detecção de impostor;
- Expor as possibilidades de ataques de faces;
- Categorizar os métodos de estado-da-arte encontrados na literatura existentes na detecção de falsificação de faces;
- Apresentar uma linha do tempo dos trabalhos científicos mais relevantes na literatura;
- Realizar uma análise comparativa das bases de dados utilizadas na literatura;
- Discutir as tendências e perspectivas futuras da área de detecção de impostor.

1.3 CONTRIBUIÇÕES

As principais contribuições dessa pesquisa são:

- Uma taxonomia com base nos descritores e classificadores utilizados em cada pesquisa expressa na literatura selecionada;
- A evolução dos métodos propostos existentes na literatura nos últimos 8 anos;
- Uma discussão sobre o viés das métricas utilizadas para avaliar os sistemas de detecção de impostor; e
- Uma abordagem das tendências atuais e perspectivas futuras da área de detecção de impostor.

Um estudo sistemático foi submetido ao periódico internacional, *Pattern Recognition*[1], com o seguinte título: *A comprehensive review on face spoofing detection: trends, open issues and perspectives.*

[1] http://www.journals.elsevier.com/pattern-recognition/

1.4 METODOLOGIA

Este estudo baseou-se em uma busca de artigos científicos nas seguintes bases de dados: Scopus[2], IEEE[3] (*Institute of Electrical and Electronics Engineers*), *Engineering Village*[4] e Portal de Periódicos CAPES[5]. Nessas bases foram consultados os artigos contendo todos os anos de publicação com as seguintes palavras-chave: *face recognition, face spoofing detection, face liveness detection, countermeasure against face spoofing attacks* e *face anti-spoofing*. A partir da consulta dos artigos, percebeu-se que os textos sobre o tema começaram a ser publicados em 2007, indo até os dias atuais. A escolha dos artigos foi realizada de acordo com os seguintes critérios: (i) deve seguir o mesmo protocolo da base de dados; (ii) deve mencionar seus resultados usando pelo menos uma das métricas discutidas na Seção 4.1 (mais detalhes no Capítulo 4) e (iii) deve ser comparável a outros trabalhos usando a mesma base de dados.

1.5 ESTRUTURA DA DISSERTAÇÃO

O presente trabalho está organizado da seguinte forma:

- **Capítulo 2** apresenta a fundamentação teórica sobre os sistemas baseados em RFI a fim de prover o arcabouço necessário para o entendimento dos sistemas de detecção de impostores.
- **Capítulo 3** mostra um estudo sistemático de detecção de impostor facial, além de fundamentar os principais tipos de ataques e uma taxonomia dos métodos existentes na literatura.
- **Capítulo 4** descreve a análise quantitativa dos métodos propostos mais relevantes sobre detecção de falsificação de faces na literatura, explicando algumas métricas utilizadas nas bases de dados e comparação dos resultados obtidos pelos métodos. Apresenta tendências atuais e discute propostas para questões abertas.
- **Capítulo 5** conclui a pesquisa e apresenta propostas para trabalhos futuros.

[2]http://www.scopus.com/
[3]http://ieeexplore.ieee.org/Xplore/home.jsp
[4]http://www.engineeringvillage.com/
[5]http://www-periodicos-capes-gov-br.ez10.periodicos.capes.gov.br/

Capítulo

2

RECONHECIMENTO FACIAL EM IMAGENS

O reconhecimento facial em imagens tem por objetivo identificar automaticamente um indivíduo específico, utilizando como referência uma base de dados de imagens de faces previamente cadastradas. Nos últimos anos, diversos trabalhos têm demonstrado uma evolução significativa nesta área, propondo vários tipos de técnicas e abordagens (RAGHAVENDRA et al., 2013), (WRIGHT; HUA, 2009), (PINTO; DICARLO; COX, 2009), (GHIASS et al., 2014), (CAO et al., 2010).

Neste capítulo, será introduzido um breve histórico, bem como as principais categorias de métodos de reconhecimento e tipo de análise realizada sobre os indivíduos em diferentes cenários na aquisição dos dados imagéticos. Em seguida, serão abordadas as etapas encontradas em um *pipeline* típico de um sistema de identificação de rostos. Por fim, serão apresentadas algumas técnicas propostas para tais sistemas, existentes na literatura.

2.1 INTRODUÇÃO AO RECONHECIMENTO FACIAL EM IMAGENS

O interesse em técnicas automatizadas para reconhecimento facial tem fomentado diversos trabalhos de pesquisa nesta área (SAMAL; IYENGAR, 1992), (ZHAO et al., 2003), (GHIASS et al., 2014). Pode-se atribuir este interesse a razões como: disponibilidade de máquinas com alto poder de processamento e tamanho reduzido, e a necessidade crescente de aplicações relacionados à segurança, apenas para citar alguns exemplos.

Pesquisas nesta área foram iniciadas entre as décadas de 60 e 70 (BLEDSOE, 1964), (KANADE, 1973), (KELLY, 1970). No início, as tarefas de reconhecimento facial utilizavam sistemas semi-automatizados, com a participação de humanos para localizar características da face, tais como olhos, nariz, boca e orelhas (KANADE, 1973), (KELLY, 1970). Mais tarde, as técnicas de extração de características nas imagens e de redução das dimensões dessas características foram sugeridas em (SIROVICH; KIRBY, 1987), buscando acelerar o processo de reconhecimento e torná-lo mais automático. No início da década de 90, Turk et al. (1991) utilizaram o método proposto em (SIROVICH; KIRBY, 1987) para a identificação automática de faces humanas com um desempenho computacional próximo ao tempo de execução. O trabalho em (CHELLAPPA; WILSON;

SIROHEY, 1995) apresenta estudos da época sobre o reconhecimento facial em tempo real voltados para as áreas comerciais e governamentais. Mais recentemente, ocorreram significativos avanços nesta área com pesquisas utilizando imagens tridimensionais (3D) (YUAN; LU; YAHAGI, 2005), (CHEN; YAO; CHAM, 2007), (INAN; HALICI, 2012), (NIINUMA; HAN; JAIN, 2013).

A área de reconhecimento facial em imagens tem aplicações em alguns segmentos como vídeo vigilância (AN; BHANU; YANG, 2012a), (AN; KAFAI; BHANU, 2012b), (GORODNICHY; GRANGER, 2014), (PRINOSIL, 2013), investigação forense (JAIN; KLARE; PARK, 2011), (PEACOCK; GOODE; BRETT, 2004) (AULSEBROOK et al., 1995) e interação homem-máquina (XU et al., 2013), (RADUCANU; DORNAIKA, 2012), (KHAN; MIYAMOTO; MORIE, 2008). Esta área pode ser dividida em duas principais categorias, com respeito ao modo como é feita a correspondência entre as imagens: verificação e identificação (JAIN; LI, 2005). A **verificação**, também conhecida como correspondência um-para-um, é responsável pela autenticação de uma pessoa, dadas duas de suas imagens: uma imagem nova e outra previamente cadastrada no sistema. Um exemplo de aplicação que usa a verificação pode ser encontrado nos serviços de imigração, onde a autenticação do passageiro é feita pela imagem obtida do seu passaporte. A **identificação**, também conhecida como correspondência um-para-todos, tem como objetivo de localizar uma imagem qualquer (não-autenticada) correspondente à face consultada em uma base de dados. O reconhecimento de faces em câmeras de vigilância é um exemplo de uma aplicação de identificação (JAIN; LI, 2005). Quanto ao modo de colaboração do usuário, a área de reconhecimento facial pode ser classificada em dois tipos: (i) cenários em que o usuário coopera (por exemplo, autenticação no celular (NG; SAVVIDES; KHOSLA, 2005), controle de acesso físico a um determinado ambiente (ZELJKOVIC et al., 2014) e sistemas de interação homem-máquina (KHAN; MIYAMOTO; MORIE, 2008)), e (ii) cenários em que o usuário não coopera, tais como sistemas de monitoramento de segurança (KAMGAR-PARSI; LAWSON; KAMGAR-PARSI, 2011).

2.2 PIPELINE DE SISTEMAS DE RECONHECIMENTO FACIAL

De um modo geral, o *pipeline* de sistemas de reconhecimento facial é composto por quatro etapas gerais: (i) localização da face e identificação de pontos relevantes na mesma, conhecidos como pontos fiduciais; (ii) normalização da imagem; (iii) extração de características e (iv) associação das características extraídas entre duas faces, conforme é mostrado na Fig. 2.1 e descritos nos itens a seguir (JAIN; LI, 2005):

- Na **localização da face**, o objetivo é detectar regiões na imagem que possuam faces a serem reconhecidas. No caso específico de se trabalhar com vídeos, pode-se realizar o rastreamento destas regiões em múltiplos quadros, com o objetivo de obter maior precisão a partir de um rastreamento da face. A detecção fornece uma estimativa da localização e escala da face, na qual é possível localizar os **pontos fiduciais**. Estes pontos são localizados em regiões peculiares da face humana, como exemplo: regiões dos olhos, nariz e boca. Por meio desses pontos são extraídas características que distinguem um indivíduo de outro.

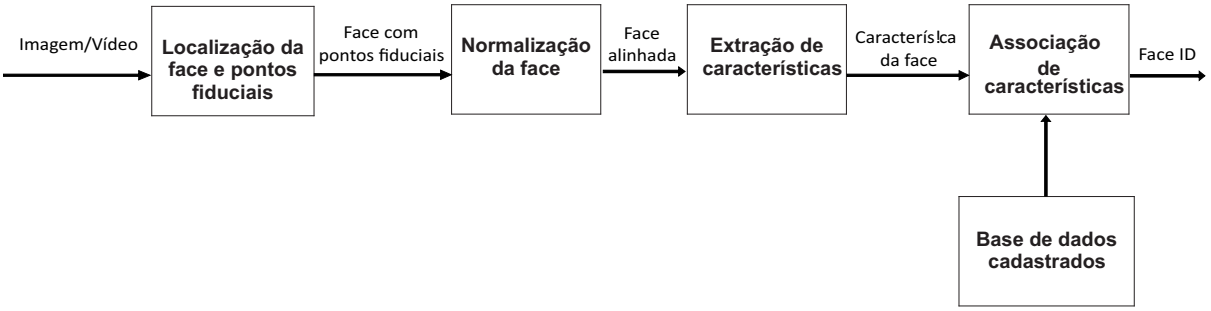

Figura 2.1 *Pipeline* geral de um sistema de reconhecimento facial. Seguindo as setas da esquerda para a direita: a imagem ou sequência de vídeo de entrada passa pela etapa de localização da face e pontos fiduciais, a fim de obter uma imagem facial com estes pontos; em seguida, é alinhada em relação à posição dos olhos pela etapa de normalização da face; após o alinhamento, características são extraídas para melhor representar a face de modo único. Na etapa de associação, os modelos criados com base nestas características são comparados com outros similares, previamente cadastrados na base de dados. Imagem adaptada de (JAIN; LI, 2005).

- O objetivo da **normalização da face** é reduzir invariâncias na imagem, tais como pose e iluminação. Nesta etapa, são realizadas operações de transformações geométricas e fotométricas na região da face. O processo de normalização geométrica tem como objetivo alinhar algumas regiões do rosto em relação à altura dos olhos. O processo de normalização fotométrica transforma a imagem de modo a deixar padrão as suas características de luminosidade e tonalidades de cor.

- A finalidade da **extração de características** é criar uma representação codificada da imagem da face. O desafio desta etapa é que a representação criada seja, ao mesmo tempo, semelhante para imagens de um mesmo indivíduo e distinta para imagens de indivíduos diferentes.

- A **associação das características extraídas entre duas faces** é a etapa de reconhecimento propriamente dita. É neste ponto que é feita uma comparação das características extraídas da imagem de entrada com as características extraídas das imagens na **base de dados cadastrada**, com o intuito de encontrar a face mais semelhante a da imagem de entrada. Esta comparação pode ser feita por uma métrica de distância no espaço das características ou por meio de classificadores.

2.3 REVISÃO DAS TÉCNICAS EXISTENTES EM RECONHECIMENTO FACIAL

Nas últimas décadas, diversas técnicas têm sido utilizadas para o reconhecimento facial. De acordo com Tolba, El-Baz e El-Harby (2006), as principais técnicas de reconhecimento de face, que se aplicam principalmente a imagens de faces frontais, são: *eigenfaces*, redes neurais artificiais, *graph matching, hidden markov models, geometric feature matching, template matching* e *3D morphable model*. A seguir cada categoria e as técnicas associadas são apresentadas.

2.3.1 Técnicas baseadas em Eigenfaces

Eigenfaces é uma das técnicas mais utilizadas para o processo de reconhecimento facial a partir de imagens (SIROVICH; KIRBY, 1987), (KIRBY; SIROVICH, 1990), (TURK; PENTLAND, 1991), (ZHAO; YANG, 1999), (PENTLAND; MOGHADDAM; STARNER, 1994), (HESELTINE; PEARS; AUSTIN, 2002), (BHOWMIK et al., 2008). As *eigenfaces* são baseadas na análise de componentes principais (PCA)[1] (SIROVICH; KIRBY, 1987), (KIRBY; SIROVICH, 1990). As *Eigenfaces* são obtidas a partir dos seguintes passos:

- Passo 1: obter uma base de dados de treinamento de imagens faciais, em que todas as imagens tenham as mesmas dimensões de largura e altura, e estejam representadas em tons de cinza. A base de dados de treino é formada por Z imagens, Γ_i, com $i \in Z$, as quais são representadas em uma matriz de dimensão $M \times N$.

- Passo 2: cada imagem da face do conjunto de treinamento varia em relação à face média, Ψ, obtida por

$$\Psi = \frac{1}{Z} \sum_{i=1}^{Z} \Gamma_i \,. \tag{2.1}$$

- Passo 3: é calculada a diferença, Φ, entre cada imagem Γ e a imagem da face média Ψ, de acordo com

$$\Phi_i = \Gamma_i - \Psi \,. \tag{2.2}$$

- Passo 4: a partir do resultado da diferença, Φ, obtemos uma nova matriz A que contém somente as variações de cada imagem da face em relação à face média, dadas por

$$A = [\Phi_1, \Phi_2, ..., \Phi_Z] \,, \tag{2.3}$$

  e o cálculo da matriz de covariância, C, é obtido por

$$C = AA^T \,. \tag{2.4}$$

- Passo 5: computar os autovalores e autovetores da matriz C é intratável em virtude da alta dimensionalidade da matriz A. Com isso, há uma necessidade de obter uma nova combinação linear, L, com o uso do PCA para redução de dimensionalidade, de acordo com

$$L = A^T A \,. \tag{2.5}$$

[1] Do Inglês, *principal component analysis* (PCA).

- Passo 6: os autovalores, λ, e autovetores, v, da matriz L são obtidos por

$$Lv = \lambda v \,, \tag{2.6}$$

onde v é determinado pela combinação linear de Z imagens de treino para formar as *eigenfaces* $\mathbf{u}_l$, dada por

$$u_l = \sum_{k=1}^{Z} v_{lk} \Phi_k \,, \tag{2.7}$$

com $l = 1, ..., Z$.

- Passo 7: as imagens de treinamento Γ_k, são projetadas no espaço vetorial de face, efetuando-se a operação

$$\omega_k = u_k^T (\Gamma_k - \Psi) \,, \tag{2.8}$$

onde $k = 1, ..., Z'$, em que Z' são os autovetores associados aos autovalores mais significativos. Estes autovetores determinam as combinações lineares das imagens de treino que dão origem às eigenfaces u_k. Os pesos ω_k representam a importância de cada *eigenface* na representação de cada imagem de treino Γ. Esses pesos são armazenados em um vetor $\Omega^T = [\omega_1, \omega_2, ..., \omega_Z]$.

- Passo 8: por fim, na fase de reconhecimento, uma nova imagem de face (teste) é identificada a partir da similaridade entre esta e a combinação linear obtida no treinamento. O procedimento de similaridade é realizado por meio do cálculo da distância Euclidiana, ε_k, entre os pesos obtidos nas imagens de treino Ω_k e teste Ω

$$\varepsilon_k^2 = \|(\Omega - \Omega_k)\|^2 \,. \tag{2.9}$$

Se a distância entre essas imagens for menor que o limiar θ_ε, então as imagens são semelhantes, caso contrário, as imagens não são similares.

As *eigenfaces* foram propostas por TURK e PENTLAND (1991) na detecção e identificação de faces; o objetivo em (TURK; PENTLAND, 1991) foi adquirir as imagens faciais sob variações de iluminação, de dimensão e de orientação da cabeça para serem processadas com a técnica *eigenfaces*. Cada imagem foi normalizada usando a localização da posição dos olhos em termos de rotação e escala. A Fig. 2.2 mostra algumas *eigenfaces* geradas a partir de 2500 imagens de face, que foram computadas sem o plano de fundo removido. Os autores obtiveram taxas de precisão de 96%, 64% e 85%, após variações de iluminação, dimensão e orientação da cabeça, respectivamente. ZHAO e YANG (1999) propuseram um método para calcular a matriz de covariância das *eigenfaces* usando três imagens faciais com diferentes condições de iluminação, para serem analisados os efeitos de iluminação. Essa análise considerou as mudanças no tipo de fonte de luz, número,

Figura 2.2 Sete *eigenfaces* foram calculadas usando uma base de dados de 2500 imagens de faces digitalizadas de dezesseis indivíduos. As imagens possuem variações de iluminação, dimensão da imagem e orientação da cabeça. Imagem retirada de (TURK; PENTLAND, 1991).

intensidade e os efeitos de luminosidade tais como: reflexo, sombra e brilho. Pentland, Moghaddam e Starner (1994) aplicaram uma extensão da técnica *eigenfaces* para *eigenfeatures* correspondente as características faciais, tais como: olhos, nariz e boca (referidas como *eigeneyes*, *eigennose* e *eigenmouth*). As *eigenfeatures* foram menos sensíveis às mudanças na aparência da face que a técnica padrão *eigenfaces*. Os resultados obtidos em (PENTLAND; MOGHADDAM; STARNER, 1994) com as *eigenfeatures* atingiram uma taxa de reconhecimento de 95% em 7.562 imagens faciais, de aproximadamente 3.000 indivíduos.

Uma variedade de técnicas de processamento de imagem para melhorar o desempenho da técnica *eigenfaces* foi proposta por (HESELTINE; PEARS; AUSTIN, 2002). As técnicas de processamento foram classificadas em quatro principais categorias: normalização de cores (por exemplo, intensidade da cor, tonalidade cor cinza, etc), estatísticas (por exemplo, brilho, media local do brilho, etc), convolução (por exemplo, suavização, borramento, contorno, etc) e combinações das técnicas (por exemplo, filtragem do contorno seguido de suavização, transformação de brilho local seguido por filtro de contorno, etc). A Fig. 2.3 ilustra as etapas desenvolvidas na abordagem proposta por processamento de imagens, cálculos para a obtenção das *eigenfaces*, o reconhecimento por meio da distância Euclidiana e análises de métricas de taxas de erro.

Uma abordagem para calcular as *eigenfaces* a partir de imagens térmicas de face em coordenadas polares foi desenvolvida em (BHOWMIK et al., 2008). Em cada imagem houve a transformação log-polar que tenta garantir a invariância às diferenças de rotação e escala. Em seguida, as imagens térmicas de treino e teste foram projetadas dentro do espaço vetorial do rosto, denominado como *eigenfaces* térmicas polares. Por fim, na etapa de classificação foi utilizado a técnica *multilayer perceptron* (MLP). A Fig. 2.4 mostra um diagrama de blocos para a tarefa de reconhecimento. Os resultados experimentais mostraram que a taxa de precisão atingiu 97.05% utilizando a base de dados $OTCBVS$[2] imagens térmicas de face.

De acordo com os trabalhos apontados acima, as *eigenfaces* mostraram ser uma técnica simples e eficiente, que utiliza a matriz de covariância aplicada nas imagens de face para serem extraídas suas características e classificadas por algum cálculo de distância. Entretanto, essa técnica tem alguma invariância com relação a mudanças na escala, rotação e luminosidade.

[2]http://vcipl-okstate.org/pbvs/bench/

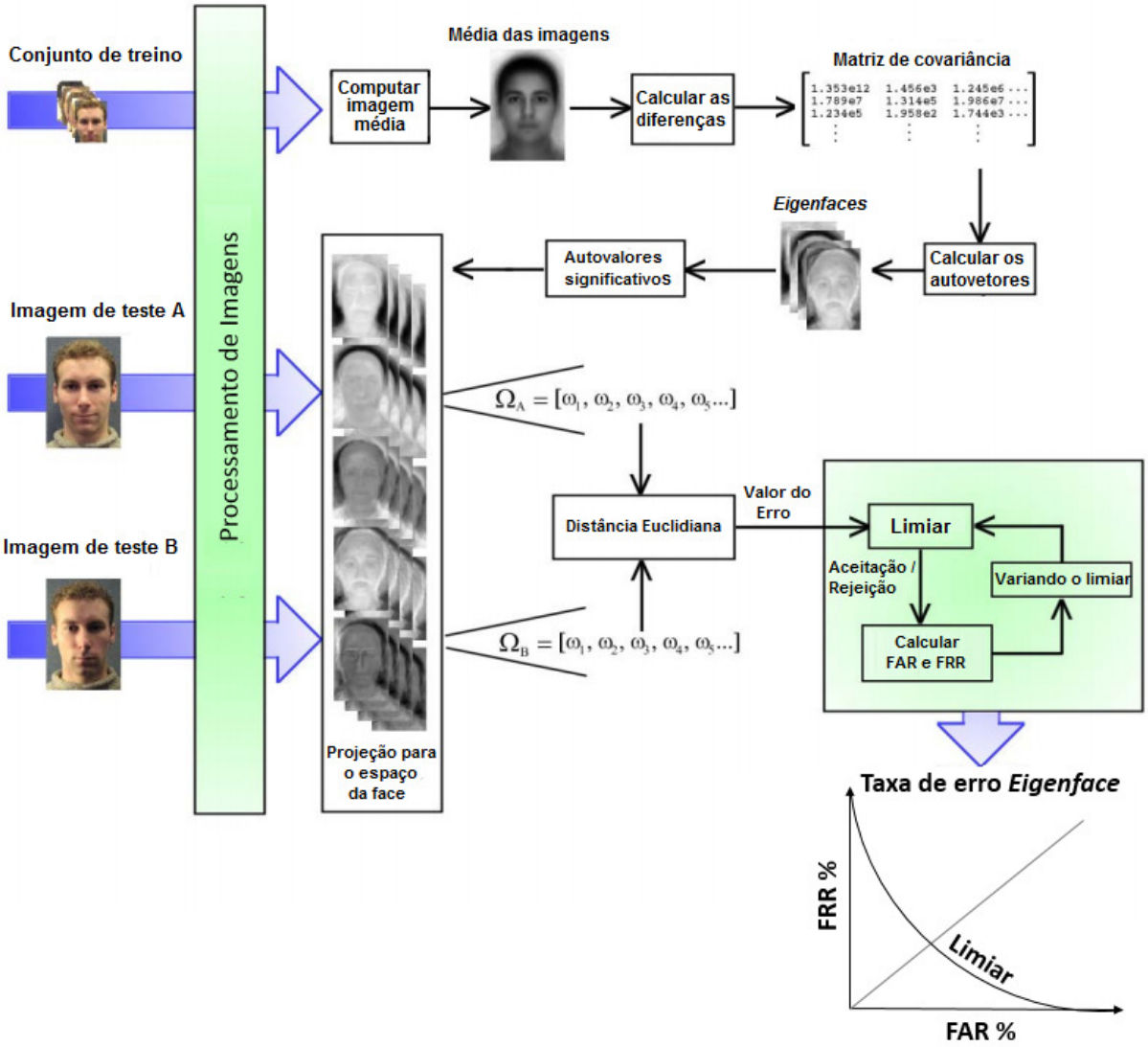

Figura 2.3 Sistema de reconhecimento facial utilizando a técnica *eigenfaces*. Seguindo as setas: as imagens da base de dados de treinamento passam por uma etapa de pré-processamento e são computadas a média dessas imagens; em seguida, são calculadas as diferenças das imagens e os autovetores da matriz de covariância que formam as *eigenfaces*; após isso, é criado um modelo de autovalores. As imagens de teste A e B foram pré-processadas e combinadas com os autovalores correspondente as imagens de treino para serem projetadas para o espaço das faces. Imagem adaptada de (HESELTINE; PEARS; AUSTIN, 2002).

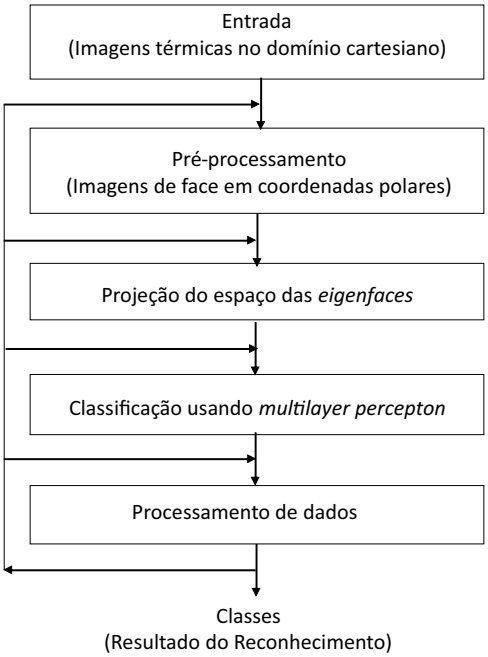

Figura 2.4 Diagrama de blocos do reconhecimento facial por meio de imagens térmicas. Seguindo as setas de cima para baixo: as imagens térmicas de faces no domínio cartesiano passam pela etapa de entrada dos dados; em seguida, as imagens faciais são representadas em coordenadas polares pela etapa de pré-processamento; após esta etapa, as *eigenfaces* são projetadas no espaço bidimensional; na etapa de classificação, os modelos criados com estas extrações de características são usados com MLP. Imagem adaptada de (BHOWMIK et al., 2008).

2.3.2 Técnicas baseadas em Redes Neurais Artificiais.

O funcionamento das redes neurais artificiais (RNA) baseia-se em uma estrutura de elementos de processamento e conexões (STONHAM, 1986), (SUNG; POGGIO, 1995), (LAWRENCE et al., 1997). O elemento principal de um processamento em uma RNA é denominado neurônio. A Fig. 2.5 apresenta um modelo de um neurônio artificial, que pode ser identificado por três elementos básicos:

- Um conjunto de **sinapses**: cada uma consiste na entrada dos sinais por meio de um peso para o neurônio. Especificamente, para cada sinal de entrada X_n, conectada ao neurônio k, uma operação de multiplicação pelo peso sináptico W_{kn} é realizada. Esse peso sináptico de um neurônio pode estar em um intervalo que contenha valores positivos e negativos;

- Um **somador** $\sum$: para somar todos os resultados obtidos da multiplicação dos sinais de entrada pelos pesos;

- Uma **função de ativação** $\varphi(.)$: responsável por obter o resultado U_k, do somatório, e delimitar o sinal de saída Y_k com um único valor, utilizado como entrada para o próximo neurônio ou como resposta da RNA.

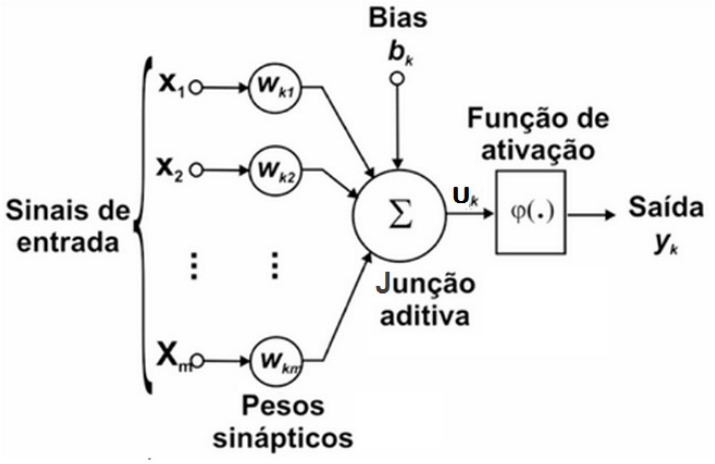

Figura 2.5 Modelo não-linear de um neurônio artificial. Seguindo as setas: os sinais de entrada X_1,X_2,...,X_n, os pesos sinápticos W_{k1},W_{k2},...,W_{kn} e um parâmetro polarizador (bias) b_k, de um neurônio artificial k, são processados em uma junção aditiva, cuja saída, U_k, é submetida a uma função de ativação para obter um valor finito Y_k. Imagem adaptada de (HAYKIN, 2000).

O modelo de neurônio da Fig. 2.5 também contém um elemento polarizador (bias), aplicado externamente ao neurônio. O bias b_k tem o objetivo de regular o valor do hiperplano de separação, composto pela combinação linear entre pesos e elementos de entrada (HAYKIN, 2000).

A organização das camadas de uma RNA define a forma como os neurônios da rede estão organizados, o que, por sua vez, define a arquitetura da rede. Essas RNAs podem ser identificadas por três classes de arquitetura:

1. **Redes alimentadas adiante com camada única**: esta classe de RNA possui uma camada de entrada contendo neurônios que se conecta a uma camada de saída. A definição de camada única refere-se à camada de saída; na camada de entrada, não é executada qualquer computação. A Fig. 2.6 mostra um exemplo de rede alimentada adiante com camada única.

2. **Redes alimentadas diretamente com múltiplas camadas** (MLP)[3]: esta rede se diferencia pela presença de uma ou mais camadas ocultas, as quais têm a função de processar os sinais de entrada antes de transmiti-los aos neurônios de saída. A Fig. 2.7 ilustra um exemplo de rede totalmente conectada, onde todos os neurônios da camada anterior estão conectados a todos os outros neurônios da camada seguinte. Caso contrário, a rede é dita parcialmente conectada.

3. **Redes recorrentes**: esta classe se diferencia das outras redes alimentadas anteriores por ter pelo menos um laço de realimentação. As conexões de realimentação são utilizadas em uma situação onde os neurônios ocultos bem como os neurônios de

[3]Do Inglês, *feed forward multilayer perceptron* (MLP).

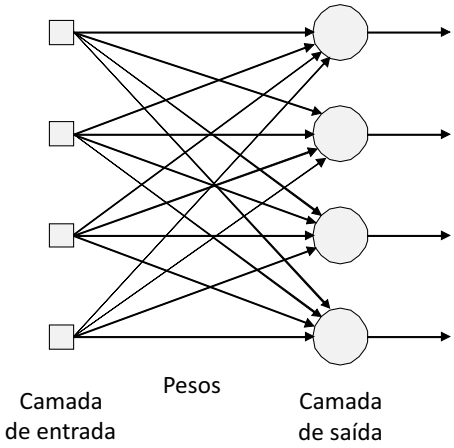

Figura 2.6 Rede alimentada adiante com uma única camada de neurônios. Seguindo as setas: os quatro neurônios da camada de entrada com os seus pesos são conectados a quatro neurônios da camada de saída. Imagem adaptada de (HAYKIN, 2000).

saída são realimentados para a sua própria entrada. A Fig. 2.8 mostra um exemplo de redes recorrentes.

As redes MLP's foram utilizadas em alguns trabalhos para reconhecimento de faces em imagens (BOUGHRARA; CHTOUROU; AMAR, 2012), (BOUGHRARA et al., 2014). Pode haver arquiteturas de RNAs híbridas, tal como encontrado em (LAWRENCE et al., 1997). Essa forma é uma combinação de uma representação de amostra de imagem

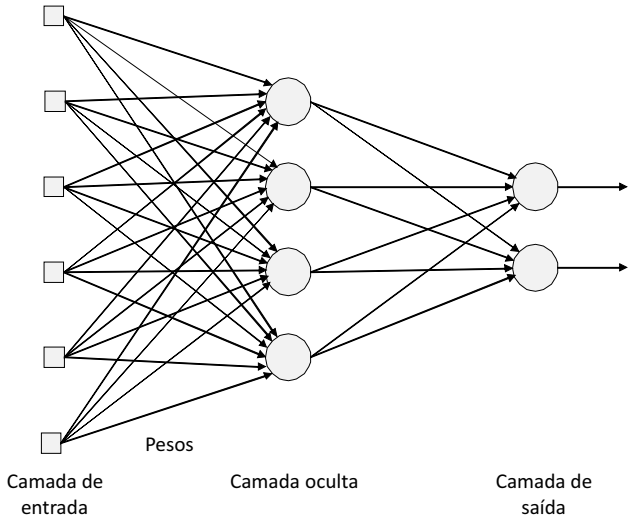

Figura 2.7 Rede alimentada adiante com múltiplas camadas de neurônios. Seguindo as setas da esquerda para direita: seis neurônios da camada de entrada com os seus pesos são conectados para quatro neurônios da camada oculta, e depois são processadas em dois neurônios da camada de saída. Imagem adaptada de (HAYKIN, 2000).

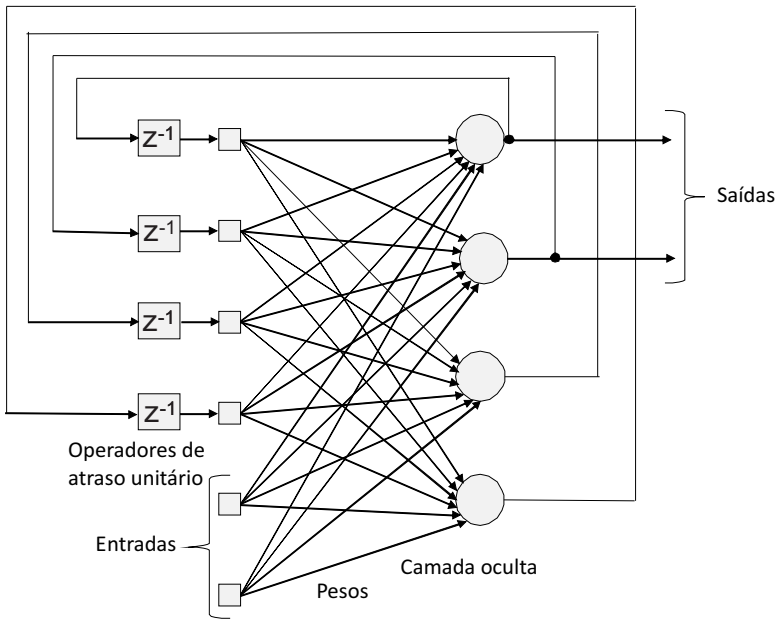

Figura 2.8 Rede recorrente com neurônios ocultos. Seguindo as setas da esquerda para direita: os sinais de entrada com os seus respectivos pesos são processados nos quatro neurônios da camada oculta, e depois são realizadas as conexões de realimentação, que tem uma importância satisfatória na aprendizagem da rede por meio de operadores de atraso unitário representado por Z^{-1}, no qual resulta em um comportamento não-linear. Imagem adaptada de (HAYKIN, 2000).

local entre uma rede neural por Mapa Auto-Organizável (SOM)[4] e uma rede neural por convolução (CNN)[5]. As redes SOM são capazes de reduzir a dimensão de um conjunto de dados de amostras, obtendo uma representação de tamanho inferior à original; tais mapas são capazes de manter as relações de vizinhança dos dados de entrada. Outra característica deste tipo de rede é que elas utilizam treinamento não-supervisionado, onde a rede procura semelhanças baseando-se apenas nos padrões de entrada. A rede neural por convolução (CNN) representa as imagens de faces com invariância a operações de translação, rotação, escala e deformação da imagem de entrada. Em (LAWRENCE et al., 1997), os resultados atingiram uma taxa de reconhecimento de 96,2%, quando o método proposto foi avaliado sobre 400 imagens faciais de 40 indivíduos na base de dados *ORL*[6].

Uma abordagem de rede probabilística baseada em rede neural (PDBNN)[7] foi desenvolvida por Lin, Kung e Lin (1997) para um sistema de reconhecimento automático de face. Este sistema executa a detecção da face, a localização dos olhos e o reconhecimento em um funcionamento próximo do tempo real.

[4]Do Inglês, *self organized map* (SOM).
[5]Do Inglês, *convolutional neural network* (CNN).
[6]http://www.camorl.co.uk/facedatabase.html
[7]Do Inglês, *probabilistic decision based neural network* (PDBNN).

2.3.3 Técnicas baseadas em Graph Matching

É uma técnica aplicada para a tarefa de reconhecimento de faces, onde a imagem de uma face é representada por um grafo. Em tais grafos, os pontos fiduciais obtidos na face são representados por nós, e as relações entre tais pontos são definidas a partir das arestas dos grafos (LADES et al., 1993), (WISKOTT; MALSBURG, 1996). Em cada ponto fiducial, um filtro de Gabor é utilizado para representar a região em torno do ponto. Esta representação é então armazenada nos nós do grafo, em que os filtros de Gabor têm como objetivo extrair as características de textura presentes na face (JI; CHANG; HUNG, 2004). A Fig. 2.9 ilustra uma representação do grafo sob uma face.

Uma variação de *graph matching* é a *Elastic Graph Matching* (EGM), que é uma abordagem utilizada no reconhecimento de face para tratar variações de poses e de expressões faciais (WISKOTT et al., 1997), (DUC; FISCHER; BIGÜN, 1999), (ZHANG; YAN; LADES, 1997). Wiskott et al. (1997) introduziram uma abordagem *bunch graph* para extração de características a partir de imagens de face. Esta abordagem utiliza um componente *wavelet* para cada ponto fiducial da face (olhos, boca, etc.), assim criando um grafo da imagem da face. Esses componentes são vetores de características com informações de frequência e orientação para cada nó rotulado do grafo. A Fig. 2.10 mostra grafos para encontrar e reconhecer imagens de faces. Wiskott et al. (1997) avaliaram a taxa de reconhecimento a partir de duas bases de dados, *Feret*[8] e *Bochum*[9]. Würtz (1997), propôs um método de reconhecimento de face eficiente sob mudanças do plano de fundo, pequenas deformações e translação. Duc, Fischer e Bigün (1999) propuseram as deformações do EGM utilizando filtros Gabor para autenticação de faces. Os filtros de Gabor foram extraídos com seis diferentes orientações e três escalas. O método proposto obteve uma taxa de erro de 6,1% utilizando a métrica *equal error rate* (EER). Zhang et al. (1997) avaliaram três técnicas: *elastic matching, eigenfaces* e RNA no reconheci-

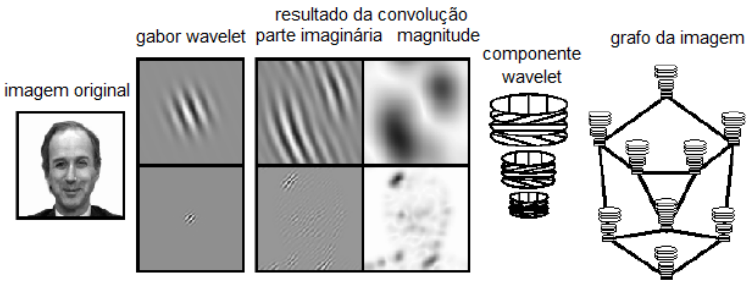

Figura 2.9 Um exemplo de representação de grafo em uma imagem de face. Uma imagem de face passa por uma transformada Gabor *wavelet* resultando em uma convolução com um conjunto de *wavelets*, onde foram computadas 12 coeficientes (3 frequências x 4 orientações). O conjunto de componentes *wavelets* constitui um grafo da imagem, usada para representar uma face. Imagem adaptada de (WISKOTT et al., 1997).

[8]http://www.nist.gov/itl/iad/ig/colorferet.cfm
[9]http://www.ini.rub.de/pages/contact

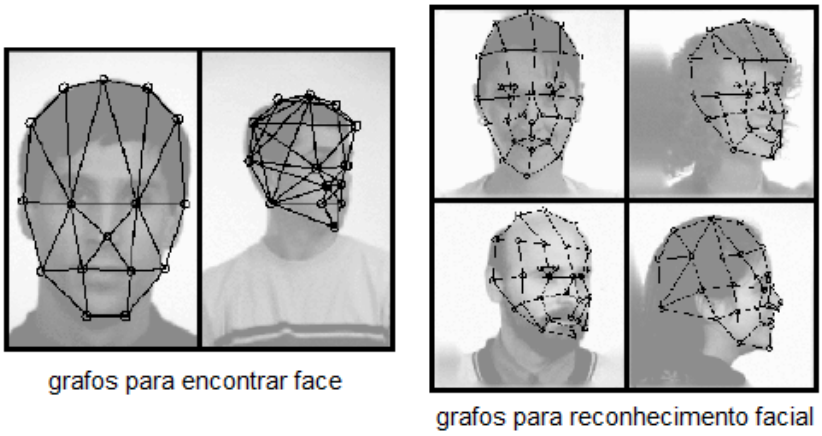

Figura 2.10 Imagens de grafos para diferentes posições de face. Os nós são posicionados pela técnica EGM. As duas imagens na esquerda possuem diferentes tamanho de face, onde foram selecionadas no procedimento de encontrar um rosto. As imagens à direita já estão redimensionadas para o tamanho normal e foram utilizadas no processo de reconhecimento facial por terem mais nós no grafo. Imagem adaptada de (WISKOTT et al., 1997).

mento de faces a partir de quatro bases de dados diferentes, tais como: *MIT*[10], *Olivetti*[11], *Weizmann*[12] e *Bern*[13], onde a técnica *elastic matching* obteve taxas de reconhecimento maiores e iguais em relação a técnica *eigenfaces* e superior a RNA. A vantagem da *elastic matching* é a invariância em diferentes cenários de iluminação, posição da face e expressão facial.

Uma variação do EGM chamada de *Morphological EGM* (MEGM), foi avaliada para identificação de faces frontais descrita em (KOTROPOULOS; TEFAS; PITAS, 2000). Essa variação utiliza transformações morfológicas, tais como: erosão e dilatação, que na saída da transformação representa um vetor de características dos nós do grafo da imagem facial. Os resultados alcançados na base de dados *Ibermatica* (KOTROPOULOS et al., 1999) mostrou que a melhor taxa de erro (EER) foi de 20% com a normalização na região da face antes de ser processada pela MEGM. Essa normalização foi aplicada sobre os valores do pixels na região da pele do rosto a partir da detecção representada em formato de elipse. Uma outra extensão do EGM denominada de *Generalized EGM* (GEGM) foi proposta em (SHIN; KIM; CHOI, 2007), que obteve desempenho superior ao método convencional EGM em diferentes escalas e rotações. GEGM foi utilizado com os parâmetros de deformações entre os grafos correspondente de faces, assim otimizando os nós e arestas para o processo de reconhecimento facial. Para tal reconhecimento foram utilizadas 940 imagens de face da base de dados do *Feret*[14], próximas da posição frontal da face com um ângulo de inclinação de ± 22.5 graus.

[10] ftp://whitechapel.media.mit.edu/pub/images
[11] http://www.scikit-learn.org/stable/datasets/olivetti_faces.html
[12] http://www.wisdom.weizmann.ac.il/∼ vision/FaceBase/
[13] http://www.ph.tn.tudelft.nl/PRInfo/data/msg00010.html
[14] http://www.nist.gov/itl/iad/ig/colorferet.cfm

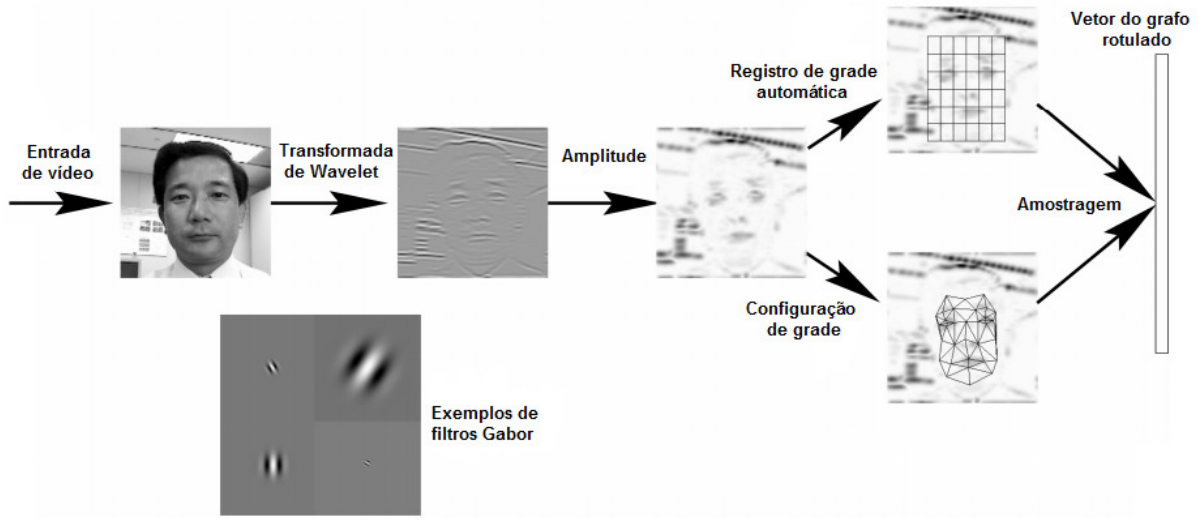

Figura 2.11 A representação do EGM extraída por meio do filtro Gabor de uma imagem facial. Seguindo as setas da esquerda para direita: os *frames* do vídeo são processados pela transformada de Gabor *wavelet* com diferentes filtros. Em seguida, foi obtida a amplitude das transformadas de *wavelet* e depois avaliada em dois tipos de grade na região da face: de forma retangular e nós de grafos ajustáveis por meio de pontos fiduciais. Após isso, a amostragem dessas grades são inseridas no vetor de grafos rotulados. Imagem adaptada de (LYONS; BUDYNEK; AKAMATSU, 1999).

Um método para avaliar as imagens faciais baseadas em EGM e em análise de discriminantes lineares (LDA)[15] foi proposto em (LYONS; BUDYNEK; AKAMATSU, 1999). O EGM foi aplicado com a utilização do filtro Gabor para classificar imagens de face com base em gênero, etnia e expressões faciais. A Fig. 2.11 mostra uma representação do EGM por meio do filtro de Gabor, bem como os componentes *wavelet*. A classificação foi realizada a partir da base de dados ATR[16], o qual conseguiu uma taxa de desempenho de reconhecimento de 92% para gênero, 95% para etnia e 91% para expressões faciais.

Os trabalhos que utilizam a combinação de grafo possuem como base a extração e correspondência de características posicionadas sobre pontos fiduciais da face. Esses pontos são os nós ou vértices que contém componentes *wavelets*, e suas arestas são conectadas entre esses pontos, que por sua vez é representada por um grafo.

2.3.4 Técnicas baseadas em Hidden Markov Models (HMMs)

Um modelo de Markov é um processo estocástico, onde a distribuição de probabilidade evolui de um estado para outro dependentemente somente do seu estado anterior. Este modelo é formado por probabilidades de transição de estados, onde os estados são representados em termos de seus vetores probabilísticos, os quais podem variar no tempo, de maneira discreta ou contínua (SAMARIA; FALLSIDE, 1993),(SAMARIA; YOUNG,

[15]Do Inglês, *linear discriminant analysis* (LDA).
[16]http://www.hip.atr.co.jp/

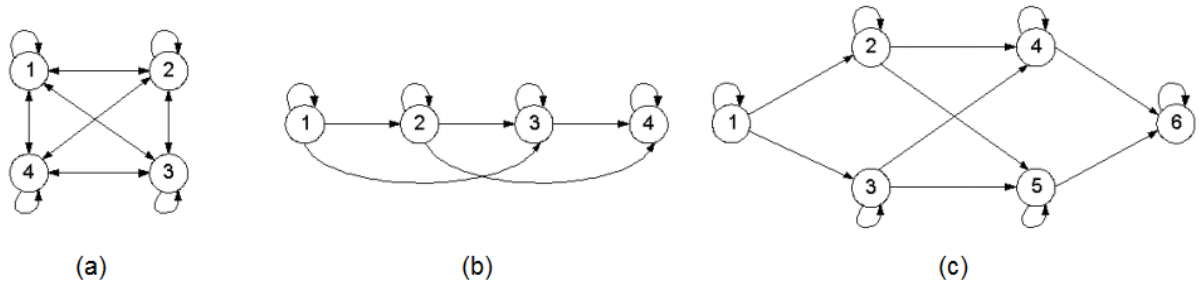

Figura 2.12 Tipos de estrutura dos modelos de Markov: a) modelo sem restrições; b) modelo sequencial; c) modelo paralelo. Imagem retirada de (YACOUBI, 1996).

1994).

Quando os modelos de Markov são utilizados em um espaço de estados desconhecidos, onde é possível definir um processo estocástico considerando uma aproximação desse espaço, o processo recebe o nome de modelos ocultos de Markov (HMM)[17]. O HMM consiste de um processo duplamente estocástico formado por uma variável oculta (*hidden*), mas que se manifesta por meio de uma outra variável estocástica que gera a sequência de símbolos observados (RABINER; JUANG, 1986). De modo geral, existem dois tipos de modelos para os HMM's (YACOUBI, 1996):

- Nos **modelos sem restrições**, todas as transições possíveis entre os estados do HMM são permitidas. Isto é factível, se não houver restrição de nulidade a nenhum dos valores da matriz de transição dos estados (um exemplo deste modelo pode ser visto na Fig. 2.12(a));

- Os **modelos esquerda-direita** podem ser classificados em modelos sequenciais e paralelos. Os **modelos sequenciais** operam segundo uma evolução em série por meio de seus estados, e pode ocorrer a conexão de um estado atual com estado posterior não sequencial no desenvolvimento do HMM (um exemplo deste modelo pode ser visto na Fig. 2.12(b)). Para os **modelos paralelos**, diversas trajetórias através do HMM são permitidas, sabendo que cada uma desses trajetórias conecta um ou vários estados do modelo (um exemplo deste modelo pode ser visto na Fig. 2.12(c)).

Nefian e Hayes III (1998a) utilizaram os HMM's tanto para detecção como para reconhecimento de faces frontais. No método proposto em (NEFIAN; HAYES III, 1998a), a imagem da face foi dividida em cinco blocos (cabelo, testa, olhos, nariz e boca), onde cada bloco é representado como um estado no HMM, conforme ilustrado na Fig. 2.13. Esses blocos são extraídos dos coeficientes da Transformada de Karhunen-Loeve (TKL), conhecido como *principal component analysis* (PCA). A TKL é utilizada para encontrar os vetores mais significativos por uma redução de dimensionalidade, onde esses vetores extraídos de cada bloco são associados a um estado do HMM e utilizado para obter as

[17]Do Inglês, *hidden markov models (HMM)*.

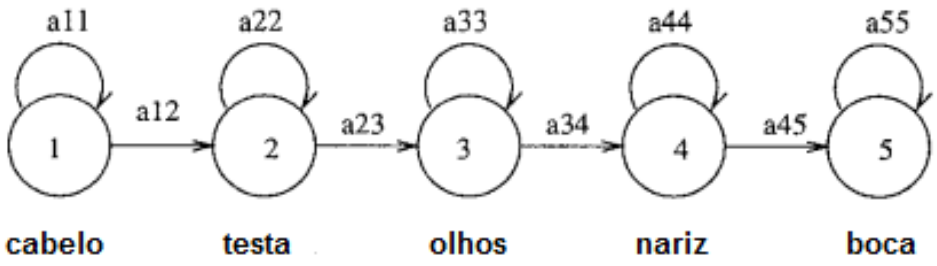

Figura 2.13 Cada estado do HMM é associado a uma região da face. Seguindo as setas da esquerda para direita: as regiões do cabelo, testa, olhos, nariz e boca são representadas como um conjunto de estados do HMM e as probabilidades (autômato finito) são computadas a partir das transições entre os estados. Imagem adaptada de (NEFIAN; HAYES III, 1998b).

estimativas iniciais da matriz de probabilidades de símbolos de observação. A taxa de reconhecimento de faces foi de 86% utilizando a base de dados *Olivetti*[18], que consiste de 400 imagens de 40 indivíduos, tendo 10 imagens para cada indivíduo.

O método em (NEFIAN; HAYES III, 1998b) foi aplicado com base no HMM para reconhecimento de face, que utilizou para a extração de características os coeficientes da transformada discreta do cosseno[19] na imagem da face distribuída em cinco blocos, com tais blocos, representando cada região da face (cabelo, testa, olhos, nariz e boca). Essa forma de extração reduziu substancialmente o tamanho do vetor de características. Os resultados obtidos com a técnica HMM atingiram uma taxa de reconhecimento de 84%, enquanto as *eigenfaces* tiveram 73%, sobre a base de dados *Olivetti*[20].

Samaria e Harter (1994) utilizaram uma extensão da técnica HMM chamada de pseudo-2D HMM (P2D-HMM) (KUO; AGAZZI, 1993). Essa extensão é um modelo probabilístico com uma estrutura bidimensional dos estados, que representam um "super-estado", onde cada "super-estado"possui um HMM, conhecido como o "sub-estado". A taxa de erro foi de 5% em seus experimentos, utilizando a base de dados *Olivetti*[21].

Le e Li (2004) propuseram um método utilizando um HMM unidimensional discreto[22] para reconhecimento de faces sob diferentes condições de iluminação, expressão, pose, oclusão e tempo de atraso. Em (LE; LI, 2004), todas as imagens de face são compartilhadas por apenas um HMM, que foi utilizado como uma forma de ponderar os vetores extraídos a partir das imagens.

O HMM aplicado a reconhecimento de faces é formado geralmente por um espaço de estados não observáveis e pela distribuição de probabilidades associadas aos estados, onde cada estado representa uma característica da face, tais como: olhos, nariz e boca. Existem algumas questões para serem tratadas sobre o HMM para reconhecimento de faces, como por exemplo: (i) a decodificação dos estados, dada uma sequência de símbolos

[18]http://scikit-learn.org/stable/datasets/olivetti_faces.html
[19]Do Inglês, *discrete cosine transform* (DCT).
[20]http://scikit-learn.org/stable/datasets/olivetti_faces.html
[21]http://scikit-learn.org/stable/datasets/olivetti_faces.html
[22]Do Inglês, *1D discrete hidden markov model* (1D-DHMM).

de observações e um HMM, que resulte em encontrar a melhor sequência de estados ocultos; (ii) a avaliação do HMM, dado um HMM determinar a probabilidade de uma dada sequência de observações.

2.3.5 Técnicas baseadas em Geometrical Feature Matching

Esta técnica é baseada no cálculo de um conjunto de características geométricas a partir de imagens de faces. Essas características são extraídas e armazenadas em um vetor que representa a posição e dimensão em regiões da face, tais como: olhos, sobrancelhas, nariz, boca e contorno do rosto (KANADE, 1973), (TAMURA; KAWAI; MITSUMOTO, 1996).

Um dos trabalhos pioneiros sobre o reconhecimento de faces de forma automatizada, usando características geométricas, foi proposto por Kanade (1973) que propõe um método para extrair pontos fiduciais da face nas regiões do nariz, da boca e dos olhos. Os resultados experimentais atingiram uma taxa de reconhecimento de 75% em uma base de dados elaborada por Kanade composta de 40 imagens de faces, sendo duas imagens de faces por cada indivíduo. Brunelli e Poggio (1993) propuseram um método para extrair as características geométricas de regiões do nariz, da boca e do queixo. A Fig. 2.14 mostra as características geométricas utilizadas para serem comparadas com as imagens de faces para reconhecimento. Os resultados atingiram uma taxa de reconhecimento de 90% em 188 imagens de 47 indivíduos numa base de dados criada por (BRUNELLI; POGGIO, 1993). Cox, Ghosn e Yianilos (1996) introduziram uma técnica de distância com base no modelo de distribuição gaussiana, onde são computados os parâmetros de média e variância. Os resultados experimentais atingiram uma taxa de reconhecimento de 95% em uma base de dados composta por uma junção de outras bases (*UCSB*[23], *Instituto Weizmann*[24] , *MIT*[25] e *NEC*[26]) totalizando 685 imagens.

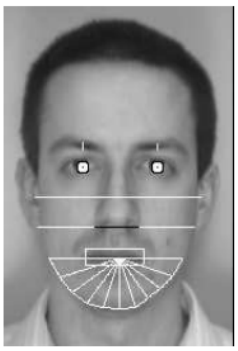

Figura 2.14 Caraterísticas geométricas utilizadas no experimento de reconhecimento de face. As regiões da face são representadas na cor branca, como: olhos, nariz, boca e queixo. Essas regiões foram extraídas pelo método proposto para reconhecimento de face em imagens. Imagem adaptada de (BRUNELLI; POGGIO, 1993).

[23] $http://engineering.ucsb.edu/faculty/profile/141$
[24] $http://www.wisdom.weizmann.ac.il/\ vision/FaceBase/$
[25] $http://vismod.media.mit.edu/vismod/demos/facerec/$
[26] $http://www.nec.com/en/global/rd/research/cl/facerecognition/technologies.html?$

A técnica de *geometrical feature matching* depende dos algoritmos de localização dos pontos fiduciais da face e pode ocorrer problemas no reconhecimento devido à baixa qualidade da imagem ou oclusão parcial da face.

2.3.6 Técnicas baseadas em Template Matching

A técnica de *template matching* consiste na medição de similaridade por meio de correlação entre as imagens armazenadas na base de dados e a imagem de entrada a ser consultada. Após a medição, é definido um limiar com base nos *templates* para o reconhecimento(BRUNELLI; POGGIO, 1993).

Esta técnica pode ser aplicada em dois tipos de *templates* para cada imagem de face (KARUNGARU; FUKUMI; AKAMATSU, 2004). A Fig. 2.15 ilustra dois *templates* de face com bordas e informação de cor no espaço YIQ. As características extraídas foram sobre os olhos, lábios e bordas da imagem da face. Os resultados obtidos na taxa de reconhecimento foi de 95,1%, utilizando a base de dados *Oulu*[27].

Brunelli e Poggio (1993) aplicaram a técnica de *template matching* em quatro regiões da face (olhos, nariz, boca e rosto) a partir de imagens de faces na posição frontal. Este *matching* foi utilizado com a medida de correlação em cada região da face. Em (BRUNELLI; POGGIO, 1993) foi comparado o desempenho da técnica *template matching* com da técnica *geometrical feature matching*, onde o *template matching* obteve um desempenho superior sobre a base de dados criada pelos próprios autores com um total de 47 indivíduos, sendo 4 imagens para cada indivíduo.

A técnica de *template matching* é utilizado para associar características de uma face ou de várias regiões desta face em uma imagem. Existem algumas questões para serem analisadas no processo de correspondência entre imagens, como por exemplo: variações de iluminação e de pose, largura e altura da face na imagem e o número de padrões a serem comparados para a tarefa de reconhecimento.

Figura 2.15 Exemplos de dois *templates* em uma face: a) bordas; b) cor. Imagem retirada de (KARUNGARU; FUKUMI; AKAMATSU, 2004).

[27]http://www.cse.oulu.fi/CMV/Downloads/Pbfd

2.3.7 Técnicas baseadas em 3D Morphable Model

Esta técnica é baseada num espaço vetorial para representação de imagens da face (VETTER; POGGIO, 1997). Essa representação é realizada de tal maneira que a combinação convexa dos vetores de forma e textura de um conjunto de imagens descreva um rosto humano. Essa técnica tem sido usualmente utilizada para o reconhecimento de face sob diferentes condições de iluminação e pose das faces analisadas.

Um método para reconhecimento de faces por meio de *3D Morphable Model* com a utilização da computação gráfica para projeção e iluminação foram propostas em (BLANZ; VETTER, 2003); os autores mostraram um algoritmo de *fitting* que estima o cenário 3D com alguns parâmetros, tais como: posição da cabeça e orientação, comprimento focal da câmera e direção da iluminação. Com esses parâmetros obtidos, o algoritmo de *fitting* foi aplicado para otimizar a reconstrução das imagens de faces no espaço tridimensional em relação a forma 3D e textura. A Fig. 2.16 mostra o *fitting* do *3D Morphable Model* nas imagens de faces para reconhecimento; essas imagens cadastradas e consultadas são analisadas por um algoritmo de *fitting* e seus coeficientes de forma α_i e textura β_i são armazenados para reconhecimento. Os resultados obtidos na taxa de identificação de faces com duas bases de dados, *CMU-PIE*[28] e *Feret*[29], foi de 95% e 95,9% respectivamente.

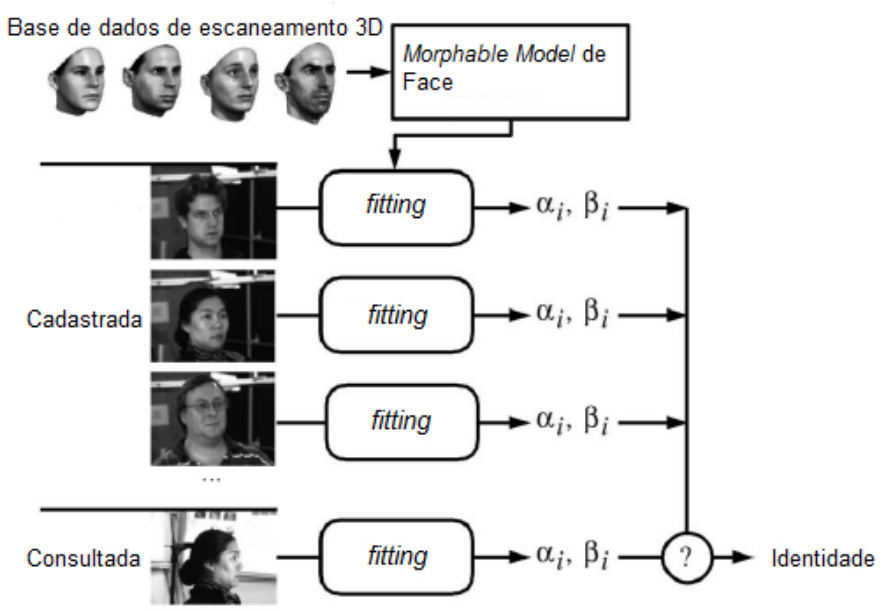

Figura 2.16 As bases de dados de imagens de faces digitalizadas em 3D são transmitidas no *3D Morphable Model* para codificar as imagens cadastradas e a imagem a ser consultada para identificação. Os coeficientes α_i, β_i do modelo da imagem consultada são comparados com os coeficientes de todas as imagens da base de dados armazenadas. Imagem adaptada de (BLANZ; VETTER, 2003).

[28]http://vasc.ri.cmu.edu/idb/html/face/
[29]http://www.nist.gov/itl/iad/ig/colorferet.cfm

Choi et al. (2008) propuseram um método invariante a pose com imagens de faces utilizando *3D morphable model* e RNA para reconhecimento. O método proposto utilizou *3D morphable model* para obter a reconstrução da face no espaço em 3D. Com a face reconstruída pôde-se extrair regiões (olhos, nariz e boca) da imagem de face no espaço em 2D a partir da face em 3D sob variações de pose. As regiões extraídas das imagens de face em 2D são utilizadas para treinar a RNA. O método proposto alcançou uma taxa de reconhecimento maior que 98% sob uma base de dados *BJUT 3D scan*[30], que consiste de 1250 imagens de 50 indivíduos. Weyrauch et al. (2004) apresentaram um método para reconhecimento de faces invariante a pose e a iluminação. Este método foi usado para computar os modelos de face no espaço em 3D a partir de três imagens de entrada de cada indivíduo. Os modelos 3D são renderizados sob variações de pose e mudanças nas condições de iluminação para construir um conjunto de imagens sintéticas para treinamento. O método atingiu uma taxa de reconhecimento de 88% em uma base de dados criada pelos próprios autores, que consiste de 2000 imagens de 10 indivíduos.

O *3D morphable model* é utilizado para representação da face no espaço em 3D a partir de imagens de faces no espaço em 2D. Esta representação possui características de formas e textura na imagem de face em 3D. Existem algumas questões para serem avaliadas sobre o *3D morphable model*, tais como: modelo de construção para o espaço tridimensional, algoritmo de *fitting* aplicado sob as imagens de faces em 2D e diferentes poses e oclusão da região da face.

2.4 CONSIDERAÇÕES FINAIS

No presente capítulo, foi apresentada a teoria geral sobre o reconhecimento facial em imagens, bem como uma breve introdução da área iniciadas entre as décadas de 60 e 70. Depois, foram discutidas as etapas comumente utilizadas em sistemas de reconhecimento facial; tais etapas são realizadas de forma sequencial, por exemplo: localização da face, normalização da face, extração de características e associação das características extraídas entre duas faces.

Foram apresentadas também as técnicas existentes na literatura sobre o tema na Seção 2.3, onde foram discutidas as principais técnicas abordadas por (TOLBA; EL-BAZ; EL-HARBY, 2006). Existem alguns pontos a serem analisados por meio dessas técnicas, por exemplo: a qualidade da imagem, as diferentes condições de iluminação e poses da região da face e a cooperação ou não do usuário no reconhecimento de faces em tempo real.

No próximo capítulo, será apresentado um estudo sistemático de revisão de literatura para a detecção de impostor facial nos sistemas de reconhecimento de faces, bem como os tipos de ataques e métodos utilizados nas aplicações de falsificação de face. Além disso, serão discutidos os métodos existentes em uma linha cronológica de impostor facial.

[30]http://www.bjpu.edu.cn/sci/multimedia/mul-lab/3dface/facedatabase.htm

Capítulo

3

DETECÇÃO DE IMPOSTOR FACIAL

Com o advento dos sistemas de reconhecimento facial, vários métodos foram elaborados para forjar tais sistemas a partir de fotos, vídeos ou máscaras. Uma nova linha de pesquisa foi criada com o objetivo de identificar os diversos tipos de ataques e tornar os sistemas de reconhecimento mais robustos.

Este capítulo representa o núcleo do estudo sistemático realizado a partir dos diversos métodos presentes na literatura. O presente capítulo trata inicialmente dos aspectos gerais das falsificações de faces, endereçando os possíveis tipos de ataques que podem ocorrer em sistemas de reconhecimento facial. Em seguida, é realizada uma descrição dos trabalhos mais relevantes na literatura nos últimos oito anos, dividindo a análise de cada método proposto a partir dos descritores e classificadores que cada um utiliza. Por fim, a evolução temporal dos trabalhos sobre técnicas de detecção de impostores faciais é apresentada a fim de apontar não somente as tendências relativas às técnicas propostas ao longo dos anos de pesquisa na área, mas também identificar perspectivas futuras.

3.1 ASPECTOS GERAIS DAS FALSIFICAÇÕES DE FACES

A princípio, detectar a face antes de detectar suas características peculiares, diminui o tempo de processamento, uma vez que muitos algoritmos se baseiam numa região de interesse da imagem. O proveito de se detectar a face, em primeiro instante, é que após esta etapa a procura pelas características faciais como olhos, nariz e boca ficarão limitadas apenas a uma determinada região da imagem. Com este conhecimento é possível estabelecer regras que diferencie uma face genuína de faces reproduzidas por impostores

Diante das informações expostas, a maioria dos trabalhos de detecção de impostor facial, foram voltados para a região da face recortada em relação a toda a imagem capturada do cenário. Essa detecção utilizam algumas técnicas de falsificação por meio de pessoas maliciosas para obterem permissão aos sistemas de reconhecimento de faces, por exemplo: ataques de impostores ocorrem quando uma pessoa tenta se passar por alguém para ter permissão para acessar um sistema de reconhecimento facial. Sendo assim, esses sistemas

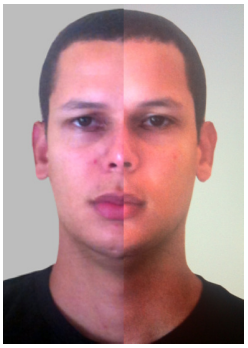

Figura 3.1 Exemplo de uma imagem da face: metade é real, a outra é falsa. Qual é a metade real ou falsa?

devem ser robustos contra tentativas de falsificação de faces por meio de fotografias ou vídeos, nos quais são duas formas usuais, de burlar, por uma pessoa maliciosa.

Como os procedimentos para replicar faces humanas são muito comuns hoje em dia (por exemplo, fotografia, gravação de vídeo e impressão 3D), a detecção de falsificação torna-se mandatória em qualquer sistema de reconhecimento facial. A Figura 3.1 ilustra a complexidade deste problema, e a seguinte pergunta pode emergir: "Qual é a metade real ou falsa?". É uma tarefa que pode ser difícil tanto para os seres humanos, quanto para sistemas baseados em técnicas de reconhecimento de padrões em imagens.

3.2 DETECÇÃO DE IMPOSTOR FACIAL

Detecção de falsificação facial (MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (SCHWARTZ; ROCHA; EDRINI, 2011a), (BHARADWAJ et al., 2013), (TIRUNAGARI et al., 2015), **detecção de vivacidade de face** (YAN et al., 2012), (PEIXOTO; MICHELASSI; ROCHA, 2011), (YANG et al., 2013), (WANG et al., 2013), (TAN et al., 2010), **medidas contra ataques de falsificação de face** (KOMULAINEN et al., 2013b), (PEREIRA et al., 2013), (KOSE; DUGELAY, 2013c), (KOSE; DUGELAY, 2013a), (KOSE; DUGELAY, 2013b) e **anti-falsificação de face** (CHINGOVSKA; ANJOS; MARCEL, 2012), (ERDOGMUS; MARCEL, 2013), (GALBALLY; MARCEL, 2014) são termos indistintamente utilizados para designar métodos para identificar um impostor utilizando disfarces faciais em sistemas de reconhecimento facial. Esses sistemas geralmente consideram os seguintes tipos de ataques de falsificação:

- A utilização de **foto impressa plana** é a mais comum, com grande potencial para acontecer, uma vez que a maioria das pessoas tem imagens faciais disponíveis na internet (por exemplo, mídia social) ou poderia ser fotografada por um impostor sem a colaboração ou permissão (um exemplo deste ataque pode ser visto na Fig. 3.2(b)).

- No ataque de **foto recortada nos olhos**, as regiões oculares de uma foto impressa são cortadas para exibir ações de olhos fechados e abertos do impostor (um exemplo

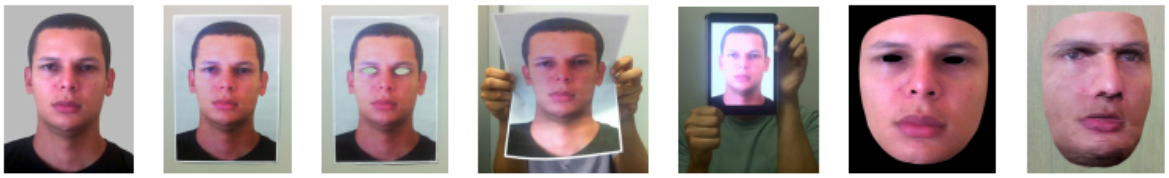

Figura 3.2 Tipos de ataques de falsificação: (a) usuário genuíno; (b) foto impressa plana; (c) foto recortada nos olhos; (d) foto distorcida, (e) reprodução de vídeo; (f) máscara usável no tamanho do rosto e (g) máscara cortada no papel.

deste ataque pode ser visto na Fig. 3.2(c)).

- Ataques de **foto distorcida** consiste em dobrar uma foto impressa em qualquer direção para simular um movimento facial (um exemplo deste ataque pode ser visto na Fig. 3.2(d)).

- Um ataque via **reprodução de vídeo** mostra quase todos os comportamentos semelhantes sobre as faces reais, com muitas características intrínsecas dos movimentos de um usuário genuíno. Este tipo de ataque tem sinais fisiológicos da vida que não são apresentados em fotos, como piscar de olhos, expressões faciais e movimentos da cabeça e boca, e inclusive pode ser facilmente realizada por meio de *tablets* ou *smartphones* grandes (um exemplo deste ataque pode ser visto na Fig. 3.2(e)).

- Ataques de **máscara** são dois tipos: máscara usável no tamanho do rosto (um exemplo deste ataque pode ser visto na Fig. 3.2(f)) e máscara cortada no papel (um exemplo deste ataque pode ser visto na Fig. 3.2(g)). Estes ataques são formados em uma estrutura facial 3D, e é um dos mais complexos ataques a serem detectados. A fabricação da máscara é um processo muito mais difícil e dispendioso de que os outros tipos de ataques, que requer dispositivos de escaneamento e impressão em 3D.

Os trabalhos analisados e categorizados foram organizados em termos de seus principais componentes, como descritores e classificadores. Descritores foram categorizados como **textura, movimento, frequência, cor, forma e reflectância**, enquanto os classificadores como **discriminante, regressão, métrica de distância e heurística**, conforme ilustrado na Tabela 3.1 que apresenta uma taxonomia dos trabalhos mais significativos na literatura sobre detecção de impostor facial. Esta taxonomia foi concebida para ajudar a compreender melhor os processos por trás de cada contramedida e para tentar encontrar as tendências gerais para diferentes tipos de ataques. As contramedidas são métodos utilizados para evitar que pessoas maliciosas com algum tipo de ataque de falsificação consigam adquirir acesso aos sistemas de reconhecimento de facial.

Tabela 3.1 Trabalhos na literatura sobre detecção de impostor facial

Descritores	**Trabalhos relacionados**
Textura	**LBP e variações** ((KOMULAINEN et al., 2013b), (CHINGOVSKA; ANJOS; MARCEL, 2012), (ERDOGMUS; MARCEL, 2013), (KIM et al., 2012), (MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (YANG et al., 2013), (KOSE; DUGELAY, 2013a), (KOSE; DUGELAY, 2014), (KOSE; DUGELAY, 2013c), (PEREIRA et al., 2013), (KOSE; DUGELAY, 2012) e Equipes IDIAP, UOULU (CHAKKA et al., 2011) e CASIA, MaskDown, LNMIIT, Muvis (CHINGOVSKA et al., 2013)), **Gabor Wavelets** ((MÄÄTTÄ; HADID; PIETIKÄINEN, 2012) e Equipe Muvis (CHINGOVSKA et al., 2013)), **GLCM** ((SCHWARTZ; ROCHA; EDRINI, 2011a) e Equipes UNICAMP (CHAKKA et al., 2011) e MaskDown, UNICAMP (CHINGOVSKA et al., 2013)), **LGS** (BASHIER et al., 2014), **ILGS** (HOUSAM et al., 2014), **LPQ** (YANG et al., 2013), **DoG** (PEIXOTO; MICHELASSI; ROCHA, 2011), (ZHANG et al., 2012), **HOG** ((MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (SCHWARTZ; ROCHA; EDRINI, 2011a), (KOMULAINEN; HADID; PIETIKAINEN, 2013a), (YANG et al., 2013) e Equipe UNICAMP (CHAKKA et al., 2011)), **HSC** ((SCHWARTZ; ROCHA; EDRINI, 2011a)) e Equipe UNICAMP (CHAKKA et al., 2011)), **CNN** (MENOTTI et al., 2015)
Movimento	**HOOF** (BHARADWAJ et al., 2013), **OFL** (KOLLREIDER; FRONTHALER; BIGUN, 2008), (KOLLREIDER; FRONTHALER; BIGUN, 2009), **Correlação de Movimento** ((KOMULAINEN et al., 2013b) e Equipe CASIA (CHINGOVSKA et al., 2013)), **GMM** ((YAN et al., 2012) e Equipes CASIA (CHAKKA et al., 2011) e LNMIIT (CHINGOVSKA et al., 2013)), **DMD** (TIRUNAGARI et al., 2015), **CRF** (PAN et al., 2007), **RASL** ((YAN et al., 2012) e Equipe CASIA (CHAKKA et al., 2011)), **HMOF** (Equipe CASIA (CHINGOVSKA et al., 2013))
Frequência	**2D-DFT** ((KIM et al., 2012), (PINTO et al., 2012), (PINTO et al., 2015) e Equipe UNICAMP (CHINGOVSKA et al., 2013)), **1D-FFT** (Equipe CASIA (CHINGOVSKA et al., 2013)), **2D-FFT** (Equipe LNMIIT (CHINGOVSKA et al., 2013)), **Haar Wavelets** ((YAN et al., 2012) e Equipe CASIA (CHAKKA et al., 2011))
Cor	**CF** (SCHWARTZ; ROCHA; EDRINI, 2011a) e Equipe UNICAMP (CHAKKA et al., 2011), **IDA** (WEN; HAN; JAIN, 2015), **IQM** ((GALBALLY; MARCEL, 2014), Equipe ATVS (CHINGOVSKA et al., 2013))
Forma	**CLM** (WANG et al., 2013)
Reflectância	***Variational Retinex*** (KOSE; DUGELAY, 2013b), (TAN et al., 2010), (KOSE; DUGELAY, 2014)
Classificadores	**Trabalhos relacionados**
Discriminante	**SVM** ((MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (KOMULAINEN et al., 2013b), (KOMULAINEN; HADID; PIETIKAINEN, 2013a), (PEREIRA et al., 2013), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (CHINGOVSKA; ANJOS; MARCEL, 2012), (KIM et al., 2012), (YANG et al., 2013), (KOSE; DUGELAY, 2013c), (ZHANG et al., 2012), (KOSE; DUGELAY, 2013a), (KOSE; DUGELAY, 2013b), (WANG et al., 2013), (KOSE; DUGELAY, 2014), (WEN; HAN; JAIN, 2015), (TIRUNAGARI et al., 2015), Equipes UOULU (CHAKKA et al., 2011) e CASIA, LNMIIT, UNICAMP (CHINGOVSKA et al., 2013)), **LDA** ((ERDOGMUS; MARCEL, 2013), (BHARADWAJ et al., 2013), (GALBALLY; MARCEL, 2014) e Equipes MaskDown, ATVS (CHINGOVSKA et al., 2013)), **MLP** (KOMULAINEN et al., 2013b), **CNN** (MENOTTI et al., 2015)
Regressão	**LLR** ((KOMULAINEN et al., 2013b), MaskDown (CHINGOVSKA et al., 2013)), **LR** ((YAN et al., 2012), Equipe CASIA (CHAKKA et al., 2011)), **SLR** (PEIXOTO; MICHELASSI; ROCHA, 2011), **SLRBLR** (TAN et al., 2010), **PLS** ((SCHWARTZ; ROCHA; EDRINI, 2011a), (PINTO et al., 2012), (PINTO et al., 2015) e Equipes UNICAMP (CHAKKA et al., 2011) e Muvis (CHINGOVSKA et al., 2013))
Métrica de Distância	**Qui-quadrado** ((KOSE; DUGELAY, 2012) e Equipe IDIAP (CHAKKA et al., 2011)), **Cosseno** (BASHIER et al., 2014), (HOUSAM et al., 2014)
Heurística	***Thresholding*** (KOLLREIDER; FRONTHALER; BIGUN, 2008), **Somatório de pesos** (KOLLREIDER; FRONTHALER; BIGUN, 2009), **Contagem de piscada** (PAN et al., 2007)

3.3 DESCRITORES

Diversos extratores de detecção de impostores em imagens são utilizados para tentar representar unicamente a face a ser analisada. A seguir serão apresentados os principais trabalhos com propostas de categorização de descritores encontrados na literatura para descrever faces nas imagens, em busca de impostores.

3.3.1 Textura

Características de textura são extraídas de imagens de faces sob o pressuposto que os rostos impressos produzem determinados padrões de textura que não existem no rosto humano. A textura é a mais forte evidência de falsificação e mais de 61% dos trabalhos relacionados na Tabela 3.1 utilizam textura isoladamente ou em combinação com outros descritores em suas contramedidas. As contramedidas são métodos desenvolvidos para proteger sistemas de reconhecimento facial contra ataques de falsificação.

Diferentes descritores de textura podem ser utilizados para detectar falsificação de faces, mas o descritor do tipo *local binary patterns* (LBP) tem sido a primeira escolha, como pode-se observar na Tabela 3.1. De fato, quase metade das trabalhos exploraram o LBP original ((KOMULAINEN et al., 2013b), (CHINGOVSKA; ANJOS; MARCEL, 2012), (ERDOGMUS; MARCEL, 2013), (KIM et al., 2012) e Equipes IDIAP (CHAKKA et al., 2011) e MaskDown (CHINGOVSKA et al., 2013)) introduzido por Ojala, Pietikäinen e Harwood (1996) ou qualquer de suas variações (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002). LBP é uma técnica de textura que analisa o padrão de tons de cinza, invariante a iluminação, que rotula todos os pixels por meio da comparação com os seus vizinhos de acordo com o valor do pixel central. Depois disso, concatena se todos os valores atribuídos em relação ao pixel central em um número binário. Por fim, os rótulos dos pixels computados são organizados em histogramas para descrever a textura, o que pode ser feito para a imagem inteira ou partes da imagem. A Fig. 3.3 mostra os passos para extração usando LBP original.

O número de vizinhos, o raio da vizinhança entre os pixels e a estratégia de codificação são todos parâmetros do LBP. O operador de textura LBP em tons de cinza pode ser utilizado em diferentes tamanhos de vizinhança, ou seja, dado um centro de pixel na imagem, um número padrão é calculado a partir da comparação do valor com os de seus

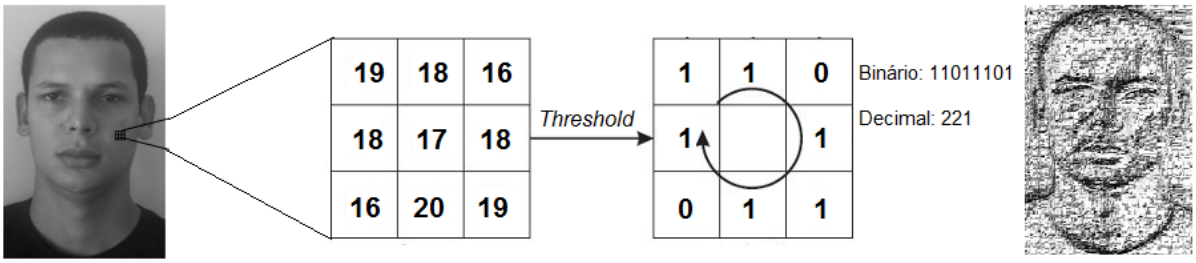

Figura 3.3 Processo de transformação do operador de análise de textura, onde é processada a imagem em tons de cinza por meio do operador LBP original.

vizinhos pela Equação 3.2. A representação do cálculo LBP é feita por:

$$LBP_{P,R} = \sum_{p=0}^{P\text{-}1} S(g_p - g_c)2^p\,, \tag{3.1}$$

$$S(x) = \begin{cases} 1, & Se \;\; x \geq 0 \\ 0, & Se \;\; x < 0 \end{cases}\,, \tag{3.2}$$

onde P, R são utilizados para designar pixels vizinhos, P o número de amostras, e R um raio de vizinhança; g_c é o valor do pixel central em nível de cinza, g_p é o valor dos vizinhos, e S designa uma função de limiar.

Uma variedade de configurações do LBP podem ser encontradas em trabalhos que envolvem a detecção de impostor facial, como: **múltiplas escalas LBP** ((MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (YANG et al., 2013), (KOSE; DUGELAY, 2013a), (KOSE; DUGELAY, 2014), (KOSE; DUGELAY, 2013c) e Equipes UOULU (CHAKKA et al., 2011) e CASIA, LNMIIT, Muvis (CHINGOVSKA et al., 2013)) que pode ser utilizado em diferentes conjuntos de vizinhança circularmente para diferentes configurações de P, R, como pode ser visto na Fig. 3.4. ***LBP variance* (LBPV)** que é invariante a rotação e foi proposto por Guo, Zhang e Zhang (2010), e no contexto de falsificação de faces foi avaliado por Kose e Dugelay (2012) que combina informações de textura e contraste a partir de imagens de faces com diferentes condições de iluminação; ***LBP from three orthogonal planes* (LBP-TOP)** ((PEREIRA et al., 2013) e Equipe MaskDown (CHINGOVSKA et al., 2013)) pode ser considerado um descritor híbrido com informações de textura e de movimento, uma vez que combinam ambas informações espaciais e temporais dentro de um único descritor; **LBP-TOP** consiste de três planos ortogonais que se cruzam no centro de um pixel na direção de XY (LBP normal), XT e YT, onde T é o eixo de tempo (a sequência dos frames) (PEREIRA et al., 2013). A Fig. 3.5 mostra os três planos ortogonais que cruzam cada pixel em uma sequência de *frames*, e três diferentes histogramas são gerados, e em seguida concatenados.

Outras técnicas de codificação de textura foram exploradas para detecção de falsificação de faces, incluindo a *local phase quantization* (LPQ) (YANG et al., 2013) que uti-

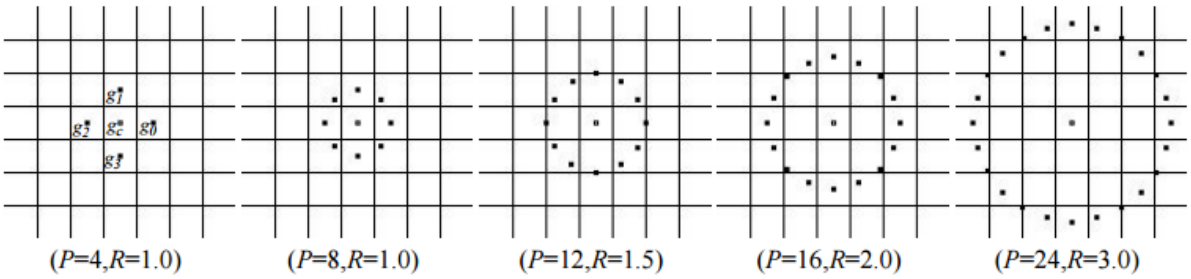

Figura 3.4 Conjuntos de vizinhança simétrico circularmente para vários P, R. Imagem retirada de (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002).

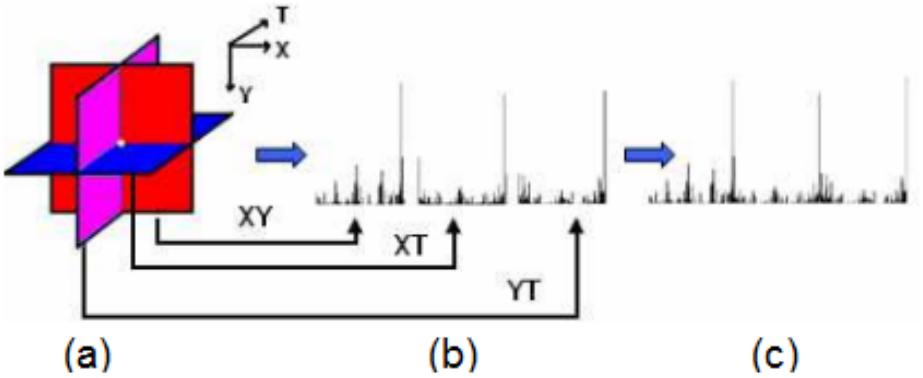

Figura 3.5 LBP-TOP computado com os seus respectivos histogramas. (a) Três planos que se intersectam de um pixel; (b) Histograma LBP de cada plano; (c) Histogramas de características concatenandos. Imagem retirada de (PEREIRA et al., 2013).

liza propriedades invariantes à desfocagem na extração das características das imagens. Este descritor utiliza informações de fase do espectro de Fourier calculado localmente para cada posição do pixel na imagem. Uma diferente técnica, a *local graph structure* (LGS) ((BASHIER et al., 2014), (HOUSAM et al., 2014)), foi utilizada para extrair características de textura por meio de comparação de um pixel alvo e os seus pixels vizinhos. O LGS é empregado com 6 pixels para formar a vizinhança do pixel alvo I(x,y). Inicialmente, o pixel alvo é comparado aos vizinhos para a região do lado esquerdo no sentido anti-horário do grafo. Se o pixel vizinho tem um valor de tom de cinza alto ou igual ao pixel alvo, logo é atribuído a 1 na extremidade que conecta os dois vértices, caso contrário atribuir 0. Em seguida é processado a região da direita, da mesma forma que da região da esquerda. A única diferença é a necessidade de avançar primeiro na horizontal e, logo, continuar o processo no sentido horário, como pode ser visto na Fig. 3.6.

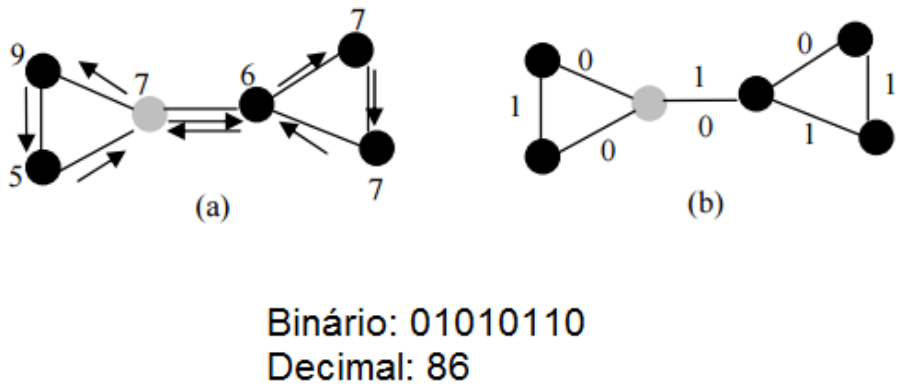

Figura 3.6 Um conjunto de pixel aplicado pela técnica LGS: (a) direção e (b) binário. Imagem adaptada de (BASHIER et al., 2014).

Histograms of oriented gradient (HOG) (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (SCHWARTZ; ROCHA; EDRINI, 2011a), (KOMULAINEN; HADID; PIETIKAINEN, 2013a), (YANG et al., 2013), primeiramente proposto por Dalal e Triggs (2005), representa a variação das orientações dos gradientes em diferentes partes da imagem, de um modo invariante à iluminação. A magnitude dos gradientes em diferentes orientações são computados em cada pixel e agrupadas em blocos; em seguida, os *bins* do histograma, as células e os blocos são normalizados. A Fig. 3.7 ilustra como as características do HOG são obtidas a partir de uma imagem de entrada. Primeiramente, a extração do HOG consiste na divisão da imagem em escala de cinza dadas por quatro etapas: (i) deslizar a janela de detecção sobre a imagem de gradientes, (ii) a partir da detecção, os blocos são extraídos por 2 × 2 células, (iii) cada célula pertencente ao bloco possui 8 × 8 pixels, e (iv) cada histograma de células constituído de 9 *bins* são representados por uma faixa de ângulos em um intervalo de 0 à 180 graus, e o conjunto de todos os histogramas dos gradientes de cada região da imagem compõe o vetor de característica final. De modo geral, na extração dos descritores HOG a partir de uma imagem em escala de cinza de dimensão M × N, são geradas duas matrizes de mesma dimensão: uma contendo a magnitude do gradiente, ($|G|$), e outra contendo a orientação dos pixels, (θ). Estes valores são computados a partir da derivada de (I_x, I_y) em cada pixel da imagem. Este procedimento é repetido ao passo que a janela de detecção desliza sobre toda a imagem, e $|G|$ e θ são computados de acordo com

$$|G| = \sqrt{I_{x^2} + I_{y^2}}\,, \tag{3.3}$$

$$\theta = arctan\frac{I_y}{I_x}\,. \tag{3.4}$$

Gabor Wavelets possuem uma boa capacidade de realçar bordas e saliências na imagem de face. Essa técnica é invariante a iluminação, rotação e escala. Além disso, ele é pouco afetado por imperfeição de fotografia, como mudanças de iluminação e ruído de imagem. Para extrair características de uma imagem, geralmente um conjunto de filtros de Gabor são utilizados em diferentes frequências (escalas) e orientações, como pode ser visto na Fig. 3.8. As características do Gabor foram exploradas no trabalho da Equipe Muvis na competição de falsificação de faces (CHINGOVSKA et al., 2013), onde computa a transformada de Gabor Wavelets em quatro escalas e seis orientações utilizando a média e o desvio padrão da magnitude da transformada dos coeficientes de Wavelet. Em (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), foram extraídos quarenta Gabor Wavelets de cinco escalas diferentes e oito orientações utilizando regiões 4 × 4 sem sobreposição.

Uma representação global, compacta e discriminante pode ser obtida em um descritor denominado de *gray level co-occurrence matrices* (GLCM) proposto por Haralick et al. (1973). A GLCM é uma matriz quadrada n × n, em tons de cinza de uma imagem $I(x, y)$, e geralmente é construída considerando as direções vertical, horizontal ou diagonal. Esta matriz de co-ocorrência é uma representação da vizinhança dos pixels de uma imagem, que descreve a ocorrência de pares de pixels de valores i e j, afastados por uma dada distância d, numa direção θ, sendo os pixels analisados dois a dois. A partir de uma matriz GLCM pode-se extrair informações de textura através de diferentes medidas de *Haralick*,

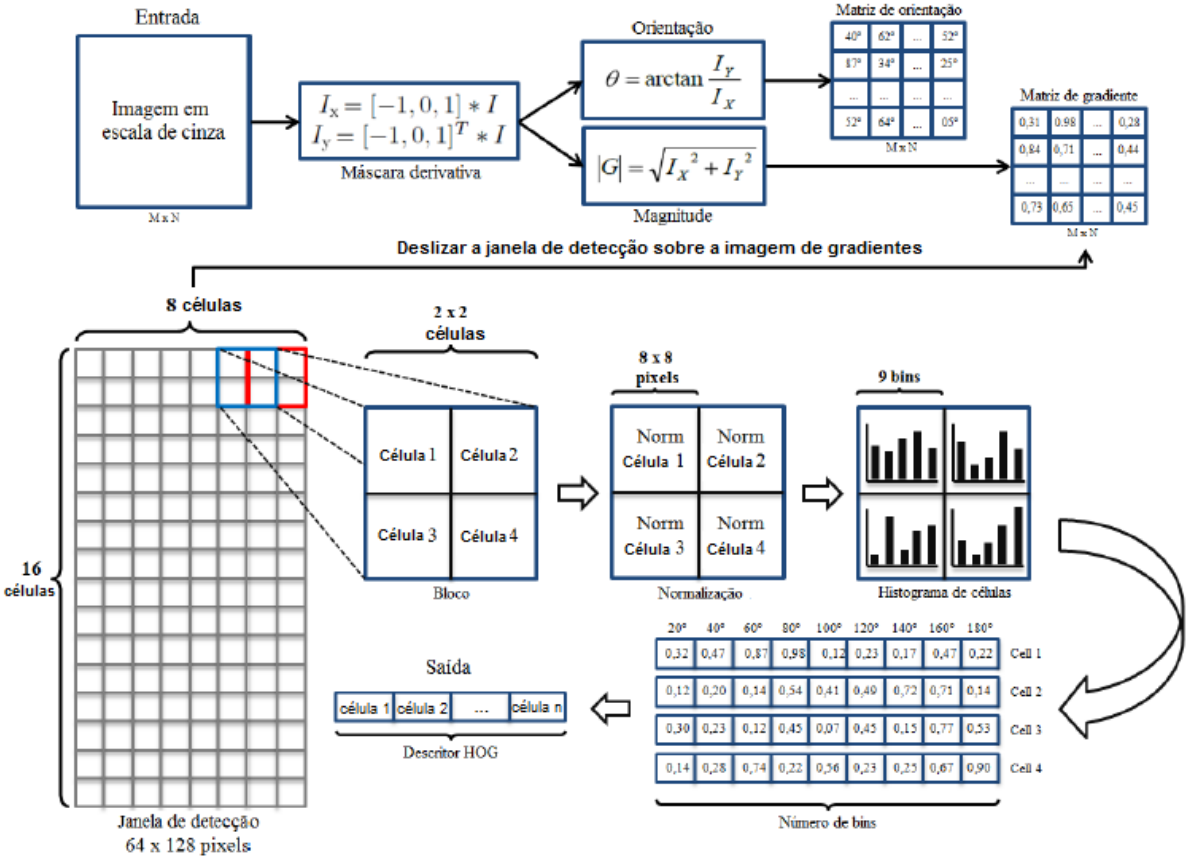

Figura 3.7 Exibição passo a passo do método do histograma de gradientes orientados (HOG). Inicialmente, a orientação e a magnitude das bordas são calculadas utilizando uma máscara centralizada [-1,0,1] em direções horizontal e vertical, sobre as imagens de entrada. Dada uma imagem em escala de cinza de dimensão M × N, são geradas duas matrizes de mesma dimensão: uma contendo a orientação dos pixels (θ) e outra contendo a magnitude do gradiente de cada pixel ($|G|$). Estes valores são calculados a partir da derivada (I_x, I_y) em cada pixel da imagem. Imagem adaptada de (OLIVEIRA et al., 2013).

por exemplo: contraste, entropia, energia, dissimilaridade, dentre outros ((SCHWARTZ; ROCHA; EDRINI, 2011a) e Equipes MaskDown, UNICAMP (CHINGOVSKA et al., 2013)). Informações de bordas podem ser exploradas para a representação de textura. A fim de descrever as bordas, a técnica de *difference of gaussians* (DoG) são utilizadas para remover as variações de iluminação, preservando os componentes de alta frequência (PEIXOTO; MICHELASSI; ROCHA, 2011) e (ZHANG et al., 2012). O objetivo é manter as altas frequências para detectar as bordas na imagem, e as informações de baixa frequência e os ruídos podem ser removidos por propriedades dos filtros Gaussianos. Um exemplo do filtro DoG aplicado nas imagens de faces genuínas e falsas pode ser visto na Fig 3.9.

Outra técnica que explora informações de bordas em objetos denominada de *histograms of shearlet coefficients* (HSC) foi proposto por Schwartz et al. (2011b), e, logo

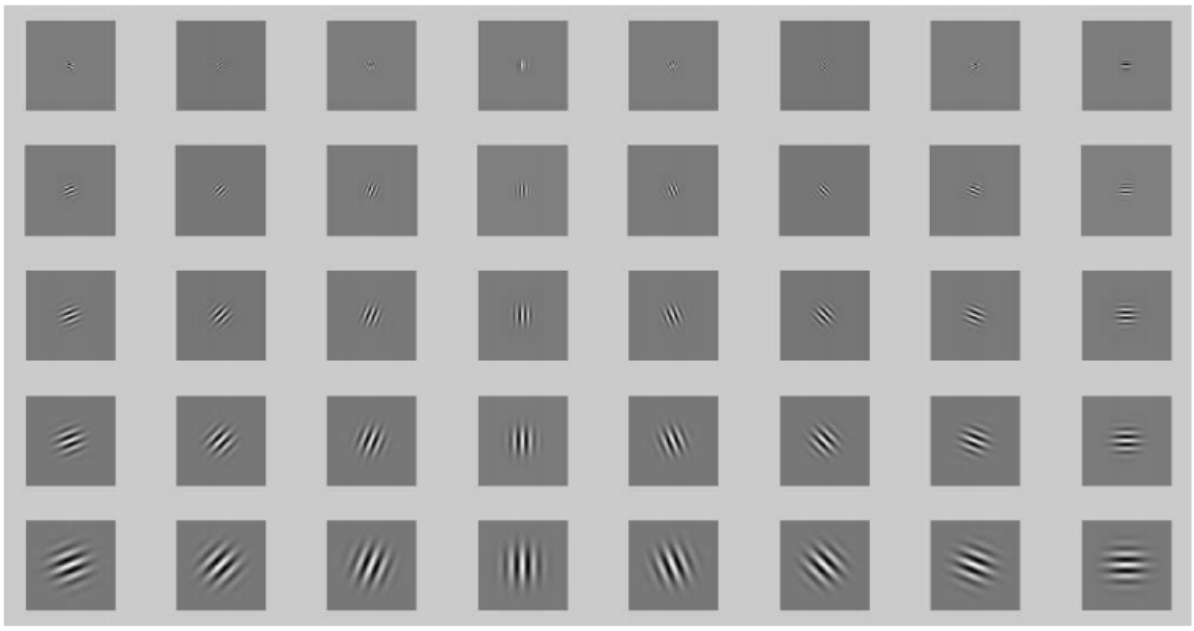

Figura 3.8 Um exemplo de 40 Gabor Wavelet com 5 escalas e 8 rotações. Imagem retirada de (SENA, 2014).

alguns trabalhos foram aplicados para detecção de impostor facial ((SCHWARTZ; ROCHA; EDRINI, 2011a) e Equipe UNICAMP (CHAKKA et al., 2011)). HSC apresenta uma decomposição multi-escala da imagem obtida pela transformada *shearlet* para extrair informações geradas pela detecção de bordas em múltiplas escalas e diferentes orientações. A transformada *shearlet* foi estabelecida em dois níveis de decomposição e oito orientações a partir da imagem facial. Como resultante, os histogramas obtidos para cada nível de decomposição são concatenados e normalizados para serem utilizados como descritor de textura (SCHWARTZ; ROCHA; EDRINI, 2011a).

Convolutional neural networks tem sido uma tendência na área de Reconhecimento de Padrões em Imagem, onde são treinadas em uma grande base de dados para fornecer

Figura 3.9 Exemplos de filtro DoG utilizado nas imagens de face. Da esquerda para a direita: imagem de face original, sua representação DoG, imagem de face falsa feita por uma foto exibida no monitor, sua representação DoG. Imagem retirada de (PEIXOTO; MICHELASSI; ROCHA, 2011).

características dinâmicas que descrevem a textura treinável a partir de imagens reais e falsas (MENOTTI et al., 2015). A CNN possui informações de múltiplas camadas para treinamentos dos dados, e pode ser observada com maior detalhes na Seção 3.4.2. Se comparada com descritores convencionais (por exemplo, de forma manual) explorados na literatura, as características aprendidas nas CNN são capazes de extrair informações mais discriminativos de forma orientada a dados.

3.3.2 Movimento

Alguns trabalhos que exploram descritores de movimento normalmente tentam detectar algumas informações presentes nas faces a fim de reconhecer um impostor, tais como: piscar de olhos, expressão facial, rotação da cabeça e movimentos dos lábios e da boca. Os descritores de movimento são o segundo tipo mais utilizado para detecção de falsificação de rostos, com mais de 25% dos trabalhos, conforme mostrado na Tabela 3.1.

Este tipo de descritor pode ser explorado a partir de duas formas diferentes para extrair as informações de movimento. A **primeira forma** foi proposta para detectar e caracterizar as variações intra-face, como piscar de olhos, expressões faciais e rotação da cabeça. Neste sentido, a técnica *conditional random fields* (CRF) foi explorada para determinar a ação de abertura e fechamento dos olhos a partir do processo estatístico do HMM e, consequentemente, detectar ataques por meio de piscar dos olhos (PAN et al., 2007). Para os movimentos faciais, *optical flow of lines* (OFL) foi utilizado para medir as variações espaço-temporal de imagens de rosto nas orientações horizontais e verticais (KOLLREIDER; FRONTHALER; BIGUN, 2008), (KOLLREIDER; FRONTHALER; BIGUN, 2009), como pode ser visto na Fig. 3.10. O *histogram of oriented optical flow* (HOOF) (BHARADWAJ et al., 2013) foi aplicado com base na orientação do gradiente das intensidades dos pixels de cada *frame* de um vídeo. Esta técnica de detecção de falsificação baseia-se nas características de movimento faciais usando *optical flow*, que calcula a magnitude e os ângulos de cada gradiente. O *histogram of magnitudes of optical flows* (HMOF) foi desenvolvida na competição de falsificação de faces (Equipe CASIA (CHINGOVSKA et al., 2013)), e são computadas a partir da técnica *optical flow* por meio da sequência dos *frames* do vídeo. Para mensurar o movimento não rígido das faces em múltiplos *frames*, foi utilizado a técnica *robust alignment sparse and low rank decomposition* (RASL) ((YAN et al., 2012), Equipe CASIA (CHAKKA et al., 2011)) para um

Figura 3.10 Exemplo de um *template* de face na posição frontal para concatenação padrão do fluxo óptico. Da esquerda para a direita: OFL horizontal; OFL vertical; magnitude da combinação do OFL. Imagem retirada de (KOLLREIDER; FRONTHALER; BIGUN, 2009).

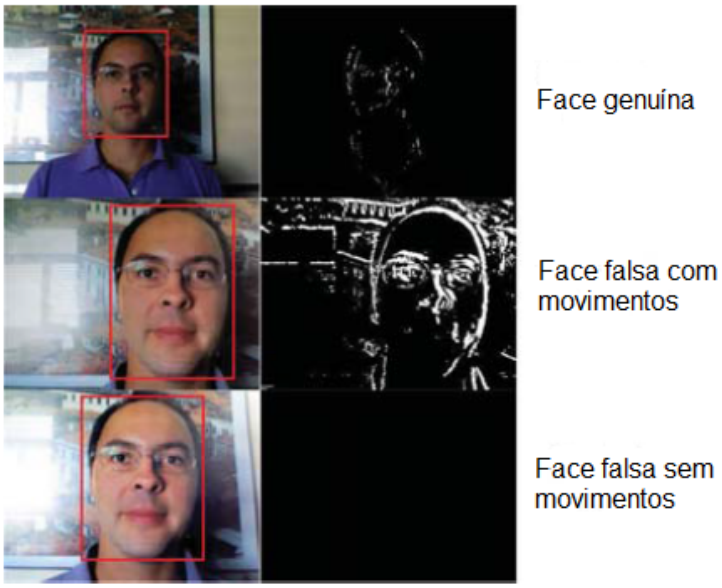

Figura 3.11 Exemplo de detecção de vivacidade entre a região da face e do plano de fundo. Da esquerda para a direita: *frame* de vídeo do cenário; a detecção do movimento do *frame*. Este processo pode ser avaliado em três formas: (i) face genuína, um *frame* de vídeo do usuário válido; (ii) face falsa com movimentos, um *frame* de vídeo de foto impressa plana segurada com as mãos do impostor e (iii) face falsa sem movimentos, um *frame* de vídeo de foto impressa plana fixada. O quadrado delimitador em vermelho corresponde à detecção automática da face. Imagem adaptada de (YAN et al., 2012).

alinhamento por lote de imagens tanto com informações da íntegra quanto parciais da face. Com a aplicação desta técnica os *frames* de vídeo ficaram mais alinhados com faces falsas. Estes resultados mostram que o RASL pode ser útil para otimizar o desempenho nos sistemas de reconhecimento facial para cenários menos ou não controlados. A **segunda forma** do descritor de movimento é avaliar a consistência da interação do usuário dentro do ambiente. Com esse propósito, a técnica de correlação de movimento foi aplicada para detectar ataques de impostores por meio de movimentos entre as regiões da face e do plano de fundo ((KOMULAINEN et al., 2013b), Equipe CASIA (CHINGOVSKA et al., 2013)), computando as regiões de interesse e normalizando os valores dos pixels a partir de múltiplos *frames* de vídeo. A técnica *gaussian mixture models* (GMM) ((YAN et al., 2012), Equipes CASIA (CHAKKA et al., 2011) e LNMIIT (CHINGOVSKA et al., 2013)) foram utilizadas para retratar o movimento no cenário com base na modelagem de subtração de plano de fundo, no qual investiga os pixels que não variam de intensidade em relação aos *frames* anteriores. A GMM pode ser representada como um somatório de pesos de múltiplas distribuições gaussianas. O trabalho de (YAN et al., 2012) apresenta a técnica GMM voltadas para detecção de movimento utilizando uma sequência de vídeo a partir de faces genuínas e duas formas de ataques são realizadas com foto impressa plana: uma contendo pequenos movimentos feitos pelo impostor segurando com as mãos a foto e a outra a foto fixada em algum ambiente, conforme ilustrado na Fig. 3.11. A

Figura 3.12 Exemplos de face genuína e ataques de falsificação em cenários controlado e adverso. Da esquerda para direita: face genuína; ataque por foto impressa plana e ataque por reprodução de vídeo a partir da imagem original da base de dados e imagem processada pela DMD. Imagem adaptada de (TIRUNAGARI et al., 2015).

textura facial de um indivíduo dentro de uma sequência de *frames* é explorada usando o *dynamic mode decomposition* (DMD) (TIRUNAGARI et al., 2015), que extrai as características por meio das *eigenfaces* (mais detalhes na Seção 2.3.1) nos *snapshots* deslocadas no espaço temporal. DMD foi utilizado em conjunto com a técnica LBP como descritor de textura, onde é aplicada para capturar indícios de presença humana numa sequência de vídeo, tais como: piscar de olhos e movimentos dos lábios. Os modos DMD foram avaliados em vídeos de faces originais e ataque de fotos e reprodução de vídeo, conforme mostrado na Fig. 3.12.

3.3.3 Frequência

Algumas contramedidas com base no descritor de frequência são eficientes na extração de características para distinguir entre uma imagem de face real e uma imagem de face reproduzida ((KIM et al., 2012), Equipes CASIA, LNMIIT, UNICAMP (CHINGOVSKA et al., 2013)). Em (LI et al., 2004) e (KIM et al., 2012) foram avaliadas as informações de frequências nas imagens de faces, nos quais os resultados mostraram que as imagens de faces falsas possuem menos componentes de alta frequência em relação as imagens de faces genuínas. No trabalho de Kim et al. (2012) foi utilizada a técnica *2D discrete Fourier transform* (2D-DFT) para extrair informações de frequência das imagens de faces. A magnitude da escala logarítmica da transformada de Fourier foi dividida em diferentes regiões de componentes de frequência. Cada região corresponde a um baixo ou alto componente de frequência. O resultado da transformada de Fourier é deslocado de modo que o componente de frequência zero encontra-se no centro do espectro, como é ilustrado na Fig. 3.13(b) a partir de uma imagem de face mostrado na Fig. 3.13(a).

Ao considerar várias imagens, o conceito de *visual rhythms* ((PINTO et al., 2012), (PINTO et al., 2015), UNICAMP (CHINGOVSKA et al., 2013)) foi utilizado para mesclar vários espectros de Fourier por meio da técnica 2D-DFT, em um único mapa que representa as informações de espaço de frequência ao longo do tempo, e, em seguida,

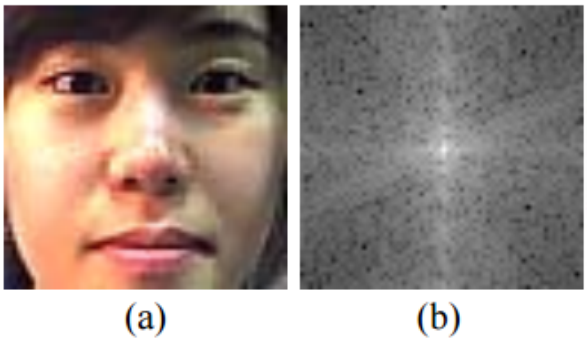

Figura 3.13 Exemplo do descritor de frequência sendo extraído por meio da 2D-DFT: (a) imagem de face genuína e (b) imagem da transformada de Fourier. Imagem retirada de (KIM et al., 2012).

as técnicas de textura são utilizadas, tais como, HOG, LBP e/ou GLCM, para a representação final da face. A Fig. 3.14 mostra um exemplo do logaritmo do espectro de Fourier aplicado em um *frame* de vídeo de uma face genuína e de uma face falsa. Pode-se observar que as Figs. 3.14(b-c) contêm as maiores respostas do sinal concentradas nos eixos da abcissa e ordenada, cuja origem está no centro do *frame*, ao contrário do logaritmo do espectro de Fourier na Fig. 3.14(a) que concentra no centro. Estas informações de frequência são relevantes para distinguir se o vídeo é real ou falso.

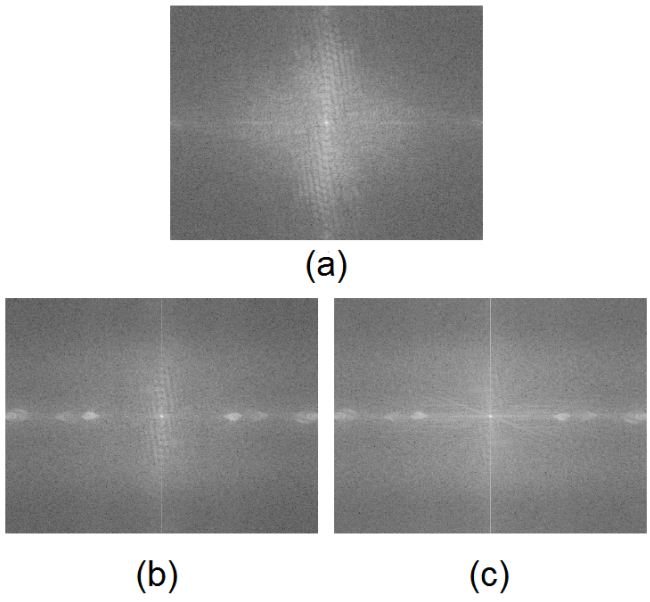

Figura 3.14 Exemplo de *frame* de vídeo do espectro de Fourier gerado a partir (a) um vídeo de usuário genuíno e (b)-(c) um vídeo de ataque considerando filtro Gaussiano e Mediano. Imagem retirada de (PINTO et al., 2012).

Color banding são mudanças bruscas causadas por impressão da foto devido a baixa qualidade ou oscilação da tela, que pode ser avaliada por meio da decomposição de *Haar Wavelets*, a fim de encontrar grandes variações unidirecionais. Os *Haar Wavelets* é um caso particular de transformada discreta de *wavelet*, onde o wavelet é um pulso quadrado variando entre 0 e 1 ((YAN et al., 2012), Equipe CASIA (CHAKKA et al., 2011)).

As técnicas *1D fast Fourier transform* (1D-FFT) (Equipe CASIA (CHINGOVSKA et al., 2013)) e *2D fast Fourier transform* (2D-FFT) (Equipe LNMIIT (CHINGOVSKA et al., 2013)) foram exploradas por meio do descritor de frequência na segunda competição para analisar ataques de falsificação de faces, onde foram extraídas os componentes de frequência do espectro de Fourier para distinguir imagens de face genuína e de face falsa.

3.3.4 Cor

Embora as cores não permaneçam constantes devido às variações de iluminação, determinadas características dominantes nas cores são consideráveis para discriminar faces genuínas e falsas. Nesse contexto, histogramas de *color frequency* (CF) descrevem a distribuição de cores em uma imagem (SCHWARTZ; ROCHA; EDRINI, 2011a); para tal, estes histogramas são calculados nas imagens por diferentes blocos, como realizado no HOG, utilizando três *bins* para codificar o número de pixels com a mais alta magnitude de gradiente em cada canal do espaço de cor RGB.

A forma de avaliar diferentes imagens de faces, em um certo instante de tempo pode ser útil para capturar sinais de vida pelas técnicas de *image distortion analysis* (IDA) (WEN; HAN; JAIN, 2015) e *image quality measures* (IQM) ((GALBALLY; MARCEL, 2014) e Equipe ATVS (CHINGOVSKA et al., 2013)). IDA foi investigada pela extração de características por meio dos espaços de cor HSV e RGB, suavização e intensidade da iluminação. No trabalho de (WEN; HAN; JAIN, 2015) foi aplicada a técnica IDA nas deformações intrínsecas das imagens, tais como: condições não controladas de iluminação no cenário e baixa qualidade da resolução. A Fig. 3.15 ilustra um exemplo de uma das características extraídas no desenvolvimento da técnica IDA chamada de reflexão especular, onde são mostradas imagens de face real e falsa antes e após do processamento da técnica. O propósito da medidas de qualidade nas imagens (IQM) é mostrar, que os

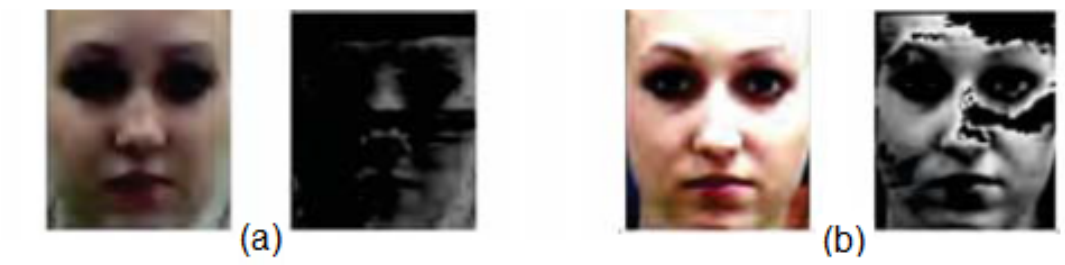

Figura 3.15 Uma ilustração das características de reflexão especular: (a) Uma imagem de face genuína e a detecção do componente de iluminação; (b) Uma face falsa reproduzida por vídeo e a detecção do componente de iluminação. Imagem retirada de (WEN; HAN; JAIN, 2015).

menores valores obtidos pelas medidas da qualidade produzido com filtragem Gaussiana são de amostras de faces falsas. Em (GALBALLY; MARCEL, 2014) e na Equipe ATVS (CHINGOVSKA et al., 2013) foram utilizadas algumas medidas de qualidade (IQM) para distinguir se a imagem de face a ser verificada é genuína ou falsa. Estas medidas foram computadas a partir do filtro passa-baixa Gaussiana nas imagens, com o objetivo de fornecer um grau de deformação das imagens de face, como correlação, bordas, diferença de pixel.

3.3.5 Forma

As informações de forma podem ser úteis para identificar um ataque de foto impressa plana, pois uma face possui características geométricas, tais como: olhos, nariz e boca.

Estas características são investigadas em uma geometria facial no espaço 3D, em vista disso fica evidente quando uma face é real ou falsa, pois os ataques de faces falsas normalmente estão no espaço bidimensional. *Constrained local models* (CLM) pode ser realizada através de contornos ao redor e dentro da face. Esta técnica é utilizada para detectar *landmarks* faciais em uma sequência de vídeo. Esses *landmarks* definem uma estrutura 3D esparsa que descreve a planaridade da face (WANG et al., 2013). Como mostrado na Fig. 3.16, as estruturas recuperadas a partir de faces autênticas usualmente contém informações suficientes no espaço em 3D, enquanto as estruturas recuperadas a partir de imagens falsas são geralmente plana em profundidade.

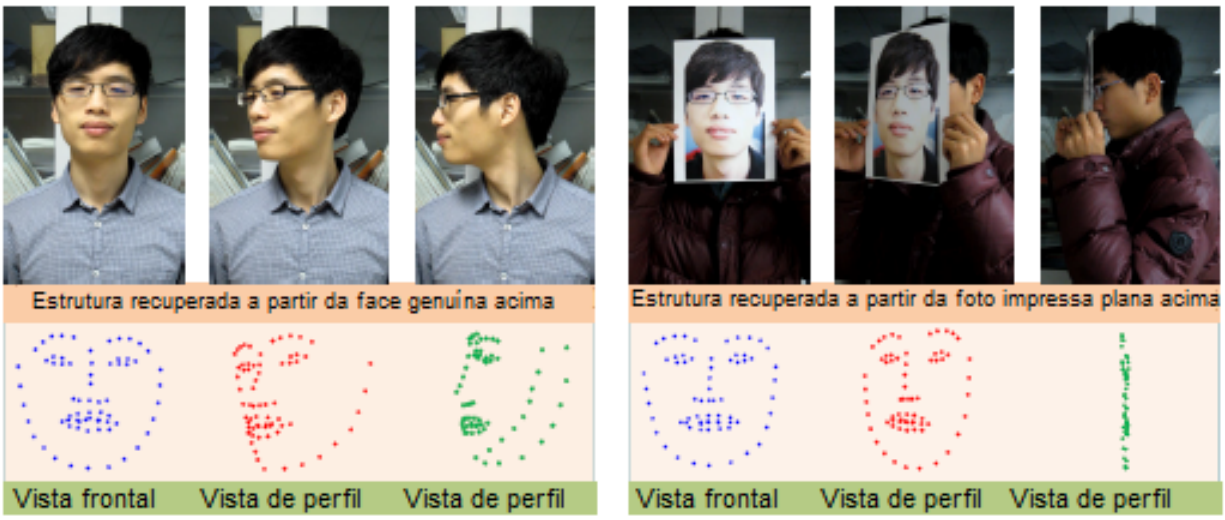

Figura 3.16 Uma comparação das estruturas de faces em 3D esparsa entre face genuína e falsa. Existem diferenças significativas nessas estruturas recuperadas, que pode ser observado pela extração de características na região da face. Imagem retirada de (WANG et al., 2013).

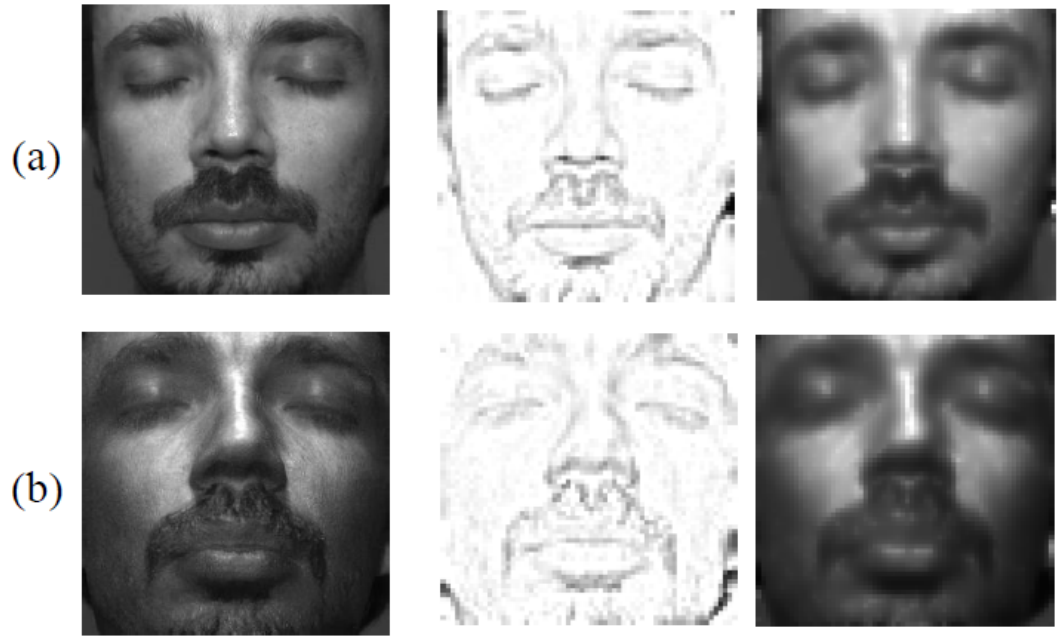

Figura 3.17 Exemplo de componentes de reflectância e iluminação do algoritmo *Retinex*. (a) uma imagem de face real com informações de textura, reflectância normalizada e iluminação; (b) uma imagem de ataque de máscara com a mesmas informações extraídas na imagem de face real. Imagem retirada de (KOSE; DUGELAY, 2013b).

3.3.6 Reflectância

Considerando que faces genuínas e falsas se comportam diferentemente nas mesmas condições de iluminação, as informações de reflectância pode ser utilizada para distinguilas. Para conseguir essa avaliação, o algoritmo *variational retinex* decompõem uma imagem de entrada em componentes de reflectância e iluminação (KOSE; DUGELAY, 2013b), (KOSE; DUGELAY, 2014), (TAN et al., 2010). A Fig. 3.17 mostra exemplos de imagens de face genuína e ataque de máscara, onde foram extraídas suas características por componentes de reflectância e iluminação utilizando o algoritmo de *variational retinex*.

3.4 CLASSIFICADORES

Diversos classificadores de detecção de impostor facial em imagens são utilizados para distinguir conjuntos de faces reais e de faces falsas. A seguir serão apresentados os trabalhos relevantes com metodologias de classificadores encontrados na literatura.

3.4.1 Discriminante

A ideia por trás do classificador discriminante baseia-se em distinguir classes de faces reais e de faces falsas. Este tipo de classificador é explorado na detecção de falsificação de faces e mais de 64% dos trabalhos explorados na Tabela 3.1 utilizam os classificadores discriminante em seus métodos.

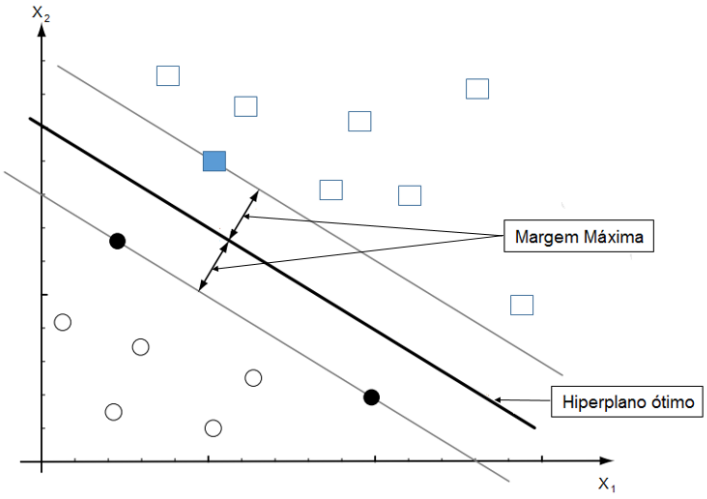

Figura 3.18 Um treinamento do SVM consiste em encontrar um hiperplano ótimo, por exemplo: aquele com a distância máxima a partir dos padrões de treinamento mais próximos. Os três vetores de suporte são os mais próximos a distância do hiperplano; tais vetores de suporte são mostrados em dois pontos (pretos) e um quadrado (azul) sólidos.

Support vector machines (SVM) é uma das técnicas mais comum de classificação na detecção de falsificação de faces, e frequentemente apresenta um desempenho superior em relação a outros tipos de classificadores (VAPNIK; VAPNIK, 1998). Para alcançar isso, SVM mapeia os elementos extraídos pelos descritores em uma dimensão maior. Isso é feito com o intuito de encontrar hiperplanos lineares para separar os descritores de faces genuínas e falsas. A Fig. 3.18 mostra um treinamento do SVM, que consiste em encontrar um hiperplano ótimo de separação entre as classes de faces reais e falsas. Vale ressaltar que o limite de decisão deve estar o mais afastado dos dados de ambas as classes. Algumas funções de kernels do SVM são utilizadas amplamente em diferentes trabalhos de ataques de falsificações de faces, como **kernel linear** ((KOMULAINEN; HADID; PIETIKAINEN, 2013a), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012), (PINTO et al., 2012), (KOSE; DUGELAY, 2013c), (KOSE; DUGELAY, 2013b), (KOSE; DUGELAY, 2013a), (KOSE; DUGELAY, 2014) e Equipe CASIA (CHINGOVSKA et al., 2013)) e **kernel não linear** que são usados quando as classes de imagens de ataques de faces e de genuínas não são linearmente separáveis. As funções de kernel não lineares são aplicadas para ampliar a precisão de classificação de faces; por exemplo o **kernel base radial** ((MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (PEREIRA et al., 2013), (CHINGOVSKA; ANJOS; MARCEL, 2012), (PINTO et al., 2012), (ERDOGMUS; MARCEL, 2013), (BHARADWAJ et al., 2013), (KIM et al., 2012), (WEN; HAN; JAIN, 2015) e a Equipe UNICAMP (CHINGOVSKA et al., 2013)) e o **histograma de kernel de interseção** (TIRUNAGARI et al., 2015). Em alguns trabalhos sobre a detecção de falsificação de faces, os autores não descrevem o tipo de kernel do SVM utilizado nos experimentos (KOMULAINEN et al., 2013b), (YANG et al., 2013), (ZHANG et al., 2012), (WANG et al., 2013).

Uma forma diferente de utilizar o classificador SVM, baseia-se na combinação com modelo estatístico HMMs (Equipe LNMIIT (CHINGOVSKA et al., 2013)). Este tipo de

classificador obteve um desempenho satisfatório na taxa de erro com o percentual de 0% na segunda competição de falsificação de faces a partir das extrações dos descritores de textura (LBP), movimento (GMM) e frequência (2D-FFT).

Como uma alternativa para abordagens lineares, o classificador *linear discriminant analysis* (LDA) ((ERDOGMUS; MARCEL, 2013), (BHARADWAJ et al., 2013), (GALBALLY; MARCEL, 2014) e Equipes MaskDown, ATVS (CHINGOVSKA et al., 2013)) é aplicado nos vetores de características geradas nas amostras de faces, que distingue as classes de faces reais e de faces falsas. Este classificador procura uma transformação linear por meio da maximização da distância entre-classes e minimização da distância intra-classe.

Uma das classes de arquitetura da RNA, a MLP foi utilizada para avaliar se o movimento excessivo (ataque de foto impressa plana segurando a mão) ou nenhum movimento (ataque de foto impressa plana fixada em um suporte) possui variações durante a sequência de N frames de vídeo (KOMULAINEN et al., 2013b). A classificação da MLP foi realizada a partir dos dados extraídos da correlação de movimento, onde obteve uma taxa de erro de 11.20% utilizando a métrica *half total error rate* (HTER).

Atualmente, as RNA por meio de aprendizagem profunda, aplica-se as CNN's em uma ampla base de dados; para tal estes dados são utilizados em duas dimensões devido ao formato da imagem para processamento das camadas, empilhadas, de convolução e de *max-pooling*. A CNN utiliza as características da MLP com compartilhamento de pesos e conexões locais entre diferentes camadas, onde todos os pesos em todas as camadas de uma rede CNN são aprendidos por meio do treinamento. Uma CNN é formada para treinar e aprender representações de invariância a escala, translação, rotação e transformações afins (LECUN; KAVUKCUOGLU; FARABET, 2010). Os elementos presentes na CNN são: camada de convolução, camada de *max-pooling* e classificador totalmente conectada. A camada de convolução é invariante a translação, que utiliza o compartilhamento de pesos; a camada de *max-pooling* tem por objetivo selecionar os mapas de características invariantes, e como resultante gera um outro mapa com resolução menor que a torna invariante a pequenas translações; ao término das camadas anteriores, os mapas de características são transformados de duas dimensões para uma dimensão e usado no treinamento no classificador totalmente conectada. Esta forma de classificação de detecção de falsificação tem sido uma tendência nos trabalhos envolvendo face (MENOTTI et al., 2015).

3.4.2 Regressão

As técnicas de classificação com base em regressão utilizam um modelo preditivo obtido a partir da extração de características por descritores nas classes de faces genuínas e falsas. Este modelo preditivo ocorre de forma supervisionada, capaz de treinar, enquanto aprende a mapear as classes de faces e a identificar padrões entre as entradas e saídas. Este tipo de classificação têm sido utilizada para detecção de falsificação de faces, nos quais diferentes técnicas de regressão são explorados na literatura, tais como: ***linear logistic regression* (LLR)** ((KOMULAINEN et al., 2013b) e Equipe MaskDown (CHINGOVSKA et al., 2013)), ***logistic regression* (LR)** ((YAN et al., 2012) e Equipe CASIA (CHAKKA et

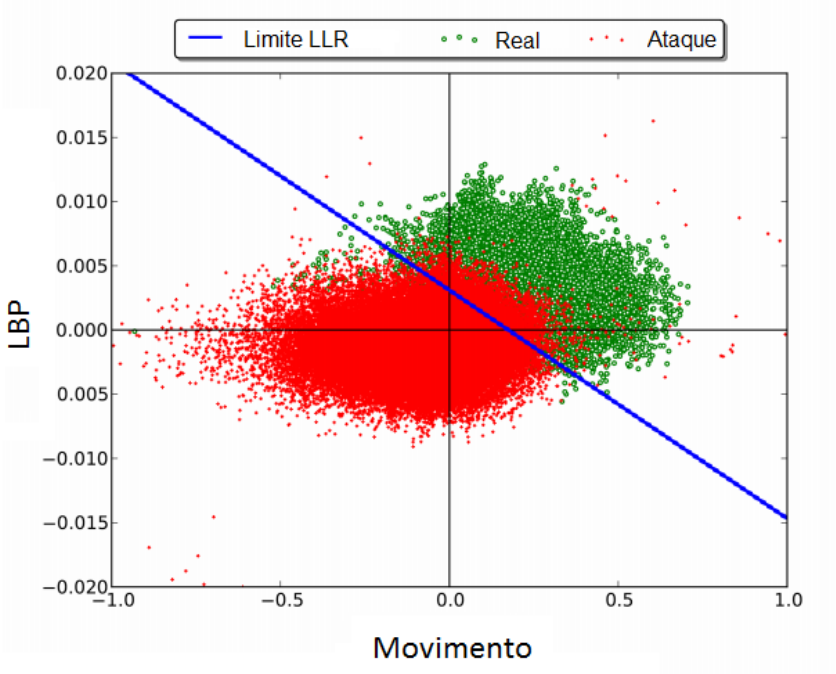

Figura 3.19 O gráfico de dispersão demonstra um desempenho satisfatório na separabilidade linear, que pode ser obtido através da combinação dos dois descritores: movimento e textura (LBP). Estes descritores, concatenados com a técnica LLR, permitem a robustez à detecção dos ataques de impostor facial. Imagem adaptada de (KOMULAINEN et al., 2013b).

al., 2011)), ***sparse logistic regression*** **(SLR)** (PEIXOTO; MICHELASSI; ROCHA, 2011), ***sparse low rank bilinear logistic regression*** **(SLRBLR)** (TAN et al., 2010), ***partial least squares*** **(PLS)** ((SCHWARTZ; ROCHA; EDRINI, 2011a), (PINTO et al., 2012), (PINTO et al., 2015) e Equipe Muvis (CHINGOVSKA et al., 2013)).

Em (KOMULAINEN et al., 2013b), foi utilizada a técnica LLR para combinação das informações extraídas por meio de dois descritores; para essa combinação foram aplicadas correlação de movimento e textura (LBP), por meio de classificadores MLP e SVM, respectivamente. O gráfico de dispersão na Fig. 3.19 descreve esses dois descritores com a utilização da técnica LLR. A técnica LLR foi elaborada na segunda competição de falsificação pela Equipe MaskDown (CHINGOVSKA et al., 2013), onde foi aplicada após a classificação por meio da técnica LDA com três descritores de textura, tais como: LBP, GLCM e LBP-TOP. Esses descritores foram computados para cada *frame* de vídeo separadamente apenas na região da face. *Logistic regression* foi desenvolvida por Yan et al. (2012) e Equipe CASIA (CHAKKA et al., 2011) para combinar informações de movimento e de frequência, com o objetivo de avaliar a detecção de impostor facial explorando três cenários: (i) **análise de movimento não rígido**, aborda a verificação dos movimentos faciais como o piscar dos olhos em uma face genuína; (ii) **consistência face e plano de fundo**, verifica indícios de movimentos faciais únicos devido ao fato de no movimento da face e do plano de fundo ocorrerem baixa consistência para faces genuínas e faces falsas e (iii) **análise de *color banding***, apresenta ruídos nas imagens

de faces falsas devido a baixa qualidade ou oscilação da tela no momento de produzir tais imagens.

Uma outra forma de classificação de regressão pode ser realizada sob as mesmas condições de iluminação para avaliar imagens de faces denominada de SLRBLR. No trabalho de Tan et al. (2010) foi explorada essa regressão em imagens com reflectância e iluminação, e aborda duas técnicas para extrair essas características da imagem, sendo a reflectância com base em *variational retinex*, e a iluminação baseada na técnica DoG na identificação de faixas de frequências médias-altas. Em seguida, a partir dos trabalhos de Tan et al. (2010) e Peixoto, Michelassi e Rocha (2011) utilizaram-se o modelo SLR para imagens filtradas com a técnica DoG. Este modelo analisa diferentes condições de iluminação e regiões de alta frequência para detecção de imagens feitas por impostores.

Alguns trabalhos foram explorados para regressão baseado em PLS ((SCHWARTZ; ROCHA; EDRINI, 2011a), (PINTO et al., 2012), (PINTO et al., 2015) e Equipe Muvis (CHINGOVSKA et al., 2013)). Este tipo de regressão, é calculado a partir da transformação linear nas características extraídas por descritores usando métodos de ponderação.

3.4.3 Métrica de distância

A métrica de distância tem a função de medir a dissemelhança entre duas amostras de faces nos sistemas de detecção de impostor facial. As métricas Qui-quadrado (χ^2) ((KOSE; DUGELAY, 2012) e Equipe IDIAP (CHAKKA et al., 2011)) e a distância do cosseno aplicada com o algoritmo de vizinho mais próximo (BASHIER et al., 2014), (HOUSAM et al., 2014) são alternativas comuns para este fim, e elas são utilizadas para computar a distância cumulativa de uma face a ser reconhecida e uma face na base de dados de faces registradas para decidir se a face é autêntica ou impostora.

3.4.4 Heurística

As técnicas de heurística são baseadas na proximidade progressiva de um certo valor, que diferencie as classes de faces; para tal, há necessidade de cálculos matemáticos com um certo limiar para a separabilidade entre classes de faces reais e falsas. Algumas abordagens nesses contexto foram propostas, como contagem de piscada dos olhos (PAN et al., 2007), que verifica a ação de abertura e fechamento dos olhos; Aplicando-se um limiar (*thresholding*) a partir dos movimentos dos olhos, onde se avalia o índice de variação dos olhos de múltiplos *frames* de vídeo (KOLLREIDER; FRONTHALER; BIGUN, 2008) e o somatório de pesos dos movimentos nas direções horizontal e vertical da região da face nos *frames* a partir do OFL, tendo como referência o frame inicial com a face centralizada (KOLLREIDER; FRONTHALER; BIGUN, 2009). Esses são alguns exemplos de heurística encontrados na literatura.

3.5 EVOLUÇÃO TEMPORAL DE IMPOSTOR FACIAL

zzA maioria dos esforços para resolver o problema da detecção de falsificação de face tem sido realizada nos últimos oito anos. A Fig. 3.20 mostra um esquema cronológico dos

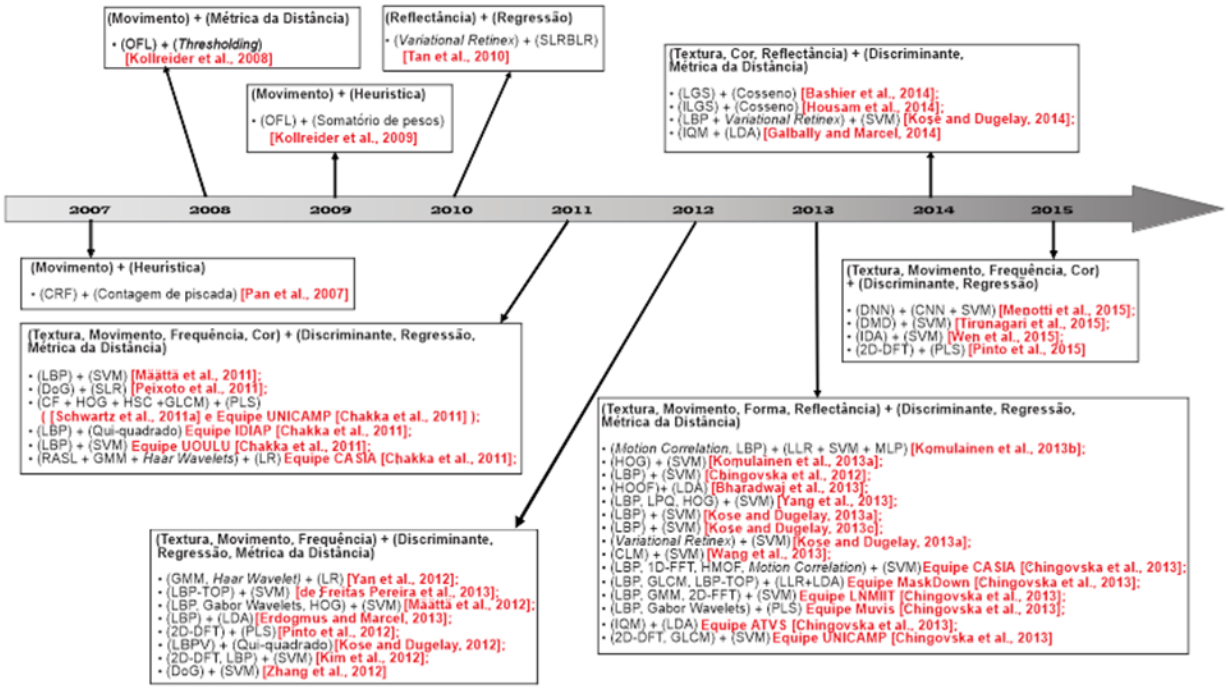

Figura 3.20 Linha do tempo dos métodos de impostor facial nos últimos 8 anos.

trabalhos mais relevantes no estado-da-arte para demonstrar a convergência de descritores e classificadores ao longo do tempo.

De 2007 a 2010, os trabalhos sobre detecção de impostor facial foram focados principalmente na análise de movimento ou reflectância, desde que ambos os tipos de descritores baseiam-se em uma observação compreensível: faces impressas não se comportam ou refletem a luz como faces reais. Embora tais medidas defensivas têm persistido até os dias atuais, outro indício presente nas imagens tem crescido substancialmente na literatura: a textura do rosto. Como apontado por Tan et al. (2010), uma face falsa utilizada por uma pessoa maliciosa é capturado por uma câmera duas vezes, enquanto uma face genuína uma vez; as primeiras pesquisas produzem artefatos para técnicas de movimento, mas sem resultados expressivos sobre o tema. Esses artefatos são muito perceptíveis em técnicas de codificação de textura, em que parecem ser uma forma eficaz para capturar e descrever faces falsas.

Em termos de classificação, no início os trabalhos foram aplicadas técnicas de heurísticas e métrica de distância, e, em seguida, as pesquisas de forma discriminante com base em SVM tornou-se cada vez mais frequente, a ponto de dominar a literatura de impostor facial nos últimos anos. Isso não é inesperado, visto que o SVM obteve uma ampla atenção em muitas outras tarefas de aprendizagem de máquina, tais como **diagnóstico médico** (SWEILAM; THARWAT; MONIEM, 2010), **reconhecimento de objetos** (MURALIDHARAN; CHANDRASEKAR, 2011) e **análise de mercado** (HUANG; NAKAMORI; WANG, 2005). Na verdade, mesmo se considerarmos apenas aplicações de processamento na região da face, existem várias maneiras de explorar o classificador SVM: **reconhecimento de face** (TEFAS; KOTROPOULOS; PITAS, 2001), **detecção de face**

(OSUNA; FREUND; GIROSI, 1997), **extração de pontos fiduciais na face** (RAPP et al., 2011), **análise da expressão facial** (KOTSIA; PITAS, 2007) e assim por diante. Apesar dos trabalhos com SVM proporcione resultados com alta precisão, uma possível orientação futura para classificação é a utilização de métodos de *deep learning*, existentes nos trabalhos de detecção de faces (ZHANG; ZHANG, 2014) e reconhecimento facial (TAIGMAN et al., 2014), bem como na detecção de impostores (MENOTTI et al., 2015).

3.6 CONSIDERAÇÕES FINAIS

Neste capítulo, foram abordadas as definições de impostor facial, apresentando os tipos de ataques que os indivíduos maliciosos utilizam para burlar os sistemas de reconhecimento facial, bem como as características de cada tipo. Além disso, foi realizado um levantamento detalhado dos trabalhos mais relevantes na literatura sobre impostor facial, abordando as técnicas que foram utilizadas nas categorias de descritores e classificadores. E, por fim, uma abordagem temporal dos métodos aplicados na falsificação de face nos últimos oito anos.

No próximo capítulo, serão detalhadas as métricas utilizadas nos trabalhos mais significativos na literatura. Em seguida, tais métricas serão avaliadas sobre algumas bases de dados de impostor facial composta por faces reais e falsas, de forma a favorecer análises comparativas. Além disso, serão apresentadas tanto uma discussão das pesquisas apontadas neste capítulo, como também algumas tendências e perspectivas futuras para detecção de impostor facial no ambiente científico e comercial.

Capítulo

4

ANÁLISE COMPARATIVA DOS MÉTODOS DE DETECÇÃO DE IMPOSTORES

A partir dos métodos avaliados sobre detecção de impostor facial, foi realizada uma análise comparativa com o objetivo de mostrar o desempenho por meio de métricas sobre as bases de dados.

4.1 MÉTRICAS AVALIADAS

Em sistemas de reconhecimento facial que abrange também a detecção de impostor facial, existem várias métricas que são comumente utilizadas para a avaliação de desempenho (VERLINDE; CHOLLET; ACHEROY, 2000), (BENGIO; MARIÉTHOZ; KELLER, 2005), (TOH; KIM; LEE, 2008), (ERDOGMUS; MARCEL, 2014). Um sistema de detecção de falsificação está sujeito a dois tipos de erro: o primeiro, um impostor pode ser aceito como um usuário genuíno (denominado, em Inglês, *number false acceptance* (NFA)); o segundo, um usuário genuíno pode ser considerado como um impostor (denominado, em Inglês, *number false rejection* (NFR)). As probabilidades desses dois erros

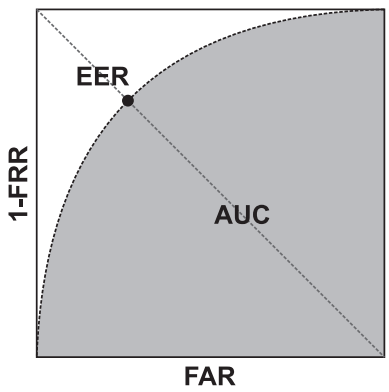

Figura 4.1 Relação entre as métricas sobre a curva ROC.

Tabela 4.1 Métricas comumente aplicadas na avaliação de falsificação de faces.

Métrica	**Significado**	**Equação**	**Tipo**
FAR	*False Acceptance Rate*	$FAR = \frac{NFA}{\#Impostor}$	Erro
FRR	*False Rejection Rate*	$FRR = \frac{NFR}{\#Genuino}$	Erro
EER	*Equal Error Rate*	$EER = (FAR = FRR)$	Erro
HTER	*Half Total Error Rate*	$HTER = \frac{FAR + FRR}{2}$	Erro
ACC	*Accuracy*	$100 \cdot \left(1 - \frac{FAR \cdot \#Impostor + FRR \cdot \#Genuino}{\#Impostor + \#Genuino}\right)$	Acerto
AUC	*Area Under Curve*	$Area = \int_a^b f(x)\,dx$, onde $f : [a,b] \rightarrow \mathbb{R}$	Acerto

ocorrerem são denominadas de *false acceptance rate* (FAR) e *false rejection rate* (FRR), respectivamente. Estas taxas são inversamente proporcionais (VERLINDE; CHOLLET; ACHEROY, 2000). Uma curva *receiver operating characteristics* (ROC) é obtida pelo cálculo de todos os possíveis pares de valores de FAR e FRR (ERDOGMUS; MARCEL, 2014), como ilustrado na Fig. 4.1. A integral de uma curva ROC é conhecida como a *area under curve* (AUC), isto é, a área cinzenta-cheia na Fig. 4.1. Além disso, o ponto da curva ROC, onde FAR e FRR são iguais, é denominado de *equal error rate* (EER) (VERLINDE; CHOLLET; ACHEROY, 2000). O ponto em que a média de FAR e FRR é mínima, chama-se *half total error rate* (HTER) (ERDOGMUS; MARCEL, 2014). Por fim, a medida de *accuracy* (ACC) considera ambos os usuários genuínos e impostores, juntamente com o FAR e FRR (ERDOGMUS; MARCEL, 2014). A Tabela 4.1 mostra o sumário das métricas usualmente aplicadas na análise de falsificação de faces.

A métrica ACC pode levar a uma análise de desempenho tendenciosa, pois a maioria das bases de dados consideradas não são balanceadas (isto é, o número de imagens de impostores e genuínas são diferentes) e maiores detalhes serão discutidas nas Seções 4.2 e 4.4. As outras métricas usadas são baseadas em uma avaliação da FAR e FRR, separadamente, por isso elas são mais confiáveis para uma análise comparativa. Por estas razões, os trabalhos pesquisados foram comparados utilizando métricas de acordo com a seguinte ordem de preferência: EER, HTER, AUC e ACC, de acordo com a utilização de cada métrica em cada trabalho analisado.

4.2 BASE DE DADOS DE IMPOSTOR FACIAL

Algumas bases de dados foram criadas nas competições de falsificações de face (ANJOS; MARCEL, 2011), (CHINGOVSKA; ANJOS; MARCEL, 2012), e, a partir dos resultados obtidos, contribuíram como referência para trabalhos existentes ou posteriores.

Apesar do interesse no estudo das fragilidades dos sistemas de reconhecimento facial, bases de dados ainda são escassas para avaliar os sistemas na prática. A explicação para este fato poderá estar no ponto de vista técnico e jurídico. Estes pontos foram abordados no estudo proposto por (GALBALLY; MARCEL; FIERREZ, 2014), onde a discussão ocorre a partir de: (i) **uma perspectiva técnica**, a aquisição dos dados relacionados a falsificação apresenta um desafio agregado para as dificuldades usuais encontradas na aquisição das bases de dados biométricas padrão, tais como: custo, tarefa demorada, necessidade de recursos humanos, cooperação do indivíduo, dentre outros. Outra situação a ser verificada é a geração de artefatos falsos (máscara, íris impressa), que, em alguns casos, são entediantes e demorados para serem produzidos; (ii) **uma perspectiva jurídica**, os dados devem estar protegidos, causando dificuldades na distribuição das bases de dados de biometria aos grupos de pesquisas e indústrias. Estas restrições legais forçam a maioria dos laboratórios a trabalhar no campo de falsificação para adquirir seus próprios conjunto de dados, geralmente pequenos, para avaliar os seus métodos de proteção. Apesar desses esforços, eles podem acarretar em limitação científica, uma vez que os resultados obtidos não podem ser comparados ou reproduzidos por outros pesquisadores.

Dentre as bases de dados de impostor facial, sete bases públicas disponíveis foram escolhidas para avaliar os métodos de falsificação de face. Para ataques 2D, ***NUAA Photograph Imposter*** (TAN et al., 2010), ***Yale Recaptured*** (PEIXOTO; MICHELASSI; ROCHA, 2011), ***Print-Attack*** (ANJOS; MARCEL, 2011), ***Replay-Attack*** (CHINGOVSKA; ANJOS; MARCEL, 2012) e ***Casia*** (ZHANG et al., 2012) são as mais conhecidas e utilizadas. As bases de dados de ***Kose e Dugelay*** (KOSE; DUGELAY, 2013c) e ***3D Mask Attack*** (ERDOGMUS; MARCEL, 2013) são as duas criadas para avaliar ataques de máscara facial. Estas bases de dados compõem a maioria dos cenários de ataque, mostrando como os ataques poderão se aprimorar. As características gerais de cada base de dados encontram-se resumidas na Tabela 4.2 e serão descritas nas seções seguintes.

4.2.1 NUAA

A base de dados *NUAA Photograph Imposter*[1] (TAN et al., 2010) foi uma das primeiras bases publicamente disponíveis para avaliação de detecção de falsificação de face. Nesta base, as imagens foram coletadas por webcams acessíveis em três sessões, em diferentes ambientes e em condições de diferentes iluminações, com um intervalo de duas semanas entre cada sessão. Estas três sessões foram divididas da seguinte forma: as duas primeiras sessões foram destinadas ao conjunto de treino, enquanto a última sessão foi para o conjunto de teste. O ataque avaliado é a fotografia impressa, que pode ser plana ou distorcida. Estes ataques de foto foram preparadas utilizando papel A4 e uma impressora

[1] $http://parnec.nuaa.edu.cn/xtan/NUAAImposterDB\_download.html$

Tabela 4.2 Visão geral das bases de dados disponíveis de falsificação de faces

Ano	Base de dados	#Indivíduo	#Real/Falsa	Tipos de ataques
2010	*NUAA Photograph Impostor* (TAN et al., 2010)	15	5105/7509	1. Foto impressa plana 2. Foto distorcida
2011	*Yale Recaptured* (PEIXOTO; MICHELASSI; ROCHA, 2011)	10	640/1920	1. Foto impressa plana
2011	*Print-Attack* (ANJOS; MARCEL, 2011)	50	200/200	1. Foto impressa plana
2012	*Replay-Attack* (CHINGOVSKA; ANJOS; MARCEL, 2012)	50	200/1000	1. Foto impressa plana 2. Reprodução de vídeo
2012	*Casia Face Anti-Spoofing* (ZHANG et al., 2012)	50	150/450	1. Foto distorcida 2. Foto recortada nos olhos 3. Reprodução de vídeo
2013	*Kose e Dugelay* (KOSE; DUGELAY, 2013c)	20	200/198	1. Máscara
2013	*3DMAD* (ERDOGMUS; MARCEL, 2013)	17	170/85	1. Máscara

colorida.

4.2.2 Yale

O principal objetivo da base de dados *Yale Recaptured*[2] (PEIXOTO; MICHELASSI; ROCHA, 2011) foi obter as imagens de impostores em 64 diferentes condições de iluminação. Desta forma, métodos baseados em textura são comumente explorados sobre esta base de dados. Imagens estáticas foram coletadas com uma distância de 50 centímetros entre o monitor LCD e a câmera.

4.2.3 Print-Attack

A base de dados *Print-Attack*[3] (ANJOS; MARCEL, 2011), foi utilizada para avaliar diferentes métodos na primeira competição de detecção de falsificação de faces (CHAKKA et al., 2011). Esta base de dados, foi criada por meio de uma foto impressa plana de um usuário genuíno para um sensor de aquisição de duas forma: a primeira é realizada pela mão, ou seja, o impostor detém a foto usando as mãos e a segunda a foto é fixada no suporte, isto é, as fotos estão presas em uma parede ou em alguma estrutura.

[2] *http* : *//ic.unicamp.br/ rocha/pub/downloads/*2011*icip/*
[3] *https* : *//www.idiap.ch/dataset/printattack/downloadproc*

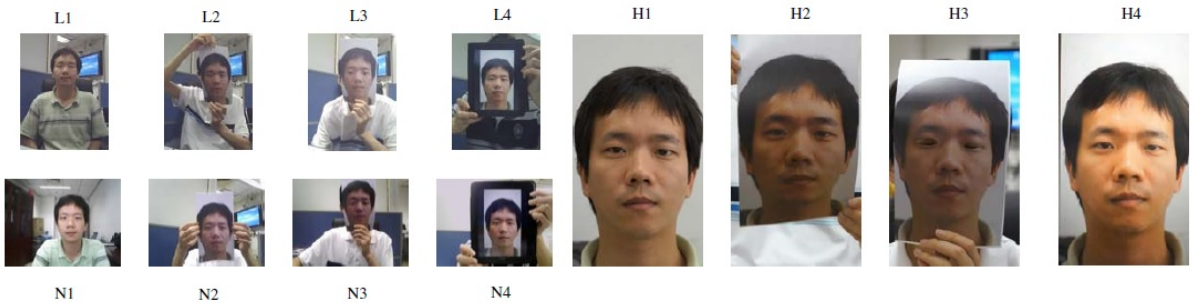

Figura 4.2 Um conjunto de vídeos completo para um indivíduo. As quatro imagens na parte superior esquerda representam os vídeos de baixa qualidade, a parte inferior esquerda são os vídeos com qualidade normal, e a parte da direita são os vídeos de alta qualidade. Para cada qualidade, da esquerda para direita são representadas nessa ordem por genuíno, ataques de foto distorcida, ataque de foto recortada nos olhos e ataque de reprodução de vídeo. Imagem retirada de (ZHANG et al., 2012).

4.2.4 Replay-Attack

A base de dados *Replay-Attack*[4] (CHINGOVSKA; ANJOS; MARCEL, 2012) é uma extensão da base *Print-Attack* para avaliar a falsificação em vídeos e fotos, e foi utilizado na segunda competição de detecção de falsificação de faces (CHINGOVSKA et al., 2013). Consiste de 1300 videoclipes de ataques de fotos e vídeos. Todas as imagens e vídeos foram coletadas sob diferentes condições de iluminação, e três modos de ataques diferentes foram considerados: foto impressa em alta resolução e reprodução de vídeos, usando uma tela de telefone móvel com baixa resolução, e um visor do iPod de 1024 x 768 pixels.

4.2.5 Casia

A base de dados *Casia Face Anti-Spoofing*[5] (ZHANG et al., 2012) contém diferentes tipos de ataques e uma variedade de qualidades de imagens classificadas como baixa, normal e alta. A Fig. 4.2 ilustra estas variedades para um indivíduo. Esta base de dados apresenta três tipos de ataques: foto distorcida, foto recortada nos olhos e reprodução de vídeo.

4.2.6 Kose e Dugelay

A base de dados colecionadas por *Kose e Dugelay* (KOSE; DUGELAY, 2013c), é uma base de dados não gratuita de máscaras de face, criado pela empresa MORPHO[6]. Os indivíduos foram capturados por um scanner 3D, que utiliza uma tecnologia de luz estruturada para obter imagens genuínas da forma e textura da face. Depois disso, as máscaras para essas imagens passam por um processo de fabricação pela Sculpteo[7] com impressão 3D, e, a seguir recapturadas pelo mesmo sensor para obter imagens de impostores.

[4] *https://www.idiap.ch/dataset/replayattack/downloadproc*
[5] *http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp*
[6] *http://www.morpho.com/*
[7] *http://www.sculpteo.com/en/*

4.2.7 3DMAD

A base de dados de ataque de máscara 3D (3DMAD)[8] (ERDOGMUS; MARCEL, 2013) foi a primeira base de dados disponível publicamente para ataques de máscara, e consiste em sequências de vídeo gravadas por uma câmera RGB-D. As máscaras foram fabricadas em dois tipos: usável no tamanho da face e cortada no papel usando os serviços de ThatsMyFace[9], e uma imagem frontal e duas imagens de perfil de cada indivíduo foram necessárias.

4.3 ANÁLISE COMPARATIVA DAS ABORDAGENS EXISTENTES NA LITERATURA

Os resultados reportados dos trabalhos foram agrupados de acordo com as bases de dados utilizadas em seus experimentos. Todos os valores numéricos neste estudo são os mesmos valores apresentados em suas obras originais, que seguiram o mesmo protocolo de avaliação.

Comparar diferentes trabalhos é uma tarefa difícil, uma vez que a maioria das vezes não temos acesso aos códigos-fonte originais e bases de dados desenvolvidas pelos autores (KIM et al., 2012); além disso, reproduzir códigos e resultados experimentais é uma tarefa complexa. Por esta razão, foi decidido realizar uma comparação usando os resultados relatados nos artigos reunidos. No entanto, na determinação do melhor método baseado nos resultados divulgados, é possível cometer erros, mesmo quando comparando trabalhos que usam a mesma base de dados, especialmente se esta base está propensa a ser tendenciosa (TORRALBA; EFROS et al., 2011). Estritamente falando, além de uma base de dados disponível comum, é de fundamental importância seguir o mesmo protocolo e ter as mesmas métricas quando diferentes contramedidas forem comparadas.

Tendo em conta as bases de dados apresentadas na Seção 4.2, alguns critérios foram adotados para selecionar quais trabalhos devem ser considerados em nossa análise: (i) deve seguir o mesmo protocolo da base de dados; (ii) deve relatar seus resultados usando pelo menos uma das métricas discutidas na Seção 4.1 e (iii) deve ser comparável a outros trabalhos usando a mesma base de dados. Por conta disso, alguns trabalhos avaliados sobre algumas bases de dados foram removidos: ***NUAA*** (CHINGOVSKA; ANJOS; MARCEL, 2012), (HOUSAM et al., 2014), (BASHIER et al., 2014), ***Print-Attack*** (YAN et al., 2012), (YANG et al., 2013), ***Casia*** (CHINGOVSKA; ANJOS; MARCEL, 2012), (TIRUNAGARI et al., 2015), (GALBALLY; MARCEL, 2014), (WEN; HAN; JAIN, 2015) e ***Kose e Dugelay*** (KOSE; DUGELAY, 2013a). As Tabelas 4.3 - 4.9 foram construídas para auxiliar a análise dos resultados selecionados. Vale ressaltar que, às vezes, foi necessário tomar conclusões indiretamente comparando métricas diferentes, como na Tabela 4.1.

A Tabela 4.3 apresenta os resultados dos métodos utilizando a base de dados *NUAA*, e as métricas mais comuns foram EER, AUC e ACC. Esta base de dados apresenta um número não equilibrado de amostras positivas e negativas, podendo ocorrer resultados

[8] $https://www.idiap.ch/dataset/3dmad/download-proc$
[9] $http://www.thatsmyface.com/Products/products.html$

Tabela 4.3 Resultados dos métodos sobre a base de dados *NUAA*

Referência	Características	Classificador	EER (%)	AUC	ACC (%)
(TAN et al., 2010)	*Variational Retinex*	SLRBLR	-	0.94	-
(MÄÄTTÄ; HADID; PIETIKÄINEN, 2011)	LBP	SVM	2.90	0.99	98.00
(PEIXOTO; MICHELASSI; ROCHA, 2011)	DoG	SLR	-	-	93.20
(SCHWARTZ; ROCHA; EDRINI, 2011a)	CF + HOG + HSC + GLCM	PLS	8.20	0.96	-
(MÄÄTTÄ; HADID; PIETIKÄINEN, 2012)	**LBP + Gabor Wavelets + HOG**	**SVM**	**1.10**	**0.99**	**-**
(KOSE; DUGELAY, 2012)	LBPV	χ^2	11.97	-	-
(YANG et al., 2013)	LBP + LPQ + HOG	SVM	1.90	0.99	97.70

Tabela 4.4 Resultados dos métodos sobre a base de dados *Yale Recaptured*

Referência	Características	Classificador	ACC (%)
(PEIXOTO; MICHELASSI; ROCHA, 2011)	DoG	SLR	91.70
(MÄÄTTÄ; HADID; PIETIKÄINEN, 2012)	**LBP + Gabor Wavelets + HOG**	**SVM**	**100.00**

tendenciosos na utilização ACC. Peixoto et al. (2011) não relataram tanto EER e AUC, mas sua ACC mostra que eles não alcançaram o melhor desempenho. Como pode ser observado na Tabela 4.3, métodos com alto valor de AUC tem baixa EER. Embora AUC não nos permite diferenciar entre os métodos propostos por (MÄÄTTÄ; HADID; PIETIKAINEN, 2011), (MÄÄTTÄ; HADID; PIETIKÄINEN, 2012) e (YANG et al., 2013), EER mostra claramente que Määttä et al. (2012) alcançaram o melhor desempenho da base de dados *NUAA*.

Um resumo dos resultados dos métodos para a base de dados *Yale Recaptured* é apresentado na Tabela 4.4. Alguns trabalhos usaram esta base de dados para comparar em termos de taxa de precisão (ACC), a única métrica em comum a todos eles. Uma vez que esta base de dados é altamente desbalanceada (por exemplo, a razão de 1:3), ACC não seria a métrica mais recomendada. No entanto, para esta comparação não é um problema por causa do desempenho mais elevado relatado por Määttä et al. (2012), com 100% de ACC, o que significa que ambas as classes foram perfeitamente classificadas.

Como indicado na Seção 4.2, a base de dados do *Print-Attack* foi usada como referência na primeira competição de detecção de falsificação (CHAKKA et al., 2011), em que três equipes alcançaram pontuação perfeita (por exemplo, IDIAP (CHAKKA et al., 2011), UOULU (CHAKKA et al., 2011) e CASIA (CHAKKA et al., 2011)). Mais tarde, as pesquisas de Määttä et al. (2012) e Tirunagari et al. (2015) também alcançaram um desempenho satisfatório de 0% de HTER (ver Tabela 4.5). Conforme mostrado na Tabela 4.2, as bases de dados: *NUAA*, *Yale Recaptured* e *Print-Attack* exploram exclusivamente os ataques de fotografia impressa. Dado que o trabalho de Määttä et al. (2012) alcançou as menores taxas de erro em relação aos demais usando a mesma abordagem, é seguro assumir que múltiplas características de textura (por exemplo, LBP, Gabor wavelets e HOG) e um classificador SVM são suficientes para detectar ataques de fotografia sobre essas bases de dados.

A base de dados *Replay-Attack* foi usada na segunda competição de detecção de falsi-

Tabela 4.5 Resultados dos métodos sobre a base de dados *Print-Attack*

Referência	Características	Classificador	HTER (%)
(**Equipe IDIAP (CHAKKA et al., 2011)**)	**LBP**	χ^2	**0.00**
(**Equipe UOULU (CHAKKA et al., 2011)**)	**LBP**	**SVM**	**0.00**
(**Equipe CASIA (CHAKKA et al., 2011)**)	**RASL + GMM + Haar Wavelets**	**LR**	**0.00**
(Equipe UNICAMP (CHAKKA et al., 2011) e (SCHWARTZ; ROCHA; EDRINI, 2011a))	CF + HOG + HSC + GLCM	PLS	0.63
(MÄÄTTÄ; HADID; PIETIKÄINEN, 2012)	**LBP + Gabor Wavelets + HOG**	**SVM**	**0.00**
(BHARADWAJ et al., 2013)	HOOF	LDA	0.62
(TIRUNAGARI et al., 2015)	**DMD**	**SVM**	**0.00**

Tabela 4.6 Resultados dos métodos sobre a base de dados *Replay-Attack*

Referência	Características	Classificador	HTER (%)
(CHINGOVSKA; ANJOS; MARCEL, 2012)	LBP	SVM	15.16
(PEREIRA et al., 2013)	LBP-TOP	SVM	7.60
(KOMULAINEN et al., 2013b)	Correlação de Movimento + LBP	LLR + SVM + MLP	5.11
(BHARADWAJ et al., 2013)	HOOF + LBP	LDA	1.25
(**Equipe CASIA (CHINGOVSKA et al., 2013)**)	**LBP + 1D-FFT + HMOF + Correlação de Movimento**	**SVM**	**0.00**
(Equipe MaskDown (CHINGOVSKA et al., 2013))	LBP + GLCM + LBP-TOP	LLR + LDA	2.50
(**Equipe LNMIIT (CHINGOVSKA et al., 2013)**)	**LBP + GMM + 2D-FFT**	**SVM**	**0.00**
(Equipe Muvis (CHINGOVSKA et al., 2013))	LBP + Gabor Wavelets	PLS	1.25
(Equipe ATVS (CHINGOVSKA et al., 2013))	IQM	LDA	12.00
(Equipe UNICAMP (CHINGOVSKA et al., 2013))	2D-DFT + GLCM	SVM	15.62
(GALBALLY; MARCEL, 2014)	IQM	LDA	15.20
(MENOTTI et al., 2015)	CNN	CNN	0.75
(TIRUNAGARI et al., 2015)	DMD	SVM	3.75
(WEN; HAN; JAIN, 2015)	IDA	SVM	7.41
(PINTO et al., 2015)	2D-DFT	PLS	14.27

ficação (CHINGOVSKA et al., 2013) e ambas as equipes, CASIA e LNMIIT, obtiveram 0% de HTER, conforme visto na Tabela 4.6. Esta base de dados tem um número desbalanceado de imagens reais e falsas (uma razão de 1:5), mas isso não influência a análise uma vez que todos os trabalhos relatam seus resultados usando a mesma métrica.

A base de dados *CASIA* é caracterizada pelo maior número de tipos de ataques, como apresentado na Tabela 4.2, mas apresenta um baixo número de amostras de faces reais e falsas. A Tabela 4.7 apresenta os resultados desta base de dados, e Zhang et al. (2012) propuseram um método com o melhor desempenho, alcançando resultados quase perfeitos (como, 0.06% EER). Em (KOMULAINEN; HADID; PIETIKAINEN, 2013a) e (YANG et al., 2013), utilizando também o classificador SVM, os autores não obtiveram um desempenho satisfatório, apresentando taxas de EER de 3.30% e 11.80%, respectivamente.

A base de dados de *Kose e Dugelay* foi utilizada para ataques de máscaras de face, com número de amostras de usuários reais e falsos aproximadamente iguais. A Tabela 4.8

Tabela 4.7 Resultados dos métodos sobre a base de dados *Casia*

Referência	Características	Classificador	EER (%)
(ZHANG et al., 2012)	**DoG**	**SVM**	**0.06**
(KOMULAINEN; HADID; PIETIKAINEN, 2013a)	HOG	SVM	3.30
(YANG et al., 2013)	LPQ + LBP + HOG	SVM	11.80

Tabela 4.8 Resultados dos métodos sobre a base de dados *Kose e Dugelay*

Referência	Características	Classificador	AUC	ACC (%)
(KOSE; DUGELAY, 2013a)	LBP	SVM	0.95	88.10
(KOSE; DUGELAY, 2013b)	*Variational Retinex*	SVM	0.97	94.47
(KOSE; DUGELAY, 2013c)	LBP	SVM	0.98	93.50
(KOSE; DUGELAY, 2014)	**LBP + *Variational Retinex***	**SVM**	**0.99**	**98.99**

Tabela 4.9 Resultados dos métodos sobre a base de dados *3DMAD*

Referência	Características	Classificador	HTER (%)
(ERDOGMUS; MARCEL, 2013)	LBP	LDA	0.95
(MENOTTI et al., 2015)	**CNN**	**CNN**	**0.00**

mostra os resultados obtidos pelos criadores desta base. O melhor desempenho sobre essa base de dados foi alcançado pelo método baseado em descritores de textura e reflectância e um classificador SVM (KOSE; DUGELAY, 2014).

A base de dados *3DMAD* possui o número de imagens falsas menor que o número de imagens reais de faces, e os métodos propostos foram avaliados com a métrica HTER. Portanto, o desbalanceamento não é um problema para a avaliação desses métodos. A Tabela 4.9 sumariza os resultados adquiridos, sendo o trabalho de Menotti et al. (2015) que obteve o melhor desempenho na taxa de erro com CNN. Métodos de lidar com reprodução de vídeo e ataques de máscara não dependem somente de descritores de textura, e podem ser explorados por diferentes características (ex., movimento, frequência e reflectância) para reduzir o erro de classificação. Percebe-se que o SVM é ainda o classificador mais utilizado pelos trabalhos levantados.

Estas sete bases de dados foram selecionadas por serem acessíveis, e a maioria gratuita, proporcionado assim uma análise comparativa entre os métodos analisados. Além disso, foram avaliados diferentes métodos com métricas de erro e acerto; para tal, em cada base de dados foi destacada o melhor resultado com suas respectivas características extraídas e classificadores.

4.4 DISCUSSÕES SOBRE OS RESULTADOS ENCONTRADOS NA LITERATURA

Percebe-se que diversos trabalhos apresentados, desde o surgimento até os dias atuais mostraram um progresso notável para detectar ataques de impostores em sistemas de reconhecimento facial. Depois de examinar todos os métodos existentes nos últimos oito anos, foi possível reconhecer as tendências atuais, perspectivas e questões abertas de como os pesquisadores estão tentando projetar suas soluções para este problema, assim como os desafios que ainda precisam ser resolvidos.

4.4.1 Tendências atuais

Conforme abordado na Seção 3.5, descritores baseados em textura (p. ex., LBP) e classificação discriminante (p. ex., SVM) têm prevalecido nos trabalhos de detecção de impostor. As técnicas de textura combinadas com o classificador SVM atingiram o melhor desempenho em cinco das sete bases de dados analisadas (*NUAA, Yale Recaptured, Print-Attack, CASIA, e 3DMAD*). Quanto às duas bases restantes (*Replay-Attack e Kose e Dugelay*), LBP e SVM ainda estão presentes, mas combinados com outros descritores (movimento, frequência ou reflectância).

A detecção automática de impostores ainda segue o mesmo *pipeline* de outros sistemas de Reconhecimento de Padrões em Imagem: extração de características e classificação das características extraídas por um classificador supervisionado. Este fato é notável ao se analisar a Tabela 4.10, onde o desempenho dos melhores trabalhos, sobre as sete bases de dados mais utilizadas na literatura, é apresentado.

Atualmente, as técnicas de *deep learning* já começam a serem exploradas no contexto de sistemas de reconhecimento facial (HUANG; LEE; LEARNED-MILLER, 2012), (SUN; WANG; TANG, 2013), (FAN et al., 2014), (ZHANG; ZHANG, 2014), (TAIGMAN et al., 2014), (GOSWAMI et al., 2014), (ZHI-PENG; YAN-NING; HAI-YAN, 2014) e, consistentemente, vêm superando outros métodos existentes no estado-da-arte. Recentemente, Menotti et al. (2015) adotaram a técnica de *deep learning* para a detecção de impostor facial, e avaliaram o desempenho do método proposto sobre duas bases de dados: *Replay-Attack* e *3DMAD*. Na primeira, os resultados foram comparáveis ao estado-da-arte, e, na segunda, o melhor resultado foi obtido. Estes resultados iniciais certamente incentivam futuras investigações nesta direção.

4.4.2 Perspectivas

Princípios de *transfer learning* (YUA et al., 2014) ainda não foram explorados até o momento para detecção de impostores faciais. Esta abordagem permite a incorporação, a qualquer momento, de novas amostras para a construção do modelo existente, tornando o modelo treinado mais flexível para a classificação de novos ataques no futuro, sem requalificação de todo o classificador.

Tabela 4.10 Desempenho dos melhores trabalhos sobre diferentes bases de dados

Referência	**Características**	**Classificador**	**Base de dados**
(MÄÄTTÄ; HADID; PIETIKÄINEN, 2012)	LBP + Gabor Wavelets + HOG	SVM	*NUAA* *Yale* *Print-Attack*
Equipe IDIAP (CHAKKA et al., 2011)	LBP	χ^2	*Print-Attack*
Equipe UOULU (CHAKKA et al., 2011)	LBP	SVM	*Print-Attack*
Equipe CASIA (CHAKKA et al., 2011)	RASL + GMM + Haar wavelets	LR	*Print-Attack*
(TIRUNAGARI et al., 2015)	DMD	SVM	*Print-Attack*
Equipe CASIA (CHINGOVSKA et al., 2013)	LBP + 1D-FFT + HMOF + Correlação de Movimento	SVM	*Replay-Attack*
Equipe LNMIIT (CHINGOVSKA et al., 2013)	LBP + GMM + 2D-FFT	SVM	*Replay-Attack*
(ZHANG et al., 2012)	DoG	SVM	*CASIA*
(KOSE; DUGELAY, 2014)	LBP + *Variational Retinex*	SVM	*Kose e Dugelay*
(MENOTTI et al., 2015)	CNN	CNN	*3D Mask Attack*

Sistemas biométricos multimodais, ou seja, aqueles baseados na combinação de diferentes características humanas (p. ex., face, íris e impressão digital), ao mesmo tempo, têm demonstrado algum indício de confiabilidade na detecção de impostores (JOHNSON; TAN; SCHUCKERS, 2010), (RODRIGUES; LING; GOVINDARAJU, 2009), (RODRIGUES; KAMAT; GOVINDARAJU, 2010), (AKHTAR et al., 2012), (BIGGIO et al., 2012). Em sistemas de detecção de impostor baseados exclusivamente em informações de faces, a inclusão da multimodalidade de descritores (p. ex., textura, profundidade e temperatura) também aumentar a eficiência do sistema (HERMOSILLA et al., 2012), (DHAMECHA et al., 2013), (ERDOGMUS; MARCEL, 2013), (MENOTTI et al., 2015).

Alguns sensores disponíveis comercialmente são capazes de capturar uma imagem tanto no espaço de cor RGB[10], quanto um mapa de profundidade[11] a um baixo custo, e podem ser usados para melhorar as contramedidas atuais e eventualmente torná-las mais confiáveis em aplicações industriais (DRYANOVSKI et al., 2012), (LITOMISKY, 2012). Outra alternativa é a utilização de câmeras térmicas[12] para o reconhecimento de faces (HERMOSILLA et al., 2012), (DHAMECHA et al., 2013). Essas câmeras foram utilizadas para capturar as imagens das faces a partir da radiação térmica trasmitida pela face, no qual consegue detectar o calor gerado pelo corpo humano (BUDDHARAJU et al., 2007). As câmeras térmicas possibilita com eficiência distinguir se uma face é genuína ou impostora, pois possui um espectro térmico que permite observar as diferentes temperaturas em torno da face, tais como: boca e olhos têm temperaturas superiores em relação ao nariz e orelhas. Visto que é complexo forjar essas temperaturas tão específicas na face. Entretanto, o alto custo para adquirir essas câmeras encarecem no sistema de detecção de impostor facial.

4.4.3 Questões abertas

Os resultados dos métodos da literatura analisados foram avaliados a partir de diferentes métricas (p. ex., ACC, AUC, HTER e EER), as quais foram computadas para cada uma das sete bases de dados disponíveis, consideradas neste trabalho. Alguns resultados obtidos a partir das métricas mencionadas são perfeitos ou quase perfeitos. Este fato não demonstra que a detecção de impostor facial é um problema resolvido, havendo muito o que ser explorado no âmbito acadêmico e comercial. Na verdade, tal fato pode indicar a falta de uma base de dados desafiadora que permita uma análise mais completa dos métodos propostos sob todos os ataques praticáveis. Bases de dados desafiadoras, tais como a *VIPER* (MA; LI; CHANG, 2014) e *Caltech-256* (GRIFFIN; HOLUB; PERONA, 2007), podem ser vistas em outros problemas de Reconhecimento de Padrões em Imagem, tais como: re-identificação de pessoas e detecção de objetos, onde os melhores resultados não estão acima de 40% na taxa de precisão. As grandes bases de dados, consideradas desafiadores, estão mais perto de cenários reais, e mais propensas a promover avanços. Além de uma grande quantidade de imagens e/ou vídeos, uma base de dados que reflita um cenário de aplicação real deve abranger todos os tipos de ataques apresentados na

[10]Do Inglês, *Red Green Blue* (RGB).
[11]http://www.xbox.com/en-US/xbox-360/accessories/kinect
[12]http://www.flircameras.com/flir_a-series_a320.htm

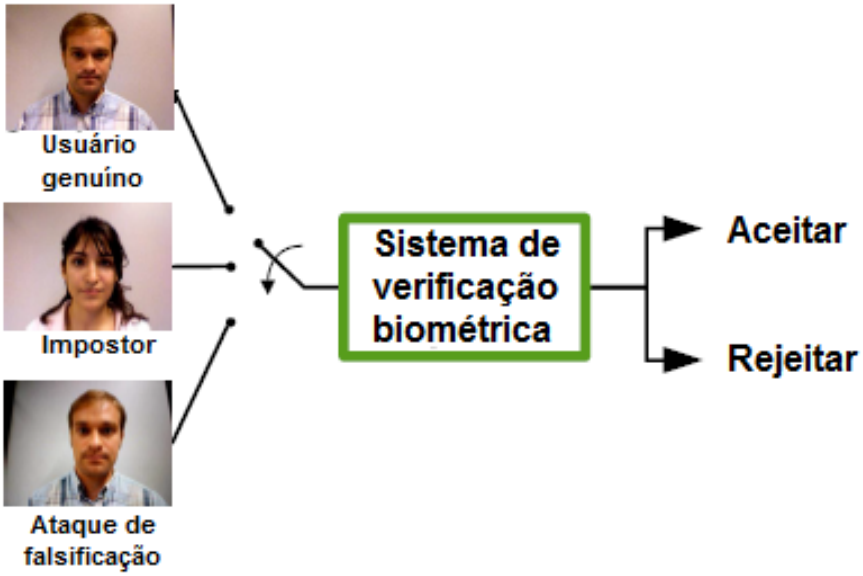

Figura 4.3 Sistema de verificação biométrica facial. Seguindo as setas da esquerda para a direita: a imagem do usuário genuíno ou impostor ou ataque de falsificação passa pelo sistema de verificação biométrica; em seguida, o resultado de cada imagem processada pelo sistema pode ser uma das duas opções: aceitar ou rejeitar. Imagem adaptada de (CHINGOVSKA; ANJOS; MARCEL, 2014).

Seção 3.1, permitindo uma análise mais factível do desempenho dos sistemas de reconhecimento de face e detecção de impostor. Por fim, a base de dados deve ser diversificada em termos de etnia, idade e sexo, apresentando cenários do mundo real com diferentes ambientes, condições de iluminação e comportamentos humanos.

A falta de um protocolo de avaliação padrão para métodos de detecção de impostor ainda é uma questão em aberto. Atualmente, a maioria dos pesquisadores utilizam duas métricas, HTER e EER, para mensurar o percentual de desempenho a fim de evitar resultados tendenciosos. Chingovska, Anjos e Marcel (2014) apresentaram um protocolo de avaliação para sistemas biométricos sobre ataques de falsificação, que analisa simultaneamente os resultados do reconhecimento e detecção de falsificação por meio de *expected performance and spoofability curves* (EPSC). Para se computar esta métrica, uma base de dados é dividida em três categorias: usuários genuínos, impostores zero esforço (a própria face do impostor sem utilizar imagem do usuário genuíno) e ataques de falsificação. A Figura 4.3 ilustra estas categorias em um sistema de verificação de faces sob os ataques de falsificação. No entanto, o método de avaliação proposto depende de uma probabilidade prévia dos ataques de falsificação, ou uma relação estimada entre a proporção de falsos positivos de impostores zero esforço e de ataques de falsificação. Assim, uma métrica de avaliação mais intuitiva e auto-explicativa também é necessária para fomentar futuros esforços de avaliações mais coerentes.

4.5 CONSIDERAÇÕES FINAIS

No presente capítulo, foi apresentada uma análise comparativa de métodos de impostores, apresentando as métricas de erro e acerto que os autores utilizaram para avaliar seus métodos propostos na detecção de falsificação de faces. Além disso, foi realizado um levantamento detalhado das bases de dados sobre o tema. E, por fim, uma abordagem comparativa dos resultados reportados nos artigos coletados na literatura.

Posteriormente, foi abordada uma discussão dos resultados apresentados, bem como algumas tendências e perspectivas para detecção de impostor facial na comunidade científica e no ambiente comercial. Em seguida, foi apontado alguns desafios atuais e futuros para avaliar os métodos de detecção de falsificação, assim como uma abordagem na criação de bases de dados robustas para avaliação.

Capítulo 5

CONCLUSÃO

A utilização de sistemas de reconhecimento facial torna-se frequente em ambientes corporativos. A disseminação de tais sistemas abre espaço para o surgimento de ataques de faces por pessoas maliciosas, capazes de burlar os sistemas de reconhecimento de faces e obter acesso às informações. Os ataques comumente utilizados baseiam-se em informações 2D, determinadas por fotos e vídeos, e são realizados por meios de câmeras, *smartphones* ou *tablets*. O processo de detecção de impostor facial na tarefa de reconhecimento de faces se faz essencial para uma segurança e confiança em tais sistemas.

Neste trabalho, realizou-se um estudo sistemático de trabalhos de detecção de impostor facial. Todo o estudo foi motivado por questionamentos de como é possível aplicar sistemas de detecção de impostor em aplicações reais. Para tal, foi analisado o problema da detecção de impostor através do levantamento dos trabalhos existentes no estado-da-arte.

Em geral, os ataques de falsificação são um desafio em termos de segurança para sistemas de reconhecimento de faces e há diferentes abordagens neste campo para encontrar métodos robustos. É comum a ênfase dos trabalhos de detecção de impostor facial em ataques 2D, por meio de apresentação de fotos impressas ou reprodução de vídeos gravados. Entretanto, os ataques 3D têm sido recentemente estudados devido aos avanços tecnológicos na reconstrução e impressão 3D. Embora alguns trabalhos tenham resultados perfeitos nas sete bases de dados avaliadas, há uma lacuna considerável em transpor as pesquisas acadêmicas para aplicações do mundo real. Posto isso, espera-se que os pesquisadores concentrem seus esforços na criação de bases de dados mais robustas e métodos de avaliação mais independentes, de agora em diante.

Os métodos aplicados na tarefa de detecção de impostor facial são vistas como um composto de descritores e classificadores, respectivamente, para extração de características e classificar a face real da falsa. A aplicação de sistemas com textura e a classificação por meio do SVM mostram os métodos favoráveis para um desempenho satisfatório tanto para as métricas que avaliam a precisão quanto para as que avaliam o erro. Essa abordagem, sinalizada no presente estudo, é capaz de apresentar os métodos mais eficientes a

partir das bases de dados, mantendo uma boa taxa de precisão. Comparando a outros estudos de detecção de impostor facial existente na literatura, o presente estudo sistemático apresenta-se mais detalhado e abrangente em análise quantitativa dos métodos, tipos de ataques, bases de dados e em questões abertas, gerando vertentes para novas metodologias sobre o tema. A principal crítica às abordagens na área de detecção de impostor facial diz respeito à falta de bases de dados complexas para avaliação computacional dos métodos propostos.

Como trabalho futuro, pretende-se investigar métodos de *deep learning* para detecção de impostor facial e uma base de dados com maior quantidade e variedade de amostras.

REFERÊNCIAS

AKHTAR, Z. et al. Evaluation of serial and parallel multibiometric systems under spoofing attacks. In: IEEE. *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*. [S.l.], 2012. p. 283–288.

AN, L.; BHANU, B.; YANG, S. Face recognition in multi-camera surveillance videos. In: IEEE. *Pattern Recognition (ICPR), 2012 21st International Conference on*. [S.l.], 2012a. p. 2885–2888.

AN, L.; KAFAI, M.; BHANU, B. Face recognition in multi-camera surveillance videos using dynamic bayesian network. In: IEEE. *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*. [S.l.], 2012b. p. 1–6.

ANJOS, A.; MARCEL, S. Counter-measures to photo attacks in face recognition: a public database and a baseline. In: IEEE. *Biometrics (IJCB), 2011 international joint conference on*. [S.l.], 2011. p. 1–7.

ANWAR, M.; IMRAN, A. A comparative study of graphical and alphanumeric passwords for mobile device authentication. p. 13–18, 2015.

AULSEBROOK, W. et al. Superimposition and reconstruction in forensic facial identification: a survey. *Forensic science international*, Elsevier, v. 75, n. 2, p. 101–120, 1995.

BASHIER, H. K. et al. Face spoofing detection using local graph structure. In: ATLANTIS PRESS. *2014 International Conference on Computer, Communications and Information Technology (CCIT 2014)*. [S.l.], 2014.

BENGIO, S.; MARIÉTHOZ, J.; KELLER, M. The expected performance curve. In: *International Conference on Machine Learning, ICML, Workshop on ROC Analysis in Machine Learning*. [S.l.: s.n.], 2005.

BHARADWAJ, S. et al. Computationally efficient face spoofing detection with motion magnification. In: IEEE. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. [S.l.], 2013. p. 105–110.

BHOWMIK, M. K. et al. Classification of polar-thermal eigenfaces using multilayer perceptron for human face recognition. In: IEEE. *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on*. [S.l.], 2008. p. 1–6.

BIGGIO, B. et al. Security evaluation of biometric authentication systems under real spoofing attacks. *IET biometrics*, IET, v. 1, n. 1, p. 11–24, 2012.

BLANZ, V.; VETTER, T. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 25, n. 9, p. 1063–1074, 2003.

BLEDSOE, W. *The model method in facial recognition*. [S.l.], 1964.

BOUGHRARA, H.; CHTOUROU, M.; AMAR, C. B. Mlp neural network based face recognition system using constructive training algorithm. In: IEEE. *Multimedia Computing and Systems (ICMCS), 2012 International Conference on*. [S.l.], 2012. p. 233–238.

BOUGHRARA, H. et al. Face recognition based on perceived facial images and multilayer perceptron neural network using constructive training algorithm. *IET Computer Vision*, IET, v. 8, n. 6, p. 729–739, 2014.

BRUNELLI, R.; POGGIO, T. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 10, p. 1042–1052, 1993.

BUDDHARAJU, P. et al. Physiology-based face recognition in the thermal infrared spectrum. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 29, n. 4, p. 613–626, 2007.

CAO, Z. et al. Face recognition with learning-based descriptor. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. [S.l.], 2010. p. 2707–2714.

CHAKKA, M. M. et al. Competition on counter measures to 2-d facial spoofing attacks. In: IEEE. *Biometrics (IJCB), 2011 International Joint Conference on*. [S.l.], 2011. p. 1–6.

CHAKRABORTY, S.; DAS, D. An overview of face liveness detection. *arXiv preprint arXiv:1405.2227*, 2014.

CHELLAPPA, R.; WILSON, C. L.; SIROHEY, S. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, IEEE, v. 83, n. 5, p. 705–741, 1995.

CHEN, Q.; YAO, J.; CHAM, W. 3d model-based pose invariant face recognition from multiple views. *IET Computer Vision*, IET, v. 1, n. 1, p. 25–34, 2007.

CHINGOVSKA, I.; ANJOS, A.; MARCEL, S. On the effectiveness of local binary patterns in face anti-spoofing. In: IEEE. *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*. [S.l.], 2012. p. 1–7.

CHINGOVSKA, I.; ANJOS, A. Rabello dos; MARCEL, S. Biometrics evaluation under spoofing attacks. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 9, n. 12, p. 2264–2276, 2014.

CHINGOVSKA, I. et al. The 2nd competition on counter measures to 2d face spoofing attacks. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013. p. 1–6.

CHOI, H.-C. et al. Pose invariant face recognition with 3d morphable model and neural network. In: IEEE. *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*. [S.l.], 2008. p. 4131–4136.

COX, I. J.; GHOSN, J.; YIANILOS, P. N. Feature-based face recognition using mixture-distance. In: IEEE. *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*. [S.l.], 1996. p. 209–216.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. [S.l.], 2005. v. 1, p. 886–893.

DHAMECHA, T. I. et al. Disguise detection and face recognition in visible and thermal spectrums. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013. p. 1–8.

DRYANOVSKI, I. et al. Real-time pose estimation with rgb-d camera. In: IEEE. *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on*. [S.l.], 2012. p. 13–20.

DUC, B.; FISCHER, S.; BIGÜN, J. Face authentication with gabor information on deformable graphs. *Image Processing, IEEE Transactions on*, IEEE, v. 8, n. 4, p. 504–516, 1999.

DUC, N. M.; MINH, B. Q. Your face is not your password face authentication bypassing lenovo–asus–toshiba. *Black Hat Briefings*, 2009.

ERDOGMUS, N.; MARCEL, S. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In: IEEE. *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. [S.l.], 2013. p. 1–6.

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

FAN, H. et al. Learning deep face representation. *arXiv preprint arXiv:1403.2802*, 2014.

GALBALLY, J.; MARCEL, S. Face anti-spoofing based on general image quality assessment. In: IEEE. *Pattern Recognition (ICPR), 2014 22nd International Conference on*. [S.l.], 2014. p. 1173–1178.

GALBALLY, J.; MARCEL, S.; FIERREZ, J. Biometric antispoofing methods: A survey in face recognition. *Access, IEEE*, IEEE, v. 2, p. 1530–1552, 2014.

GHIASS, R. S. et al. Infrared face recognition: A comprehensive review of methodologies and databases. *Pattern Recognition*, Elsevier, v. 47, n. 9, p. 2807–2824, 2014.

GORODNICHY, D.; GRANGER, E. Target-based evaluation of face recognition technology for video surveillance applications. In: IEEE. *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014 IEEE Symposium on*. [S.l.], 2014. p. 110–117.

GOSWAMI, G. et al. Mdlface: Memorability augmented deep learning for video face recognition. In: IEEE. *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. [S.l.], 2014. p. 1–7.

GRIFFIN, G.; HOLUB, A.; PERONA, P. Caltech-256 object category dataset. California Institute of Technology, 2007.

GUO, Z.; ZHANG, L.; ZHANG, D. Rotation invariant texture classification using lbp variance (lbpv) with global matching. *Pattern recognition*, Elsevier, v. 43, n. 3, p. 706–719, 2010.

GURAV, S. M. et al. Graphical password authentication: Cloud securing scheme. In: IEEE. *Electronic Systems, Signal Processing and Computing Technologies (ICESC), 2014 International Conference on*. [S.l.], 2014. p. 479–483.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 6, p. 610–621, 1973.

HAYKIN, S. S. Redes neurais artificiais: princípio e prática. *2ª Edição, Bookman, São Paulo, Brasil*, 2000.

HERMOSILLA, G. et al. A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognition*, Elsevier, v. 45, n. 7, p. 2445–2459, 2012.

HESELTINE, T.; PEARS, N.; AUSTIN, J. Evaluation of image preprocessing techniques for eigenface-based face recognition. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Second International Conference on Image and Graphics*. [S.l.], 2002. p. 677–685.

HOUSAM, K. B. et al. Face spoofing detection based on improved local graph structure. In: IEEE. *Information Science and Applications (ICISA), 2014 International Conference on*. [S.l.], 2014. p. 1–4.

HUANG, G. B.; LEE, H.; LEARNED-MILLER, E. Learning hierarchical representations for face verification with convolutional deep belief networks. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 2518–2525.

HUANG, W.; NAKAMORI, Y.; WANG, S.-Y. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, Elsevier, v. 32, n. 10, p. 2513–2522, 2005.

IBRAHIM, R.; ZIN, Z. M. Study of automated face recognition system for office door access control application. In: IEEE. *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*. [S.l.], 2011. p. 132–136.

INAN, T.; HALICI, U. 3-d face recognition with local shape descriptors. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 7, n. 2, p. 577–587, 2012.

JAIN, A. K.; KLARE, B.; PARK, U. Face recognition: Some challenges in forensics. In: IEEE. *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. [S.l.], 2011. p. 726–733.

JAIN, A. K.; LI, S. Z. *Handbook of face recognition*. [S.l.]: Springer, 2005.

JI, Y.; CHANG, K. H.; HUNG, C.-C. Efficient edge detection and object segmentation using gabor filters. In: ACM. *Proceedings of the 42nd annual Southeast regional conference*. [S.l.], 2004. p. 454–459.

JOHNSON, P.; TAN, B.; SCHUCKERS, S. Multimodal fusion vulnerability to non-zero effort (spoof) imposters. In: IEEE. *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*. [S.l.], 2010. p. 1–5.

KAMGAR-PARSI, B.; LAWSON, W.; KAMGAR-PARSI, B. Toward development of a face recognition system for watchlist surveillance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 33, n. 10, p. 1925–1937, 2011.

KANADE, T. Picture processing system by computer complex and recognition of human faces. *Doctoral dissertation, Kyoto University*, Eurographics Association, v. 3952, p. 83–97, 1973.

KARUNGARU, S.; FUKUMI, M.; AKAMATSU, N. Face recognition using genetic algorithm based template matching. In: IEEE. *Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on*. [S.l.], 2004. v. 2, p. 1252–1257.

KELLY, M. D. *Visual identification of people by computer*. [S.l.], 1970.

KHAN, I. R.; MIYAMOTO, H.; MORIE, T. Face and arm-posture recognition for secure human-machine interaction. In: IEEE. *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*. [S.l.], 2008. p. 411–417.

KIM, G. et al. Face liveness detection based on texture and frequency analyses. In: IEEE. *Biometrics (ICB), 2012 5th IAPR International Conference on*. [S.l.], 2012. p. 67–72.

KIRBY, M.; SIROVICH, L. Application of the karhunen-loeve procedure for the characterization of human faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 12, n. 1, p. 103–108, 1990.

KOLLREIDER, K.; FRONTHALER, H.; BIGUN, J. Verifying liveness by multiple experts in face biometrics. In: IEEE. *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*. [S.l.], 2008. p. 1–6.

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

KOMULAINEN, J.; HADID, A.; PIETIKAINEN, M. Context based face anti-spoofing. In: IEEE. *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. [S.l.], 2013a. p. 1–8.

KOMULAINEN, J. et al. Complementary countermeasures for detecting scenic face spoofing attacks. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013b. p. 1–7.

KOSE, N.; DUGELAY, J.-L. Classification of captured and recaptured images to detect photograph spoofing. In: IEEE. *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on*. [S.l.], 2012. p. 1027–1032.

KOSE, N.; DUGELAY, J.-L. Countermeasure for the protection of face recognition systems against mask attacks. In: IEEE. *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. [S.l.], 2013a. p. 1–6.

KOSE, N.; DUGELAY, J.-L. Reflectance analysis based countermeasure technique to detect face mask attacks. In: IEEE. *Digital Signal Processing (DSP), 2013 18th International Conference on*. [S.l.], 2013b. p. 1–6.

KOSE, N.; DUGELAY, J.-L. Shape and texture based countermeasure to protect face recognition systems against mask attacks. In: IEEE. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. [S.l.], 2013c. p. 111–116.

KOSE, N.; DUGELAY, J.-L. Mask spoofing in face recognition and countermeasures. *Image and Vision Computing*, Elsevier, v. 32, n. 10, p. 779–789, 2014.

KOTROPOULOS, C.; TEFAS, A.; PITAS, I. Morphological elastic graph matching applied to frontal face authentication under well-controlled and real conditions. *Pattern Recognition*, Elsevier, v. 33, n. 12, p. 1935–1947, 2000.

KOTROPOULOS, C. et al. Performance assessment of morphological dynamic link architecture under optimal and real operating conditions. Citeseer, 1999.

KOTSIA, I.; PITAS, I. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, IEEE, v. 16, n. 1, p. 172–187, 2007.

KUO, S.-s.; AGAZZI, O. E. Machine vision for keyword spotting using pseudo 2d hidden markov models. In: IEEE. *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*. [S.l.], 1993. v. 5, p. 81–84.

LADES, M. et al. Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, IEEE, v. 42, n. 3, p. 300–311, 1993.

LAWRENCE, S. et al. Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, IEEE, v. 8, n. 1, p. 98–113, 1997.

LE, H.-S.; LI, H. Face identification system using single hidden markov model and single sample image per person. In: IEEE. *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. [S.l.], 2004. v. 1.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: IEEE. *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*. [S.l.], 2010. p. 253–256.

LI, J. et al. Live face detection based on the analysis of fourier spectra. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Defense and Security*. [S.l.], 2004. p. 296–303.

LI, X. et al. Poster: Arranging the layout of alphanumeric buttons–the role of passwords. In: ACM. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. [S.l.], 2014. p. 1457–1459.

LIN, S.-H.; KUNG, S.-Y.; LIN, L.-J. Face recognition/detection by probabilistic decision-based neural network. *Neural Networks, IEEE Transactions on*, IEEE, v. 8, n. 1, p. 114–132, 1997.

LITOMISKY, K. Consumer rgb-d cameras and their applications. *Rapport technique, University of California*, p. 20, 2012.

LYONS, M. J.; BUDYNEK, J.; AKAMATSU, S. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 12, p. 1357–1362, 1999.

MA, B.; LI, Q.; CHANG, H. Gaussian descriptor based on local features for person re-identification. In: SPRINGER. *Computer Vision-ACCV 2014 Workshops*. [S.l.], 2014. p. 505–518.

MÄÄTTÄ, J.; HADID, A.; PIETIKAINEN, M. Face spoofing detection from single images using micro-texture analysis. In: IEEE. *Biometrics (IJCB), 2011 international joint conference on*. [S.l.], 2011. p. 1–7.

MÄÄTTÄ, J.; HADID, A.; PIETIKÄINEN, M. Face spoofing detection from single images using texture and local shape analysis. *IET biometrics*, IET, v. 1, n. 1, p. 3–10, 2012.

MEADOWCROFT, P. Card fraud–will pci-dss have the desired impact? *Card Technology Today*, Elsevier, v. 20, n. 3, p. 10–11, 2008.

MENOTTI, D. et al. Deep representations for iris, face, and fingerprint spoofing detection. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 10, n. 4, p. 864–879, 2015.

MURALIDHARAN, R.; CHANDRASEKAR, C. Object recognition using support vector machine augmented by rst invariants. *International Journal of Computer Science Issues (IJCSI)*, Citeseer, v. 8, n. 5, p. 280–286, 2011.

NEFIAN, A. V.; III, M. H. H. Hidden markov models for face recognition. In: IEEE. *Acoustics, Speech, and Signal Processing, 1998. Proceedings., 1998 IEEE International Conference on*. [S.l.], 1998a. v. 5, p. 2721–2724.

NEFIAN, A. V.; III, M. H. H. Face detection and recognition using hidden markov models. In: IEEE. *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. [S.l.], 1998b. v. 1, p. 141–145.

NG, C. K.; SAVVIDES, M.; KHOSLA, P. K. Real-time face verification system on a cell-phone using advanced correlation filters. In: IEEE. *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. [S.l.], 2005. p. 57–62.

NIINUMA, K.; HAN, H.; JAIN, A. K. Automatic multi-view face recognition via 3d model based pose regularization. In: IEEE. *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. [S.l.], 2013. p. 1–8.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, Elsevier, v. 29, n. 1, p. 51–59, 1996.

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

OLIVEIRA, L. et al. *Relatório da disciplina de Visão Computacional e Reconhecimento de Padrões*. [S.l.], 2013.

OSUNA, E.; FREUND, R.; GIROSI, F. Training support vector machines: an application to face detection. In: IEEE. *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. [S.l.], 1997. p. 130–136.

PAN, G. et al. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: IEEE. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. [S.l.], 2007. p. 1–8.

PARVEEN, S. et al. Face anti-spoofing methods. *Current Science (00113891)*, v. 108, n. 8, 2015.

PEACOCK, C.; GOODE, A.; BRETT, A. Automatic forensic face recognition from digital images. *Science & Justice*, Elsevier, v. 44, n. 1, p. 29–34, 2004.

PEIXOTO, B.; MICHELASSI, C.; ROCHA, A. Face liveness detection under bad illumination conditions. In: IEEE. *Image Processing (ICIP), 2011 18th IEEE International Conference on*. [S.l.], 2011. p. 3557–3560.

PENTLAND, A.; MOGHADDAM, B.; STARNER, T. View-based and modular eigenspaces for face recognition. In: IEEE. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. [S.l.], 1994. p. 84–91.

PEREIRA, T. de F. et al. Lbp- top based countermeasure against face spoofing attacks. In: SPRINGER. *Computer Vision-ACCV 2012 Workshops*. [S.l.], 2013. p. 121–132.

PINTO, A. et al. Using visual rhythms for detecting video-based facial spoof attacks. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 10, n. 5, p. 1025–1038, 2015.

PINTO, A. d. S. et al. Video-based face spoofing detection through visual rhythm analysis. In: IEEE. *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*. [S.l.], 2012. p. 221–228.

PINTO, N.; DICARLO, J. J.; COX, D. D. How far can you get with a modern face recognition test set using only simple features? In: IEEE. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. [S.l.], 2009. p. 2591–2598.

PRINOSIL, J. Local descriptors based face recognition engine for video surveillance systems. In: IEEE. *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*. [S.l.], 2013. p. 862–866.

RABINER, L. R.; JUANG, B.-H. An introduction to hidden markov models. *ASSP Magazine, IEEE*, IEEE, v. 3, n. 1, p. 4–16, 1986.

RADUCANU, B.; DORNAIKA, F. Pose-invariant face recognition in videos for human-machine interaction. In: SPRINGER. *Computer Vision–ECCV 2012. Workshops and Demonstrations*. [S.l.], 2012. p. 566–575.

RAGHAVENDRA, R. et al. A new perspective—face recognition with light-field camera. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013. p. 1–8.

RAPP, V. et al. Multiple kernel learning svm and statistical validation for facial landmark detection. In: IEEE. *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. [S.l.], 2011. p. 265–271.

RODRIGUES, R. N.; KAMAT, N.; GOVINDARAJU, V. Evaluation of biometric spoofing in a multimodal system. In: IEEE. *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. [S.l.], 2010. p. 1–5.

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

SAMAL, A.; IYENGAR, P. A. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition*, Elsevier, v. 25, n. 1, p. 65–77, 1992.

SAMARIA, F.; FALLSIDE, F. *Face identification and feature extraction using hidden markov models*. [S.l.]: Citeseer, 1993.

SAMARIA, F.; YOUNG, S. Hmm-based architecture for face identification. *Image and vision computing*, Elsevier, v. 12, n. 8, p. 537–543, 1994.

SAMARIA, F. S.; HARTER, A. C. Parameterisation of a stochastic model for human face identification. In: IEEE. *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*. [S.l.], 1994. p. 138–142.

SCHWARTZ, W. R.; ROCHA, A.; EDRINI, H. P. Face spoofing detection through partial least squares and low-level descriptors. In: IEEE. *Biometrics (IJCB), 2011 International Joint Conference on*. [S.l.], 2011a. p. 1–8.

SCHWARTZ, W. R. et al. A novel feature descriptor based on the shearlet transform. In: IEEE. *Image Processing (ICIP), 2011 18th IEEE International Conference on*. [S.l.], 2011b. p. 1033–1036.

SENA, E. D. R. Técnicas multilineares em reconhecimento facial. 2014.

SHIN, H.; KIM, S.-D.; CHOI, H.-C. Generalized elastic graph matching for face recognition. *Pattern Recognition Letters*, Elsevier, v. 28, n. 9, p. 1077–1082, 2007.

SIROVICH, L.; KIRBY, M. Low-dimensional procedure for the characterization of human faces. *JOSA A*, Optical Society of America, v. 4, n. 3, p. 519–524, 1987.

STONHAM, T. Practical face recognition and verification with wisard. In: *Aspects of face processing*. [S.l.]: Springer, 1986. p. 426–441.

SUN, Y.; WANG, X.; TANG, X. Hybrid deep learning for face verification. In: IEEE. *Computer Vision (ICCV), 2013 IEEE International Conference on*. [S.l.], 2013. p. 1489–1496.

SUNG, K.-K.; POGGIO, T. Learning human face detection in cluttered scenes. In: SPRINGER. *Computer Analysis of Images and Patterns*. [S.l.], 1995. p. 432–439.

SWEILAM, N. H.; THARWAT, A.; MONIEM, N. A. Support vector machine for diagnosis cancer disease: A comparative study. *Egyptian Informatics Journal*, Elsevier, v. 11, n. 2, p. 81–92, 2010.

TAIGMAN, Y. et al. Deepface: Closing the gap to human-level performance in face verification. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. [S.l.], 2014. p. 1701–1708.

TAMURA, S.; KAWAI, H.; MITSUMOTO, H. Male/female identification from 8× 6 very low resolution face images by neural network. *Pattern Recognition*, Elsevier, v. 29, n. 2, p. 331–335, 1996.

TAN, X. et al. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: *Computer Vision–ECCV 2010*. [S.l.]: Springer, 2010. p. 504–517.

TEFAS, A.; KOTROPOULOS, C.; PITAS, I. Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 23, n. 7, p. 735–746, 2001.

TIRUNAGARI, S. et al. Detection of face spoofing using visual dynamics. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 10, n. 4, p. 762–777, 2015.

TOH, K.-A.; KIM, J.; LEE, S. Biometric scores fusion based on total error rate minimization. *Pattern Recognition*, Elsevier, v. 41, n. 3, p. 1066–1082, 2008.

TOLBA, A.; EL-BAZ, A.; EL-HARBY, A. Face recognition: A literature review. *International Journal of Signal Processing*, Citeseer, v. 2, n. 2, p. 88–103, 2006.

TORRALBA, A.; EFROS, A. et al. Unbiased look at dataset bias. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. [S.l.], 2011. p. 1521–1528.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. *Journal of cognitive neuroscience*, MIT Press, v. 3, n. 1, p. 71–86, 1991.

TURK, M.; PENTLAND, A. P. et al. Face recognition using eigenfaces. In: IEEE. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. [S.l.], 1991. p. 586–591.

UDDIN, M. P. et al. Developing an efficient solution to information hiding through text steganography along with cryptography. In: IEEE. *Strategic Technology (IFOST), 2014 9th International Forum on*. [S.l.], 2014. p. 14–17.

VAPNIK, V. N.; VAPNIK, V. *Statistical learning theory*. [S.l.]: Wiley New York, 1998.

VERLINDE, P.; CHOLLET, G.; ACHEROY, M. Multi-modal identity verification using expert fusion. *Information Fusion*, Elsevier, v. 1, n. 1, p. 17–33, 2000.

VETTER, T.; POGGIO, T. Linear object classes and image synthesis from a single example image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 19, n. 7, p. 733–742, 1997.

WANG, T. et al. Face liveness detection using 3d structure recovered from a single camera. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013. p. 1–6.

WEN, D.; HAN, H.; JAIN, A. K. Face spoof detection with image distortion analysis. *Information Forensics and Security, IEEE Transactions on*, IEEE, v. 10, n. 4, p. 746–761, 2015.

WEYRAUCH, B. et al. Component-based face recognition with 3d morphable models. In: IEEE. *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. [S.l.], 2004. p. 85–85.

WISKOTT, L. et al. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 19, n. 7, p. 775–779, 1997.

WISKOTT, L.; MALSBURG, C. V. D. Recognizing faces by dynamic link matching. *Neuroimage*, Elsevier, v. 4, n. 3, p. S14–S18, 1996.

WRIGHT, J.; HUA, G. Implicit elastic matching with random projections for pose-variant face recognition. In: IEEE. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. [S.l.], 2009. p. 1502–1509.

WÜRTZ, R. P. Object recognition robust under translations, deformations, and changes in background. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 7, p. 769–775, 1997.

XU, L. et al. A facial recognition method based on 3-d images analysis for intuitive human-system interaction. In: IEEE. *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*. [S.l.], 2013. p. 371–377.

YACOUBI, A. E. *Modélisation Markovienne de l'écriture manuscrite Application à la reconnaissance des adresses postales*. Tese (Doutorado), 1996.

YAN, J. et al. Face liveness detection by exploring multiple scenic clues. In: IEEE. *Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on*. [S.l.], 2012. p. 188–193.

YANG, J. et al. Face liveness detection with component dependent descriptor. In: IEEE. *Biometrics (ICB), 2013 International Conference on*. [S.l.], 2013. p. 1–6.

YUA, H. et al. Lifelong and fast transfer learning for gesture interaction. *Journal of Information & Computational Science*, JOICS, v. 11, n. 4, p. 1023–1035, 2014.

YUAN, X.; LU, J.; YAHAGI, T. A method of 3d face recognition based on principal component analysis algorithm. In: IEEE. *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*. [S.l.], 2005. p. 3211–3214.

ZELJKOVIC, V. et al. Personal access control system using moving object detection and face recognition. In: IEEE. *High Performance Computing & Simulation (HPCS), 2014 International Conference on*. [S.l.], 2014. p. 662–669.

ZHANG, C.; ZHANG, Z. Improving multiview face detection with multi-task deep convolutional neural networks. In: IEEE. *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. [S.l.], 2014. p. 1036–1041.

ZHANG, J.; YAN, Y.; LADES, M. Face recognition: eigenface, elastic matching, and neural nets. *Proceedings of the IEEE*, IEEE, v. 85, n. 9, p. 1423–1435, 1997.

ZHANG, Z. et al. A face antispoofing database with diverse attacks. In: IEEE. *Biometrics (ICB), 2012 5th IAPR international conference on*. [S.l.], 2012. p. 26–31.

ZHAO, L.; YANG, Y.-H. Theoretical analysis of illumination in pca-based vision systems. *Pattern recognition*, Elsevier, v. 32, n. 4, p. 547–564, 1999.

ZHAO, W. et al. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, ACM, v. 35, n. 4, p. 399–458, 2003.

ZHI-PENG, F.; YAN-NING, Z.; HAI-YAN, H. Survey of deep learning in face recognition. In: *2014 International Conference on Orange Technologies*. [S.l.: s.n.], 2014.