



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
MESTRADO ACADÊMICO EM ENGENHARIA DE TELEINFORMÁTICA

GILDERLANE RIBEIRO ALEXANDRE

UMA ANÁLISE CRÍTICA SOBRE AS TÉCNICAS 3D EMPREGADAS NO
RECONHECIMENTO DE EXPRESSÕES FACIAIS APOIADA EM REVISÃO
SISTEMÁTICA DA LITERATURA

FORTALEZA

2019

GILDERLANE RIBEIRO ALEXANDRE

UMA ANÁLISE CRÍTICA SOBRE AS TÉCNICAS 3D EMPREGADAS NO
RECONHECIMENTO DE EXPRESSÕES FACIAIS APOIADA EM REVISÃO
SISTEMÁTICA DA LITERATURA

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Orientador: Prof. Dr. José Marques Soares

Coorientador: Prof. Dr. George André Pereira Thé

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A369a Alexandre, Gilderlane Ribeiro.

Uma análise crítica sobre as técnicas 3D empregadas no reconhecimento de expressões faciais apoiada em revisão sistemática da literatura / Gilderlane Ribeiro Alexandre. – 2019.
70 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2019.

Orientação: Prof. Dr. José Marques Soares.

Coorientação: Prof. Dr. George André Pereira Thé.

1. Revisão sistemática da literatura. 2. Reconhecimento de expressões faciais. 3. Imagens 3D. I. Título.
CDD 621.38

GILDERLANE RIBEIRO ALEXANDRE

UMA ANÁLISE CRÍTICA SOBRE AS TÉCNICAS 3D EMPREGADAS NO
RECONHECIMENTO DE EXPRESSÕES FACIAIS APOIADA EM REVISÃO
SISTEMÁTICA DA LITERATURA

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia de Teleinformática. Área de Concentração: Sinais e Sistemas

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. José Marques Soares (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. George André Pereira Thé (Coorientador)
Universidade Federal do Ceará (UFC)

Prof^ª. Dra. Fátima Neusizeuma Sombra de Medeiros
Universidade Federal do Ceará (UFC)

Prof. Dr. Luciano Rebouças de Oliveira
Universidade Federal da Bahia (UFBA)

Aos meus pais, que sempre me incentivaram e apoiaram com todo o amor do mundo nesta jornada.

AGRADECIMENTOS

Agradeço aos meus pais pelo amor que dedicaram a mim durante toda minha vida, por me incentivarem a cada nova empreitada e por me oferecerem todo o amparo necessário para que eu pudesse realizar este trabalho.

Ao meu companheiro, Liduino de Albuquerque Marques, pelo apoio, carinho e compreensão durante os longos anos de curso.

Aos companheiros de pesquisa Artur Rodrigues Rocha Neto, Izaías Machado Junior, Polycarpo Souza Neto e Maria Estela de Oliveira Paiva por sua amizade e por compartilharem seus conhecimentos, contribuindo tão ricamente para o meu trabalho.

Aos meus orientadores, professor Dr. José Marques Soares e professor Dr. George André Pereira Thé, pela amizade e parceria no desenvolvimento deste trabalho, sendo sempre ouvintes atentos das minhas preocupações e oferecendo apoio e encorajamento.

Ao professor Dr. Robson da Silva Siqueira por ter desenvolvido os estudos que serviram de forte alicerce para a construção deste trabalho.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

“É com o coração que se vê corretamente; o essencial é invisível aos olhos.”

(Antoine de Saint-Exupéry)

RESUMO

As imagens 3D capturam informação espacial sobre as deformações da superfície da face e permitem a superação de algumas restrições frequentemente encontradas nas imagens digitais bidimensionais (2D). As soluções para o problema do Reconhecimento de Expressões Faciais em 3D (REF 3D) combinam, em geral, técnicas de processamento digital de imagens e de aprendizagem de máquinas. As investigações neste domínio costumemente se apoiam no uso de bases de dados disponibilizadas por grupos de pesquisa ou instituições acadêmicas, sendo um particular desafio o fato de estas bases conterem um número limitado de amostras com expressões faciais de poucos indivíduos. De maneira geral, sobre as amostras são aplicadas técnicas de pré-processamento, extração e seleção de atributos e classificação automática por meio de algoritmos classificadores, restando, ainda, muitos desafios para o reconhecimento robusto de expressões humanas. Uma análise crítica das múltiplas abordagens e técnicas alternativas que vem sendo propostas para superar esses desafios pode enriquecer e dar suporte ao processo decisório sobre o desenvolvimento de aplicações e/ou permitir a identificação de eventuais lacunas para novas investigações. Neste sentido, é apresentada nesta dissertação a avaliação de 49 estudos especializados em REF 3D por meio de uma Revisão Sistemática da Literatura. Três principais aspectos são levados em consideração: o pré-processamento dos dados capturados, as decisões de representação da face e os experimentos de classificação. A investigação permitiu identificar os tipos preferenciais de dados e de regiões de interesse para representação da face. Foi também possível constatar a tendência em buscar métodos que sejam independentes do reconhecimento *a priori* de pontos característicos (*keypoints*) da face. A análise crítica permitiu, ainda, ressaltar e documentar algumas vulnerabilidades metodológicas encontradas nos procedimentos de pré-processamento e no projeto de experimentos de classificação. Todas as descobertas foram cuidadosamente documentadas com o intuito de contribuir para o processo decisório e orientação de pesquisadores em REF 3D.

Palavras-chave: Revisão sistemática da literatura. Reconhecimento de expressões faciais. Imagens 3D.

ABSTRACT

Digital three-dimensional (3D) images represent a non-invasive alternative for representation and automatic classification of human emotional expressions. 3D images capture spatial information about deformations on face surface and allow overcoming some restrictions frequently found in digital bidimensional (2D) images. Solutions for the 3D Facial Expression Recognition (3D FER) problem combine, generally, computer vision and machine learning techniques. Investigations in this domain frequently rely on the use of databases made available by research groups or academic institutions, being a particular challenge the fact of those databases have a limited number of samples of few subjects performing facial expressions. Usually, preprocessing, feature extraction and selection procedures in addition to automatic classification algorithms are applied to this data but many challenges remain for the robust recognition of human emotional expressions. A critical analysis of multiple approaches and alternative techniques that have been proposed to overcome those challenges can enrich and support decision making on application development and identify gaps for further research. In this direction, it is presented in this dissertation the evaluation of 49 3D FER specialized studies by means of a Systematic Literature Review. Three major aspects are taken into consideration: preprocessing of captured data, face representation decisions and classification experiments. The investigation allowed identifying the preferential types of data and regions of interest for face representation. Also, it made it possible to note a trend towards the adoption of methods independent of a priori detection of characteristic points (keypoints) on face. The critical analysis also allowed highlighting and documenting some methodological vulnerabilities found in preprocessing procedures and in the design of classification experiments. All findings were carefully documented in order to contribute to the decision making and guidance of researchers in 3D FER.

Keywords: Systematic literature review. Facial expression recognition. 3D images.

LISTA DE FIGURAS

Figura 1 – Sete regiões de interesse utilizadas em Lekdioui <i>et al.</i> (2017)	20
Figura 2 – Nuvem de pontos de uma face humana.	21
Figura 3 – Fluxo usual seguido em métodos tradicionais para Reconhecimento de Expressões Faciais (REF) 3D.	22
Figura 4 – Estudos publicados sobre REF 3D e indexados pelo Scopus de 2008 a 2018. A reta de regressão sobre os dados mostra uma crescente dos registros no intervalo consultado.	26
Figura 5 – Fases e atividades do processo de RSL conduzida neste trabalho.	28
Figura 6 – Resumo dos procedimentos de busca e seleção de documentos.	34
Figura 7 – Tipos de dados disponíveis nas imagens empregadas nos estudos consultados.	37
Figura 8 – Exemplos de regiões de interesse extraídas em métodos locais de representação da face.	39
Figura 9 – Proporções dos escopos de região de interesse (RoI) empregadas por tipo de dados disponíveis nas imagens.	40
Figura 10 – Número de pontos característicos requeridos pelos 49 estudos consultados.	45
Figura 11 – Ocorrência dos modos de detecção de pontos relevantes nos estudos consultados.	46
Figura 12 – A frequência das principais características de amostras de dados empregadas em experimentos com a BU-3DFE.	48
Figura 13 – Amostras de dados mais frequentes da BU-3DFE	48
Figura 14 – Proporções de treinamento e teste mais frequentes em REF 3D.	51
Figura 15 – Medianas das acurácias reportadas em REF 3D, discriminadas por emoção.	52
Figura 16 – Proporção de estudos independentes de <i>keypoints</i> por ano de publicação.	59
Figura 17 – O <i>Trade-off</i> viés-variância.	73
Figura 18 – Os intervalos de confiança de 99% para diferentes valores de N	75

LISTA DE TABELAS

Tabela 1 – Unidades de ação facial.	18
Tabela 2 – Regras de associação entre AU e expressões de emoções básicas.	19
Tabela 3 – Critérios de inclusão e exclusão	31
Tabela 4 – Formulário de extração de dados preenchido para cada estudo selecionado. .	35
Tabela 5 – 49 estudos consultados categorizados de acordo com os escopos de suas regiões de interesse (RoI).	38
Tabela 6 – Estudos que reportam a aplicação de técnicas para o pré-processamento dos <i>scans</i> 3D de faces.	41
Tabela 7 – Porção dos estudos consultados que reportam segmentação da face como parte do seu pré-processamento.	43
Tabela 8 – Bases de dados utilizadas para validação nos estudos consultados.	47
Tabela 9 – Bases de dados utilizadas para validação nos estudos consultados.	54
Tabela 10 – Quadro resumo dos resultados.	55

LISTA DE ABREVIATURAS E SIGLAS

AUs	<i>action units</i>
AV	Aversão
FACS	<i>Facial Action Coding System</i>
FE	Felicidade
FLDA	<i>Fisher's Linear Discriminant Analysis</i>
HMM	<i>Hidden Markov Model</i>
ICNP	<i>Iterative Closest Normal Point</i>
ICP	<i>Iterative Closest Point</i>
K-NN	<i>K-nearest neighbors</i>
LBP	<i>Local Binary Pattern</i>
ME	Medo
NE	Neutra
NN	<i>Nearest neighbors</i>
PCA	<i>Principal Component Analysis</i>
RA	Raiva
REF	Reconhecimento de Expressões Faciais
RF	<i>Random Forest</i>
RoI	regiões de interesse
RoI	região de interesse
RSL	Revisão Sistemática da Literatura
SIFT	<i>Scale-Invariant feature transform</i>
SU	Surpresa
SVM	<i>Support Vector Machine</i>
TR	Tristeza

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.2	Produção científica	15
1.3	Estrutura do trabalho	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	O que a face revela?	17
2.2	A detecção automática de expressões emocionais	19
2.2.1	<i>Pré-processamento</i>	22
2.2.2	<i>Representação da face</i>	22
2.2.3	<i>Experimentos de classificação</i>	23
2.3	O que se espera de um trabalho científico?	24
2.4	Considerações finais	25
3	METODOLOGIA	27
3.1	Planejamento	27
3.1.1	<i>Objetivo e questões de pesquisa</i>	28
3.1.2	<i>Fontes e estratégia de busca</i>	29
3.1.2.1	<i>Busca automática</i>	29
3.1.2.2	<i>Busca manual</i>	30
3.1.3	<i>Crítérios de seleção</i>	30
3.1.4	<i>Avaliação do protocolo</i>	32
3.2	Condução	32
3.2.1	<i>Resultados de busca e seleção</i>	33
3.2.2	<i>Extração de dados</i>	33
3.2.3	<i>Sumarização dos dados</i>	34
3.3	Considerações finais	34
4	RESULTADOS	36
4.1	Representação da face	36
4.1.1	<i>Tipos de dados e descritores</i>	36
4.1.2	<i>Regiões de interesse</i>	37
4.2	Pré-processamento	40

4.2.1	<i>Correção de pose</i>	40
4.2.2	<i>Corte ou segmentação da face</i>	42
4.2.3	<i>Suavização</i>	42
4.2.4	<i>Preenchimento de buracos</i>	43
4.2.5	<i>Detecção de pontos relevantes (keypoints)</i>	44
4.3	Experimentos de classificação	46
4.3.1	<i>Amostra de dados</i>	46
4.3.1.1	<i>BU-3DFE</i>	47
4.3.1.2	<i>Bosphorus</i>	49
4.3.1.3	<i>BU-4DFE</i>	49
4.3.1.4	<i>Outras bases de dados</i>	50
4.3.2	<i>Classificadores e estratégia de classificação</i>	50
4.4	Considerações finais	52
5	DISCUSSÕES	56
5.1	Representação da face	56
5.2	Pré-processamento	57
5.3	Experimentos de classificação	58
6	CONCLUSÕES E TRABALHOS FUTUROS	61
	REFERÊNCIAS	62
	APÊNDICES	69
	APÊNDICE A – Strings de busca genérica e adaptadas	69
A.1	<i>String de busca genérica</i>	69
A.2	<i>String de busca adaptada para base bibliográfica Scopus</i>	69
A.3	<i>String de busca adaptada para a base bibliográfica Web of Science</i>	69
A.4	<i>String de busca adaptada para a base bibliográfica IEEE Xplore</i>	70
A.5	<i>String de busca adaptada para a base bibliográfica ACM Digital Library</i>	70
	APÊNDICE B – Formulário de extração de dados preenchido	72
	APÊNDICE C – Estabilidade das estimativas em classificação de padrões	73

1 INTRODUÇÃO

As reações emocionais humanas são de grande interesse da ciência e seu estudo tem aplicações em diversas áreas, como entretenimento, saúde e segurança. As expressões faciais são apenas um dos vários meios que podem ser explorados e pelos quais se revelam as reações emocionais. Entretanto, são uma das formas mais naturais de comunicação emocional entre seres humanos e, em termos de usabilidade, uma das formas menos invasivas de coletar dados emocionais de indivíduos.

Tradicionalmente, aplicações que fazem uso de expressões emocionais a partir das faces utilizam imagens 2D estáticas ou dinâmicas e requerem equipamentos já popularizados e de baixo custo como câmeras digitais. Mais recentemente, a popularização de equipamentos de aquisição de imagens 3D tem permitido a utilização desses tipos de dados para biometria e reconhecimento de expressões emocionais.

Sandbach *et al.* (2012) apontam para a dificuldade de distinguir-se algumas regiões da musculatura facial a partir de imagens 2D ao passo em que as representações 3D permitem essa distinção. Por essas características, o reconhecimento de expressões faciais 3D (REF 3D) tornou-se uma área de crescente interesse de pesquisa.

Consequentemente, alguns trabalhos de revisão da literatura descrevem diferentes aspectos das produções científicas na área. Sandbach *et al.* (2012) apresentam uma abrangente revisão da literatura em REF 3D, com uma caracterização de diversos aspectos como aquisição, bases de dados de faces 3D disponíveis, categorização das técnicas de extração de atributos em vigor e classificação automática. Desde então, as mais recentes revisões da literatura relacionadas a faces 3D não têm enfatizado REF. Por exemplo, Corneanu *et al.* (2016) apresentam uma abrangente revisão da literatura sobre métodos multimodais de reconhecimento automático de expressões faciais. Essa abordagem ampla contribui para a compreensão do histórico e tendências na área e ajuda a posicionar as técnicas 3D dentro de um cenário de variadas outras técnicas de imagens, mas não privilegia nenhuma delas em sua discussão. Em Danelakis *et al.* (2015), estudos sobre REF 3D são consultados apenas em cenários dinâmicos (sequência de imagens 3D ou vídeos 3D). Ademais, nesses estudos, ignoram-se quais os métodos de seleção e consulta à literatura adotados.

A fim de investigar de maneira aprofundada as pesquisas desenvolvidas nesta área, apresenta-se uma nova descrição dos trabalhos em REF 3D e atualiza-se a compreensão do cenário atual de publicações com ênfase em aspectos tradicionalmente pouco abordados em

estudos secundários.

1.1 Objetivos

Nesta dissertação, tem-se como principal objetivo analisar criticamente os métodos 3D empregados no reconhecimento de expressões faciais com o apoio da condução de uma Revisão Sistemática da Literatura (RSL) no período compreendido entre 2013 e 2018, identificando oportunidades de pesquisa futura e desafios existentes na área.

Os objetivos específicos a serem alcançados são:

- elaborar um protocolo de RSL para documentar detalhadamente o processo de seleção de referências e avaliação de evidências, bem como para que este mesmo protocolo possa ser utilizado por outros pesquisadores, a fim de atualizar futuramente os resultados registrados para o período coberto por esta dissertação;
- coletar de maneira objetiva e estruturada os dados considerados relevantes nos estudos submetidos à revisão para torná-los públicos e passíveis de serem re-analisados por outros pesquisadores;
- descrever os principais aspectos relativos às decisões sobre representação da face para identificar tendências e oportunidades de proposição de novos métodos;
- realizar uma descrição detalhada das técnicas de pré-processamento requeridas por métodos de REF 3D para compreender as principais limitações impostas pelos dados disponíveis às técnicas propostas nesse domínio;
- caracterizar os principais protocolos de experimentação dos estudos em REF 3D para identificar as configurações que tornam os experimentos na área mais facilmente comparáveis e de maneira justa;

1.2 Produção científica

Ao longo do percurso trilhado para o desenvolvimento deste trabalho de mestrado, alguns dos resultados e aspectos mais relevantes das investigações realizadas foram publicados. Tais publicações, que são listadas em seguida, contribuíram sobremaneira para as conclusões expostas neste documento.

- Artigos completos publicados em periódicos internacionais
ALEXANDRE, G. R.; SOARES, J. M.; PEREIRA THÉ, G. A. Systematic review

of 3D facial expression recognition methods. **Pattern Recognition**, v. 100, p. 107108, 1 abr. 2020.

SIQUEIRA, R. S. et al. Triaxial Slicing for 3-D Face Recognition From Adapted Rotational Invariants Spatial Moments and Minimal Keypoints Dependence. **IEEE Robotics and Automation Letters**, v. 3, n. 4, p. 3513–3520, out. 2018.

- Artigo completo publicado em conferência internacional

ALEXANDRE, G. R.; THÉ, G. A. P.; SOARES, J. M. Reliability of Results and Fairness in the Comparison of Rates Among 3D Facial Expression Recognition Works In: **Lecture Notes in Computer Science**.1 ed.: Springer International Publishing, 2019, p. 391-401.

1.3 Estrutura do trabalho

Esta dissertação está organizada em seis capítulos.

No Capítulo 2 são apresentadas as fundamentações teóricas que sustentam a condução deste trabalho e que incluem o estudo da face como comunicador de mensagens, os métodos de detecção automática de expressões faciais e os fundamentos de metodologia científica.

No Capítulo 3 é apresentado em detalhes o método de condução deste trabalho por meio da descrição das fases de planejamento e condução da RSL.

Adiante, no Capítulo 4, são identificados os principais resultados provenientes dos dados coletados pela condução da RSL, incluindo as respostas objetivas às questões de pesquisa da revisão.

No Capítulo 5 são elaboradas discussões sobre os resultados que estendem a análise metodológica das técnicas investigadas.

Finalmente, no Capítulo 6 são apresentadas as conclusões e perspectivas de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Os conceitos que serviram de referencial teórico para a abordagem adotada neste trabalho se relacionam principalmente com o estudo das expressões faciais, os métodos para sua detecção automática a partir de imagens, bem como com os fundamentos de metodologia científica. O objetivo deste capítulo é apresentar como esses assuntos se relacionam e, juntos, fornecem os fundamentos para a condução deste trabalho.

2.1 O que a face revela?

As faces possuem um alto potencial comunicativo, oferecendo mais de um tipo de sinal para transmitir mais de um tipo de mensagem. A face pode revelar sinais estáticos ou dinâmicos, que por sua vez podem ser lentos ou rápidos. Os sinais estáticos estão associados a características relativamente permanentes da face como cor da pele e dos olhos, estrutura óssea e formato do rosto e posição de elementos faciais como olhos, nariz, boca, sobrancelha. Quanto aos dinâmicos, os sinais lentos são aqueles que modificam a aparência facial gradualmente com o decorrer do tempo, como, por exemplo, mudança no tônus muscular e na textura da pele, que se manifestam na forma de flacidez e aparecimento de rugas com o avançar da idade ou com condições ambientais hostis. Já os sinais rápidos são produzidos por movimentos de músculos faciais e ocasionam mudanças temporárias na aparência como o levantar das sobrancelhas, o cerramento dos lábios, o franzimento da testa, cuja duração pode ser de segundos ou frações de segundo. Todos esses tipos de sinais são utilizados para comunicar mensagens sobre aspectos diversos como: humor, idade, gênero, etnia e muitos outros traços pessoais (EKMAN; FRIESEN, 2003).

O conceito de sistema multissinal-multimensagem oferecido pela face pressupõe o entendimento das exatas relações entre os sinais e as mensagens que esses sinais comunicam. Por exemplo, os sinais lentos são mais efetivos ao comunicar a idade de um indivíduo que os sinais rápidos. Por outro lado, o estado emocional é percebido por meio de sinais rápidos e não por sinais lentos ou estáticos. Não obstante, essas relações podem ser alteradas propositalmente ou acidentalmente, por meio da modificação ou disfarce de sinais faciais. É o caso, por exemplo, de óculos escuros e cirurgias plásticas, que comprometem a identificação de sinais estáticos; das maquiagens, que disfarçam sinais lentos; e da dissimulação de expressões, que tenta mascarar sinais rápidos (EKMAN; FRIESEN, 2003).

Ao examinar em específico as mensagens emocionais transmitidas pela face, os principais pontos de interesse são os sinais faciais rápidos. Isso, entretanto, não significa que as representações estáticas das ocorrências desses sinais, como fotografias por exemplo, não possam ser utilizadas para reconhecê-los. Em um dos mais famosos experimentos sobre a universalidade das expressões emocionais, Ekman (1971) submeteu fotografias de pessoas realizando expressões faciais à inspeção por membros da tribo *Fore* da Nova Guiné, pré-letrada e supostamente nunca exposta aos meios de comunicação de massa. Nesse estudo, Ekman (1971) coleta evidências que fortalecem a hipótese de que a modificação na aparência da face causada por cada uma das emoções de um conjunto de emoções básicas é comum a todas as pessoas, produto da evolução e não da interação social. Essa hipótese já havia sido considerada cerca de cem anos antes por Darwin (1872).

Atualmente o conjunto de emoções mais amplamente aceitas como básicas ou primárias são precisamente as seis propostas por Ekman: raiva, aversão, medo, felicidade, tristeza e surpresa. Em outra importante contribuição, Ekman e Friesen (1978) propõem o *Facial Action Coding System* (FACS), a representação de movimentos faciais em termos das *action units* (AUs), permitindo posteriormente a precisa associação entre movimentos musculares faciais e expressões emocionais primárias. Na Tabela 1, são apresentadas as descrições das principais AUs associadas a expressões emocionais e na Tabela 2 são listadas as associações entre as AUs e as emoções básicas.

Tabela 1 – Unidades de ação facial.

AU	Descrição
1	Levantador de sobrancelha interna
2	Levantador de sobrancelha externa
4	Abaixador de sobrancelha
6	Levantador de bochechas
7	Apertador de pálpebras
10	Levantador de lábio superior
12	Puxador de canto do lábio
15	Depressor de canto do lábio
17	Levantador de queixo
25	Separador de lábios
26	Queda de mandíbula

Fonte: adaptado de Valstar *et al.* (2011).

Naturalmente, as faces não são os únicos meios para comunicar emoções, e diversos outros sinais, além dos faciais rápidos, podem ser agregados para estimar o estado emocional de um indivíduo. Quando sinais faciais e outros sinais fisiológicos como dilatação pupilar,

Tabela 2 – Regras de associação entre AU e expressões de emoções básicas.

Expressão facial	AU
Felicidade	1, 6, 12, 16, 25, 26
Surpresa	1, 2, 5, 26, 27
Nojo	9, 10, 17, 25, 26
Raiva	2, 4, 7, 16, 17, 23, 24, 25, 26
Medo	1, 4, 5, 7, 20, 25, 26
Tristeza	1, 4, 7, 15, 17, 25, 26

Fonte: adaptado de Konar e Chakraborty (2015).

temperatura, entre outros, são utilizados em conjunto para a identificação de emoções, fala-se em métodos multimodais. Nesse caso, o ganho de informação tem geralmente a contrapartida de necessitar de métodos invasivos para aquisição dos sinais como por exemplo sensores vestíveis. Nesta dissertação, apenas os métodos que utilizam unicamente imagens faciais para reconhecimento de expressões emocionais foram considerados.

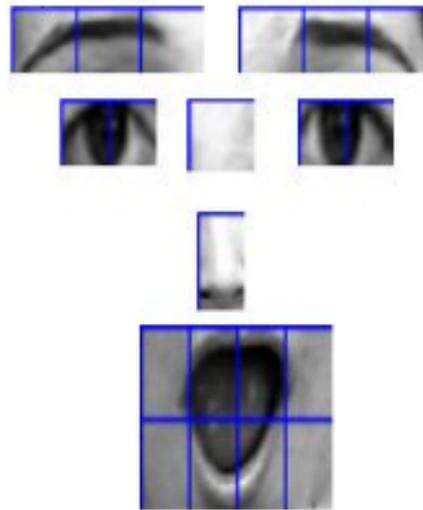
2.2 A detecção automática de expressões emocionais

A sistematização do mapeamento das emoções na face oferecida pelo FACS, além de poderosa ferramenta para o campo da psicologia que se ocupa do estudo das emoções, também implica em um cenário favorável para a automatização dessa tarefa por sistemas de computador. De fato, segundo Valstar *et al.* (2011), a detecção automática de emoções tem sido um tópico de interesse das ciências da computação desde a década de 1970. Desde então, as imagens digitais têm sido o principal objeto dos estudos propostos em REF.

Dentre as abordagens propostas em REF, uma vasta quantidade dispõe das noções de AUs e emoções básicas para a extração de descritores de intensidade e textura, agregadas a técnicas de aprendizado de máquinas para o reconhecimento automático. Por exemplo, Lekdioui *et al.* (2017) propõem uma solução para REF utilizando imagens 2D em escala de cinza. Segundo seu método, as imagens de faces são segmentadas em sete regiões de interesse que representam componentes considerados relevantes na face como olhos, região entre olhos, sobrancelhas, nariz e boca (ver Figura 1). Essas regiões de interesse, por sua vez, são divididas em blocos dos quais descritores de textura são extraídos. Os descritores de cada região de interesse são concatenados para formar o vetor de atributos que alimenta uma *Support Vector Machine* (SVM) multiclases. As imagens são classificadas como pertencentes a uma das sete classes a seguir: raiva, aversão, medo, felicidade, tristeza, surpresa e neutra (sem emoção aparente). Como resultado, a técnica

atingiu 94,04% de acurácia sobre a base de dados CK.

Figura 1 – Sete regiões de interesse utilizadas em Lekdioui *et al.* (2017)



Fonte: Lekdioui *et al.* (2017)

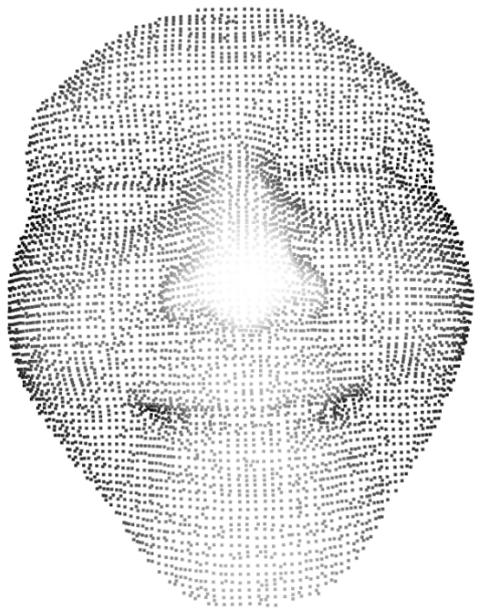
Apesar das medidas de desempenho acima de 90% que as técnicas para REF têm alcançado em diversos estudos com imagens digitais 2D nas últimas décadas, algumas dificuldades ainda são permanentes nesse domínio, como por exemplo a dependência de condições de iluminação adequadas para aquisição das imagens. Ainda, a suscetibilidade a disfarces e fraudes, por meio por exemplo do uso de maquiagens e pela apresentação de fotografias em vez do próprio indivíduo se torna uma fragilidade. Esses fatores restringem os cenários para aplicação e comprometem a confiabilidade dessas soluções.

Nos últimos anos a aplicação de métodos baseados em aprendizagem profunda (*deep learning*) tem ganhado força neste domínio, impulsionada pela alta disponibilidade de imagens digitais 2D de faces e facilidade de novas aquisições. Os métodos baseados em aprendizagem profunda se diferenciam dos métodos clássicos principalmente pelo aprendizado de características em vez de sua extração. Apesar de extremamente efetivos, os métodos baseados em aprendizagem profunda dependem de uma grande quantidade de dados para treinamento ou de técnicas de aumento de dados (*data augmentation*) para evitar o sobreajuste, além de exigir alto poder computacional para treinar esse volume de dados.

A popularização de sensores 3D tem permitido a aquisição da informação espacial da superfície dos objetos, dando origem a um novo tipo de representação, as imagens digitais

3D. Essas imagens capturam a profundidade dos objetos e são capazes de superar as principais limitações impostas pelas imagens digitais 2D, como a dependência de condições de iluminação e a suscetibilidade a disfarces pelo uso de maquiagens ou apresentação de fotografias. Na Figura 2 é apresentado um exemplo de nuvem de pontos, uma das formas de representar uma imagem digital 3D. Nessa representação, o conceito de pixel é substituído pelo conceito de ponto geométrico. O ponto, a menor unidade formadora da nuvem de pontos, não representa a intensidade da imagem, mas o conjunto de coordenadas espaciais (x, y, z) em relação ao sensor que realiza a captura. Opcionalmente, é possível associar informações de cor e textura às nuvens de pontos pela captura simultânea de imagens 2D e 3D.

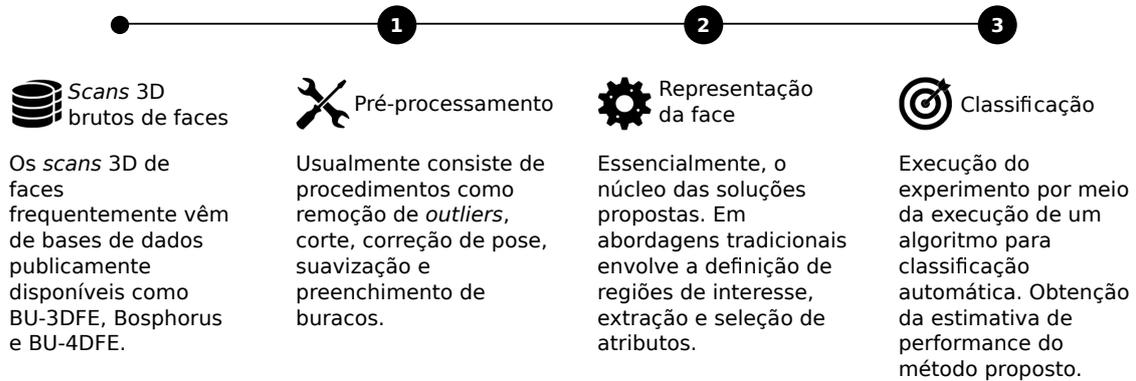
Figura 2 – Nuvem de pontos de uma face humana.



Fonte: Savran *et al.* (2008)

Em se tratando dos esforços científicos na direção do reconhecimento de expressões emocionais da face por sistemas automáticos, as soluções tradicionais são elaboradas por meio da combinação de técnicas de processamento de imagens digitais e aprendizado de máquinas. Nesse contexto, o fluxo comum seguido na proposição de soluções envolve procedimentos de pré-processamento, decisões para a representação da face e experimentos de classificação, como ilustrado na Figura 3. Essa estrutura fundamenta parte importante das decisões metodológicas adotadas neste trabalho. A seguir, são apresentadas discussões preliminares sobre cada um desses três principais procedimentos.

Figura 3 – Fluxo usual seguido em métodos tradicionais para REF 3D.



Fonte: elaborado pela autora.

2.2.1 Pré-processamento

Características intrínsecas ao processo de aquisição dos dados podem impor limitações importantes à adequada execução dos métodos propostos em REF. Chama-se de pré-processamento quaisquer procedimentos preparatórios que ajustam melhor os dados crus para torná-los mais adequados ao processamento que o segue. Por exemplo, em imagens digitais 2D, condições adversas de iluminação (ambientes muito escuros ou muito claros) podem ser tratadas, até certo ponto, pela manipulação dos histogramas. Similarmente, o ruído capturado durante a aquisição dos dados pode ser tratado com a aplicação de filtros de suavização. Em ambos os casos, esses procedimentos de preparação das imagens são considerados pré-processamento.

Similarmente, em REF 3D, o pré-processamento consiste em procedimentos que tornam as imagens digitais 3D de faces mais facilmente manipuláveis durante os procedimentos de representação da face e classificação automática. Notadamente, soluções robustas a condições adversas de aquisição e que dispensam pré-processamento também podem ser propostas para o problema de REF 3D.

2.2.2 Representação da face

Aprende-se da característica multissinal-multimensagem da face que nem todos os sinais emitidos por ela são relevantes para decodificar uma mensagem em específico. Tampouco, os sinais relevantes associados a determinada mensagem se manifestam de maneira uniforme

sobre toda a superfície da face. Ao reconhecer expressões faciais de emoções, apenas os sinais faciais rápidos, resultado de ações musculares específicas, são significativos. Isso significa que algumas regiões da face não contribuem para distinguir entre diferentes expressões emocionais. Não raro, abordagens para representar a face para um sistema automático de classificação incluem uma etapa de definição de regiões de interesse, explorando os diferentes graus de relevância que diferentes regiões da face tem para o problema de REF.

As regiões de interesse, entretanto, não são suficientes para representar a face adequadamente para um sistema de classificação. É necessário ainda representar essas regiões por valores, na forma de descritores ou vetores de características. Esses descritores, por sua vez, são calculados a partir do tipo de dados disponíveis na imagem que contém a face. Em REF 3D, os dados utilizados para descrever as deformações da face podem estar à disposição a partir de imagens digitais 3D, de suas sequências, os vídeos digitais 3D, ou da combinação desses com imagens digitais 2D adquiridas simultaneamente. Descritores específicos são necessários para capturar as características relevantes que estejam disponíveis em cada tipo de imagem. Por exemplo, enquanto a utilização de imagens digitais 2D permite o emprego de descritores de cor e textura, os dados na forma de nuvens de pontos 3D (ver Figura 2) não oferecem essa possibilidade. Em contrapartida, informações sobre a deformação da superfície da face podem ser muito bem representadas por descritores geométricos tão simples quanto as coordenadas dos pontos em uma nuvem de pontos.

Assim, a extração de descritores que representem adequadamente a face de forma a servir de entrada para um sistema automático de classificação depende do tipo de dados a disposição, bem como das regiões a partir das quais os descritores são extraídos. Naturalmente, fatores como a necessidade ou dispensabilidade de definir regiões locais de interesse, a forma como as eventuais regiões de interesse são definidas, os tipos de imagens e como são calculados os descritores são informações importantes ao caracterizar um método de representação da face.

Nesta dissertação qualquer menção à etapa de representação da face refere-se ao conjunto de decisões sobre definição de regiões de interesse, os tipos de dados das imagens e a extração e seleção de descritores ou atributos.

2.2.3 Experimentos de classificação

Os procedimentos de pré-processamento e decisões para representação da face mencionados nas subseções anteriores têm clara influência nos resultados. De fato, essas

formas de manipulação dos dados são deliberadamente projetadas para torná-los mais facilmente distinguíveis durante o processo de classificação automática. No entanto, mais sutilmente, o projeto do experimento de classificação em si também tem influência sobre os resultados alcançados.

No domínio de REF os resultados são geralmente apresentados na forma das acurácias médias do desempenho de sistemas de classificação automática. Tais valores são estimativas do desempenho real de um dado sistema, obtidas por meio da apresentação de amostras de uma população durante o treinamento do classificador. Essas estimativas são utilizadas para legitimar as contribuições apresentadas nos trabalhos, tornando-se os principais critérios de comparação entre técnicas.

A configuração experimental é, portanto, o terceiro elemento crucial para caracterização de uma abordagem, uma vez que está intrinsecamente relacionada aos resultados alcançados. Particularmente em estudos que tratam de REF 3D, além do classificador em si, outros aspectos devem ser notados para uma descrição abrangente da configuração experimental: a amostra de dados empregada, o número de rodadas de classificação e a proporção entre dados de treinamento e validação são elementos importantes.

2.3 O que se espera de um trabalho científico?

Segundo Lakatos e Marconi (2008), trabalhos científicos completos e originais permitem à comunidade de pesquisadores "a repetição das experiências mediante a descrição da metodologia empregada, do processamento utilizado e dos resultados obtidos". Sendo assim, a preocupação com aspectos metodológicos fundamentais caracteriza fortemente os trabalhos científicos. Essa preocupação permite à comunidade a continuidade de trabalhos, o seu eventual aperfeiçoamento e proposição de alternativas, o que tem sido a fundação sobre a qual se sustenta o avanço científico. Seguir e apropriadamente reportar um processo metodológico são condições básicas esperadas de uma contribuição que se pretende de impacto. De fato, segundo Marconi e Lakatos (2003), "não há ciência sem o emprego de métodos científicos".

Salvador (1986) sustenta que trabalhos científicos originais devem permitir à comunidade de pesquisadores:

- a) reproduzir as experiências e obter os resultados descritos, com a mesma precisão e sem ultrapassar a margem de erro indicada pelo autor;
- b) repetir as observações e julgar as conclusões do autor;

c) verificar a exatidão das análises e deduções que permitiram ao autor chegar às conclusões.

Apesar disso, há na literatura registros de que, no contexto de REF, muitas publicações têm falhado em reportar adequadamente seus processos metodológicos. Gong *et al.* (2009), ao tentar reproduzir os resultados obtidos em trabalhos relacionados para comparar com os seus próprios, identificaram alta instabilidade nos valores reportados. Valstar *et al.* (2011) criticam a ausência de informações sobre quais as partes exatas das bases de imagens faciais são utilizadas e sobre os protocolos de treinamento e teste conduzidos em publicações na área de REF.

2.4 Considerações finais

A identificação do estado emocional de outros seres humanos é uma tarefa complexa que tem evoluído com a espécie humana ao longo dos anos. Hoje sabe-se que as emoções desempenham um papel tão importante nas tomadas de decisão quanto a razão (GOLEMAN, 2012; KAHNEMAN, 2012). Reconhecer, interpretar e simular emoções humanas logo extrapolaram os limites da neurociência afetiva e tornaram-se também propósitos para sistemas computacionais, com aplicações em diversas áreas como segurança, entretenimento, educação e saúde (SCHULLER *et al.*, 2008; LUNESKI *et al.*, 2010).

Aliadas ao fato de que o estudo das expressões faciais tem sido um tópico de interesse científico por décadas, as conveniências oferecidas pelas representações 3D figuram entre as razões pelas quais a atividade de pesquisa em REF 3D têm crescido nos últimos anos. O gráfico da Figura 4 ilustra o número crescente de trabalhos sobre o tema “Reconhecimento de expressões faciais 3D” entre os anos de 2008 e 2018.

Naturalmente, o crescente interesse de pesquisa pela área tem levado a uma extensa variedade de abordagens que, por um lado, enriquecem as soluções disponíveis, mas, por outro, contribuem para o aparecimento de lacunas metodológicas, como as encontradas por Valstar *et al.* (2011) e Gong *et al.* (2009).

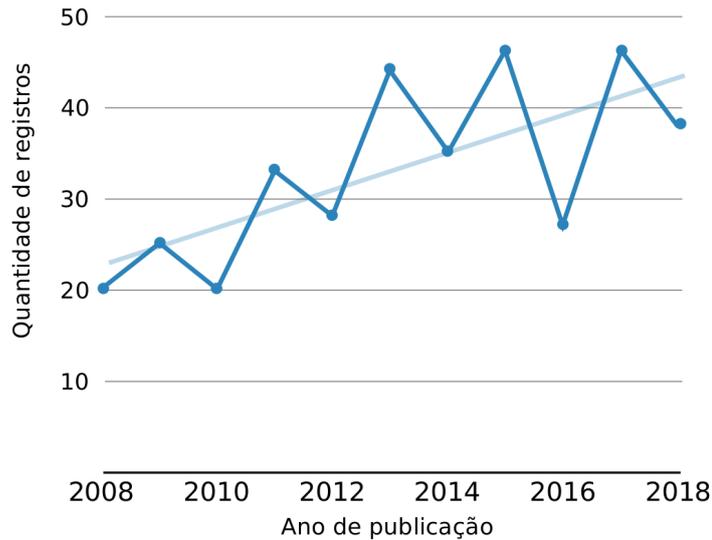
Nesse contexto, enquanto as proposições de métodos originais se tornam disponíveis, a demanda por estudos que agreguem, sumariem e estabeleçam conclusões sobre uma coleção de estudos primários é espontaneamente estimulada para a proposição de taxonomias, identificação de tendências e lacunas para contribuir com outros pesquisadores na área.

Em vista disso, neste trabalho, analisamos uma extensa variedade de estudos sobre REF 3D utilizando como ferramenta o método sistemático de revisão da literatura. Foram levados

Figura 4 – Estudos publicados sobre REF 3D e indexados pelo Scopus de 2008 a 2018. A reta de regressão sobre os dados mostra uma crescente dos registros no intervalo consultado.

Registros sobre o tópico REF 3D recuperados da base bibliográfica Scopus

Busca para o período de 2008 a 2018



Nota: *String* de busca: “3D” AND “facial” AND “expression recognition”.

Fonte: elaborado pela autora.

em consideração os três principais aspectos presentes em trabalhos originais e suas relações: 1) Representação da face; 2) Pré-processamento e 3) Experimentos de classificação. A condução da RSL, seus resultados e conclusões são apresentados nos próximos capítulos.

3 METODOLOGIA

A busca por evidências que apoiem ou refutem hipóteses acerca de determinado objeto de estudo, em geral, marca o início das investigações científicas. Isso se justifica pela necessidade de bem compreender o que tem sido produzido cientificamente sobre um tópico de pesquisa para a identificação de fundamentos, desafios e oportunidades antes da efetiva proposta de contribuição por parte de um pesquisador. De maneira geral, as evidências são obtidas por meio da investigação de uma coleção de estudos denominados primários e cujo objetivo é a proposição de soluções originais para um problema dentro de algum contexto específico. A análise e síntese dessas evidências, por sua vez, são realizadas por meio de um estudo secundário que estabelece conclusões sobre um conjunto de estudos primários. As revisões sistemáticas e não sistemáticas, bem como consultas à literatura¹ são exemplos de estudos secundários.

Uma RSL se diferencia de estudos secundários não sistemáticos por possuir um processo de condução com fases e atividades bem definidas. Muito embora as fases e atividades possam ser conduzidas de maneira iterativa, cada uma delas serve a um propósito específico dentro da revisão (FELIZARDO *et al.*, 2017). A sistematização da revisão da literatura minimiza as chances de viés do pesquisador ao selecionar e interpretar evidências. Além disso, uma RSL oferece os meios para que seja auditada e replicada, o que fortalece a confiabilidade das conclusões da investigação.

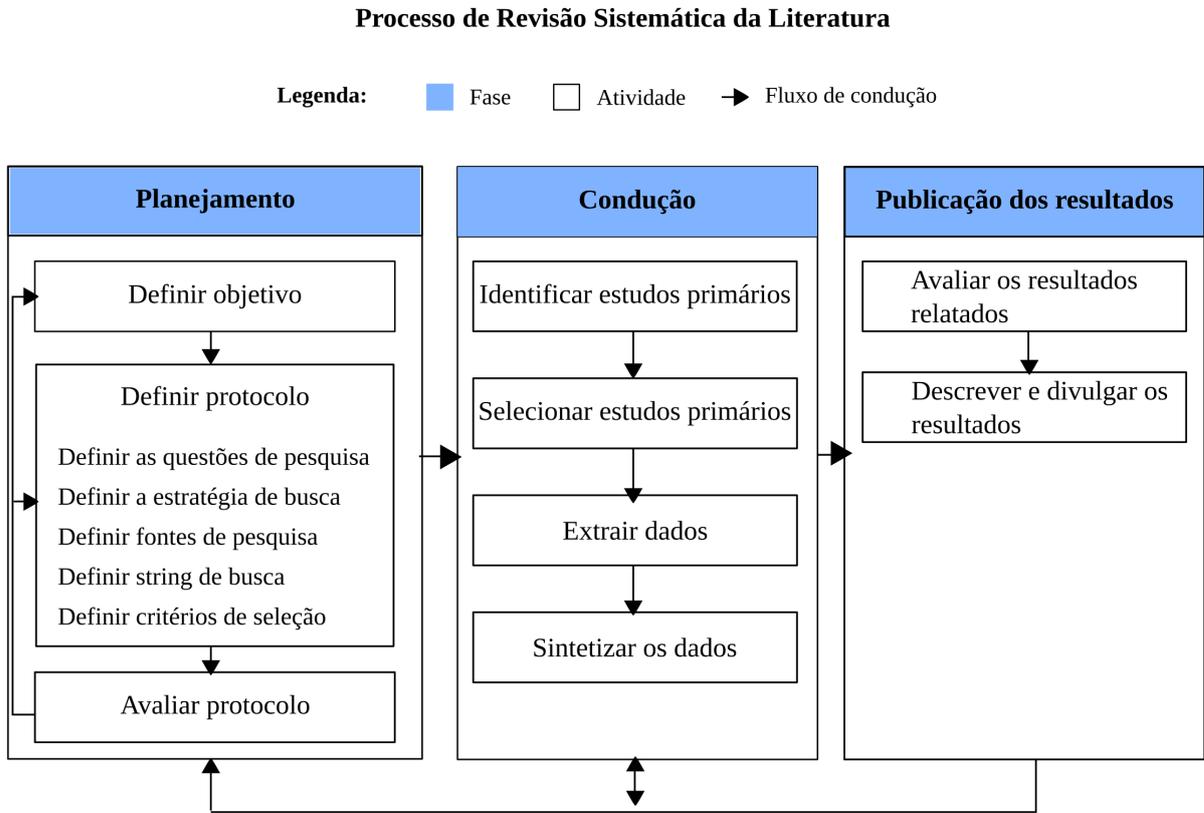
A busca por evidências realizada neste trabalho foi conduzida na forma de uma RSL baseada nas linhas guia propostas por Kitchenham (2004) e Biolchini *et al.* (2007) com adaptações. Na Figura 5 são sumarizadas as fases e atividades da revisão sistemática conduzida neste trabalho. Neste capítulo, as atividades realizadas durante o planejamento e condução da revisão são apresentadas em detalhes.

3.1 Planejamento

A existência de uma fase de planejamento é uma das características mais relevantes de uma RSL, uma vez que define e documenta como a revisão deve ser conduzida, oportunizando sua sistematização. Durante essa fase, devem ser cuidadosamente elaboradas as definições que guiarão a busca sistemática por evidências na literatura. As seções a seguir apresentam as

¹ Consultas à literatura são comumente referidas como *surveys* na literatura em língua inglesa e, neste caso, o termo não significa questionários aplicados a amostras de alguma população, o que é fundamentalmente um estudo primário.

Figura 5 – Fases e atividades do processo de RSL conduzida neste trabalho.



Fonte: Adaptado de Felizardo *et al.* (2017).

definições elaboradas neste trabalho.

3.1.1 *Objetivo e questões de pesquisa*

O objetivo da RSL conduzida neste trabalho é investigar as técnicas para o reconhecimento automático de emoções via expressões faciais 3D, desenvolvidas entre 2013 e 2018, com ênfase em três principais aspectos e suas relações: pré-processamento, representação da face e experimentos de classificação.

Uma maneira de abordar e garantir o atingimento do objetivo de pesquisa é a elaboração de questões de pesquisa, cujas repostas pretendem-se obter das conclusões da revisão. As questões de pesquisa que derivam do objetivo apresentado são as seguintes:

- QP1 Quais estratégias para representação da face são mais frequentemente utilizadas em REF 3D?
- QP2 Quais pré-processamentos têm sido requeridos para o tratamento de dados 3D com o propósito de REF?

QP3 Quais protocolos têm sido seguidos para os experimentos de classificação em REF 3D?

3.1.2 Fontes e estratégia de busca

A estratégia de busca por estudos primários adotada nesta RSL consiste na busca automática em quatro bancos de dados bibliográficos e na busca manual na lista de referências de estudos secundários recentes na área. A combinação das duas estratégias visa a aumentar a chance de alcançar estudos primários potencialmente relevantes. A seguir, os procedimentos de busca em cada uma das estratégias são detalhados.

3.1.2.1 Busca automática

As bases IEEE Xplore², ACM Digital Library³, Scopus⁴ e Web of Science⁵ foram selecionadas como fontes de busca automática por serem bases de dados reconhecidas por indexarem importantes periódicos e anais de conferências nas áreas de tecnologia e ciências. A diferença básica entre essas fontes consiste nos critérios de indexação, de forma que sua combinação visa a aumentar a abrangência das buscas. Portanto, espera-se que enquanto referências idênticas são recuperadas por bases de dados distintas, cada base também contribua com um número de referências únicas. A restrição de tempo aplicada à busca automática foi de 2013 a 2018 e também incluiu as referências disponíveis como acesso antecipado (*early access*) nas bases. O idioma dos trabalhos foi limitado ao inglês.

Finalmente, a partir de palavras-chaves e seus sinônimos extraídos do objetivo e questões de pesquisa, a seguinte *string* de busca genérica foi elaborada: **TITLE:** (3D OR “ Three dimension” OR “Three dimensional”) AND (face OR facial) AND (recognition OR recognizer OR detection OR detector OR detecting OR identification OR identifier OR identifying) AND (expression OR affect OR affection OR emotion OR emotional OR sentiment OR feeling) AND **TITLE OR ABSTRACT OR KEYWORDS:**(“key point extraction” OR “key point extraction” OR “key points extraction” OR “key points extraction” OR “feature selection” OR “features selection” OR “attribute selection” OR “attributes selection” OR descriptor OR signature OR classification OR classifying OR classifier OR application OR system).

² <https://ieeexplore.ieee.org/Xplore/home.jsp>

³ <https://dl.acm.org/>

⁴ <https://www.scopus.com/search/form.uri?display=basic>

⁵ <https://www.webofknowledge.com/>

3.1.2.2 Busca manual

A busca manual por estudos primários foi realizada pela consulta à lista de referências de estudos secundários recentes sobre o tópico REF (o termo em língua inglesa efetivamente empregado para a busca é *facial expression recognition*). Como a recuperação de um volume significativo de estudos primários é esperada a partir das buscas automáticas, a busca manual não se pretende exaustiva, mas complementar, tornando possível considerar estudos potencialmente relevantes que não puderam ser alcançados pela *string* de busca. Dessa forma, foram selecionados como fontes os estudos secundários Li e Deng (2018) e Zhang (2018), que não se limitam ao tópico REF 3D, mas incluem-no em suas discussões.

3.1.3 Critérios de seleção

Os estudos recuperados a partir das buscas automáticas e manual constituem um conjunto de trabalhos potencialmente relevantes para os objetivos da RSL. Entretanto, a inspeção dos estudos realizada durante as buscas é, por natureza, superficial, ao recuperar trabalhos apenas pela ocorrência de termos chave nos metadados. Por exemplo, uma parcela dos trabalhos recuperados, apesar de conter os termos chave, não resiste a uma análise semântica desses termos no texto. Assim, um trabalho que utiliza redes convolucionais 3D para o reconhecimento facial robusto a expressões emocionais a partir de imagens RGB presumivelmente possui os termos chave para ser recuperado nas atividades de busca, mas claramente não faz parte do conjunto de trabalhos-alvo desta RSL. Por isso, após as atividades de busca, os estudos recuperados foram submetidos à atividade de seleção para a identificação de estudos relevantes. Essa atividade consiste na procura por elementos textuais do artigo que ratifiquem sua relevância, sendo realizada, por conveniência, em dois estágios: seleção preliminar e seleção final. A seleção preliminar objetiva reduzir o número de documentos a serem lidos na íntegra de tal maneira que, apenas título e *abstract* do documento são avaliados. O subconjunto de documentos aceitos nesse estágio são posteriormente avaliados na íntegra durante a seleção final.

A fim de minimizar vieses do pesquisador em ambos os estágios de seleção, são definidos dois conjuntos de critérios de seleção: critérios de inclusão e critérios de exclusão, identificados por códigos únicos iniciados por CI e CE, respectivamente. Todos os critérios de seleção aplicados nesta revisão são listados na Tabela 3.

Os critérios de inclusão determinam as características que devem estar presentes

Tabela 3 – Critérios de inclusão e exclusão

ID	Descrição
CI01	Estudo apresenta uma técnica para o reconhecimento automático de expressões emocionais usando imagens 3D de faces.
CI02	Estudo propõe ou descreve uma aplicação ou sistema que implementa reconhecimento automático de expressões emocionais usando imagens 3D de faces.
CE01	Estudo não está disponível em texto completo (apenas o resumo disponível.)
CE02	Estudo não está disponível em inglês.
CE03	Estudo é uma duplicata dentre os estudos recuperados.
CE04	Estudo não foi publicado entre 2013 e 2018.
CE05	Estudo apresenta técnica exclusivamente dependente de imagens 2D.
CE06	Estudo não menciona a base de imagens utilizada para validar seu método OU não valida seu método com uma base de imagens 3D de faces.
CE07	Estudo aborda apenas o reconhecimento facial robusto a variações de expressão, não o reconhecimento de expressões emocionais.
CE08	Estudo não considera as seis emoções básicas: Raiva, Aversão, Medo, Felicidade, Tristeza e Surpresa.
CE09	Estudo não apresenta resultados de reconhecimento discriminados por emoção.
CE10	Estudo não atende aos critérios CI01 nem CI02.
CE11	Estudo é uma descrição de curso OU palestra OU patente OU editorial OU tutorial ou estudo secundário.

Fonte: elaborado pela autora.

no documento avaliado para que este seja aceito na seleção. Foram estabelecidos apenas dois critérios de inclusão que basicamente agrupam dois tipos de trabalhos originais nesse domínio: são aceitos trabalhos que propõem uma técnica para o reconhecimento emocional via expressão facial 3D, bem como trabalhos que descrevem um sistema ou aplicação que implementa tal técnica na prática. Os critérios de exclusão, por sua vez, especificam os aspectos que tornam o documento fora do escopo da RSL, o que inclui a condição em que nenhum dos critérios de inclusão se verificam (CE10), mas não se limita a essa.

Múltiplos critérios de exclusão podem ser atribuídos a um documento sob avaliação, de maneira que essa superposição de critérios pode ser utilizada para detalhar as razões para uma exclusão. Por exemplo, conforme CE05, os documentos que lidam unicamente com imagens RGB e, portanto, não empregam técnicas 3D, são desconsiderados. O critério CE07, por sua vez, desqualifica estudos que se limitam ao problema do reconhecimento facial e que, para isso, não consideram expressões emocionais como classes propriamente ditas. Considera-se ainda o universo de expressões emocionais dos estudos. Em CE08, o escopo de documentos é restrito àqueles que incluem nas suas investigações as seis expressões básicas: raiva, aversão, medo, felicidade, tristeza e surpresa. Assim, o estudo mencionado como exemplo anteriormente nesta subseção deveria ser excluído conforme, pelo menos, CE05 e CE07.

Finalmente, além de atender aos critérios de inclusão, um documento só é aceito se não se enquadrar em nenhum dos critérios de exclusão. Por outro lado, um critério de exclusão

é atribuído a um documento apenas se este for verificável. Em caso de dúvida, um trabalho é sempre aceito e submetido a uma investigação mais minuciosa no próximo estágio de seleção.

3.1.4 Avaliação do protocolo

A definição das estratégias de busca e seleção descritas nas Seções 3.1.2 e 3.1.3 são resultado de um processo incremental de avaliação e refinamento. A avaliação do protocolo foi realizada com respeito a dois aspectos: a definição da *string* de busca para cada base bibliográfica e a definição dos critérios de seleção.

A *string* de busca genérica, apresentada na Seção 3.1.2.1 precisou ser adaptada para cada base bibliográfica. Isso é necessário porque cada uma delas define um conjunto de regras para as consultas (*queries*) que suportam. Assim, o número máximo suportado de palavras-chave (*keywords*) ou operadores lógicos, por exemplo, pode variar entre as bases bibliográficas. Além disso, alguns recursos, tais como a possibilidade de combinar *queries* funcionam diferentemente entre as bases bibliográficas. Essas particularidades contribuem para que as *strings* de busca adaptadas sejam diferentes entre si, o que pode impactar os resultados esperados. Similarmente, o grau de assertividade dos critérios de seleção produz efeito sobre o conjunto final de documentos revisados. Portanto, buscas e seleções experimentais foram realizadas para avaliar se as adaptações da *string* de busca genérica e os critérios de seleção conduziam a resultados aceitáveis. Nesse processo, a *string* de busca foi refinada, bem como os critérios de seleção foram melhorados para clareza e completude antes da fase de condução. Para esse fim, os estudos Azazi *et al.* (2015), Yurtkan e Demirel (2014b), Hariri *et al.* (2017), Emambakhsh e Evans (2017), Hasani e Mahoor (2017) foram adotados como estudos de controle, a partir do conhecimento *a priori* do seu conteúdo. Assim, as *strings* de busca adaptadas foram consideradas aceitáveis apenas quando esses cinco estudos puderam ser encontrados no conjunto de todos os estudos recuperados das bases bibliográficas. Posteriormente, os critérios de seleção deveriam permitir a inclusão dos três primeiros, assim como a exclusão dos dois últimos. As *strings* de busca adaptadas para cada base bibliográfica podem ser consultadas no Apêndice A.

3.2 Condução

O principal artefato da fase de planejamento da RSL é o protocolo da revisão, que documenta como esta será realizada. A fase de condução da revisão consiste, portanto, na

implementação do protocolo. A atividade de identificação dos estudos (ver Figura 5) põe em prática a estratégia de busca sobre as fontes de pesquisa, detalhadas na Seção 3.1.2. A atividade de seleção dos estudos, por sua vez, põe em prática a aplicação dos estágios de seleção: preliminar e final, expostos na Seção 3.1.3. A atividade de extração de dados é então realizada sobre os resultados de seleção e reúne um conjunto de dados relevantes para sumarização. A seguir, são apresentados os detalhes das atividades de condução.

3.2.1 Resultados de busca e seleção

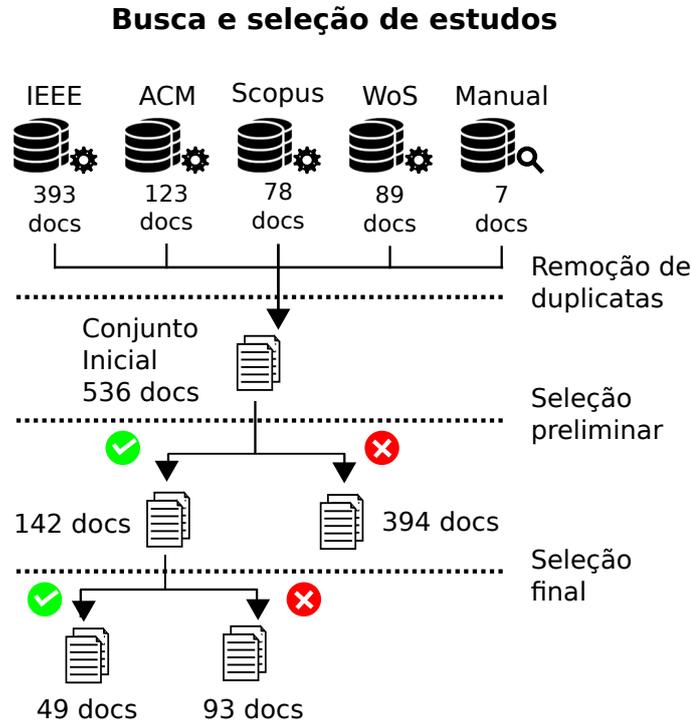
Logo após a identificação dos estudos primários nas fontes de pesquisa, o conjunto bruto de documentos recuperados é submetido à remoção de referências duplicadas, uma vez que múltiplas bases bibliográficas podem indexar um mesmo documento. O total de 536 documentos únicos permaneceram após a remoção automática de duplicatas, realizada com o apoio da ferramenta *Mendeley*⁶, e compõem o conjunto inicial de documentos candidatos à extração de dados. Conforme definido no protocolo, a seleção preliminar de documentos foi conduzida com base na leitura de apenas título e *abstract* dos 536 documentos do conjunto inicial. Cinco pesquisadores estiveram envolvidos nesse estágio e avaliaram a adequação de cada documento aos critérios de inclusão e exclusão previamente estabelecidos. Como resultado, 142 documentos foram aceitos neste estágio, isto é, indicados para o estágio seguinte de seleção. Durante o estágio de seleção final de estudos, os documentos foram avaliados por meio da leitura do texto na íntegra. Como resultado, o total de 49 estudos foram preservados para extração de dados. Na Figura 6 são sumarizados os procedimentos de busca e seleção.

3.2.2 Extração de dados

As evidências para responder às questões de pesquisa apresentadas na Seção 3.1.1 são finalmente coletadas por meio do preenchimento do formulário de extração de dados. As questões de pesquisa também serviram de guia para a elaboração da estrutura que acomoda os dados coletados dos estudos. O formulário é organizado em seis seções: (1) Metadados, (2) Representação da face, (3) Pré-processamento, (4) Dependência de *keypoints*, (5) Experimentos de classificação e (6) Resultados. Uma descrição detalhada dos campos do formulário de extração de dados é apresentada na Tabela 4.

⁶ www.mendeley.com

Figura 6 – Resumo dos procedimentos de busca e seleção de documentos.



Fonte: elaborado pela autora.

3.2.3 Sumarização dos dados

A partir do preenchimento do formulário de extração de dados, os elementos considerados relevantes para a caracterização dos estudos segundo os aspectos definidos no protocolo são sumarizados nas formas de tabelas e gráficos, que facilitam a análise dos dados coletados. A sumarização na forma de gráficos e tabelas oportuniza a elaboração de respostas diretas às questões de pesquisa da revisão. Adicionalmente, foram elaboradas discussões mais detalhadas sobre os dados coletados.

3.3 Considerações finais

O conjunto de 49 artigos resultado da condução da RSL representa uma amostra suficientemente representativa do universo bibliográfico em REF 3D sobre o qual serão elaboradas as respostas às questões de pesquisa definidas na Seção 3.1.1. O capítulo seguinte representa a terceira fase da RSL (ver Figura 5), fornecendo uma avaliação dos resultados obtidos, bem como as respostas objetivas às questões de pesquisa definidas no protocolo da revisão.

Tabela 4 – Formulário de extração de dados preenchido para cada estudo selecionado.

Dados extraídos	Descrição
1. Metadados	Título, autores e ano de publicação do estudo.
2. Representação da face	
2.1. Regiões de interesse	Grau de granularidade empregada para a definição da região de interesse. Por exemplo: superfícies, curvas, pontos.
2.2. Descritores	Os descritores calculados a partir das regiões de interesse e que compõem o vetor de atributos. Por exemplo: distância euclidiana entre pontos fiduciais.
2.3. Tipo de dados	natureza dos dados empregados para representar a face, a partir dos quais os descritores são calculados. Por exemplo: 3D, 2D+3D ou 4D.
3. Pré-processamento	Os passos de pré-processamento reportados no estudo investigado. Por exemplo: registro, remoção de <i>outliers</i> , etc.
4. Dependência de <i>keypoints</i>	
4.1. Número de <i>keypoints</i>	O número de <i>keypoints</i> requeridos no estudo (para qualquer propósito).
4.2. Modo de detecção	Os meios de definição desses pontos. Por exemplo: automaticamente pelos autores ou fornecidos no banco de imagens, etc.
4.3. Tipo de dados	A natureza dos dados a partir dos quais os <i>keypoints</i> são determinados. Por exemplo: 3D ou RGB.
4.4. Quais tarefas específicas são dependentes de <i>keypoints</i> ?	Por exemplo: pré-processamento, extração de atributos, etc.
5. Experimentos de classificação	
5.1. Banco de dados 3D de faces	O banco de dados 3D de faces utilizado para validar o método proposto no estudo.
5.2. Amostra de dados	O subconjunto de dados efetivamente empregado. Por exemplo: o número de indivíduos, os níveis de intensidade, a presença ou ausência de amostras neutras, etc.
5.3. Proporção treinamento/teste	Por exemplo: 90%/10%, 10-CV
5.4. Rodadas de classificação	Quantas vezes o classificador avaliou cada par treinamento e teste.
5.5. Classificador	O algoritmo para classificação. Por exemplo: SVM, k-NN, etc.
6. Resultados	A acurácia de reconhecimento atingida para cada classe: raiva, aversão, medo, felicidade, tristeza, surpresa, neutro.

Fonte: a autora.

4 RESULTADOS

Neste capítulo, são apresentados os resultados da avaliação dos dados coletados por meio da aplicação do formulário de extração de dados. A exposição detalhada dos resultados com respeito à representação da face, pré-processamento e experimentos de classificação é sucedida pelas respostas objetivas às questões de pesquisa.

4.1 Representação da face

Métodos tradicionais em REF 3D adotam abordagens particulares de representação da face que servem de entrada a um sistema automático de classificação. De fato, decisões sobre a forma de representação da face consistem em alguns dos aspectos mais importantes em tarefas para REF. A seguir, são sumarizados os dados mais relevantes sobre métodos tradicionais de representação da face identificados nos estudos consultados.

4.1.1 Tipos de dados e descritores

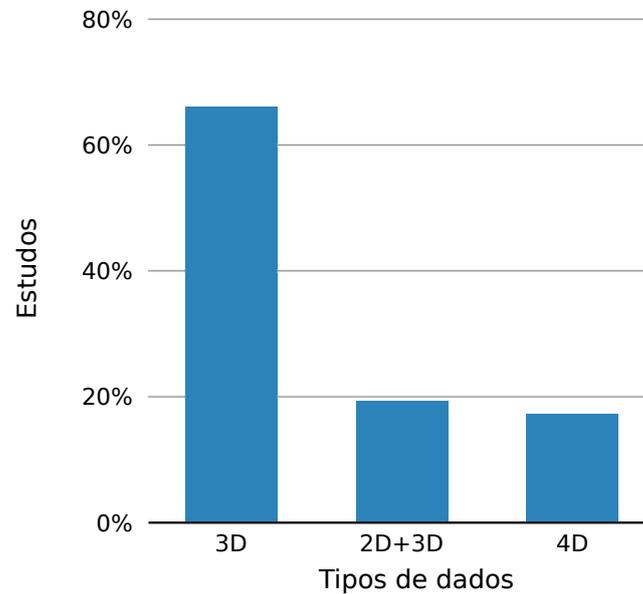
O problema de REF 3D é abordado na literatura por meio do uso de dados puramente espaciais (3D) ou por sua combinação com informação de cor e informação de variação de tempo. A denominação 2D+3D é dada às abordagens que combinam informação de imagens 2D e imagens 3D capturadas simultaneamente. Nesse caso, atributos extraídos a partir de informações de cor e textura se aliam a atributos extraídos de informações espaciais para compor um vetor de atributos. A abordagem denominada 4D emprega informação da variação percebida em sequências de *frames* ao longo do tempo e, por isso, são também chamados de vídeos 3D. A maioria dos estudos consultados propõem soluções estabelecidas apenas sobre informações espaciais de imagens 3D estáticas, em oposição às soluções combinadas, conforme apresentado no gráfico da Figura 7.

Os descritores estão diretamente relacionados aos tipos de dados disponíveis nas imagens que representam as faces. Nos estudos em que os métodos adotados fazem uso de dados puramente 3D, normalmente extraem-se descritores geométricos tais como valores de coordenadas, distâncias (euclidianas ou geodésicas) e curvaturas. Além disso, mesmo nas abordagens que utilizam unicamente dados 3D, as coordenadas geométricas da face são normalmente mapeadas em representações 2D, tais como mapas de profundidade e suas derivações: mapas de normais e mapas de índice de forma. Essa transformação permite a extração indireta de

Figura 7 – Tipos de dados disponíveis nas imagens empregadas nos estudos consultados.

Tipos de dados empregados para representação da face

Amostra de 49 estudos especializados em REF 3D



Nota: Dentre os estudos categorizados como 4D, os vídeos 3D texturizados também estão incluídos.

Fonte: elaborado pela autora.

descritores originalmente projetados para o domínio 2D a partir de dados 3D. Nos estudos em que as abordagens adotadas também lidam com informação de cor e textura, extraem-se, adicionalmente, descritores apropriados a esse tipo de dados, como *Local Binary Pattern* (LBP) e descritores baseados em *Scale-Invariant feature transform* (SIFT).

4.1.2 Regiões de interesse

No contexto de REF 3D, as RoI podem adotar os escopos global, local e híbrido, dos quais o último combina aspectos dos dois primeiros (SOLTANPOUR *et al.*, 2017; CORNEANU *et al.*, 2016). Na Tabela 5, os 49 estudos consultados são categorizados de acordo com o escopo da RoI do seu método de representação da face.

Métodos que fazem uso de RoI de escopo global consideram uma única região que normalmente compreende toda a superfície da face. Os métodos de escopo local utilizam um dos três tipos de regiões específicas sobre a face: *keypoints*, curvas e superfícies locais; ou ainda uma combinação delas, como em Sheng *et al.* (2016) que considera ambos *keypoints* e superfícies locais.

Tabela 5 – 49 estudos consultados categorizados de acordo com os escopos de suas RoI.

Global	Local			Híbrido
	Keypoints	Curvas	Superfícies	
An e Ruan (2016)	Aly <i>et al.</i> (2014)	Amor <i>et al.</i> (2014)	Berretti <i>et al.</i> (2013)	Aly <i>et al.</i> (2016)
Gaoyun An <i>et al.</i> (2013)	Azazi <i>et al.</i> (2014a)	Maalej <i>et al.</i> (2013)	Derkach e Sukno (2017)	Jan e Hongying Meng (2015)
Fu <i>et al.</i> (2016)	Azazi <i>et al.</i> (2014b)	Samad e Iftekharruddin (2016)	Han e Ming (2014)	
Li <i>et al.</i> (2017b)	Azazi <i>et al.</i> (2015)	Zhen <i>et al.</i> (2017)	Hariri <i>et al.</i> (2017)	
Li <i>et al.</i> (2017a)	Elhoufi <i>et al.</i> (2016)		Hayat <i>et al.</i> (2013)	
Savran e Sankur (2017)	Jazouli <i>et al.</i> (2017)		Lemaire <i>et al.</i> (2013)	
Trimech <i>et al.</i> (2016)	Khashman e Conkbayir (2013)		Li <i>et al.</i> (2015)	
Zeng <i>et al.</i> (2013)	Suja <i>et al.</i> (2016)		Li <i>et al.</i> (2015)	
	Tao e Matuszewski (2016)		Moeini <i>et al.</i> (2016)	
	Yurtkan e Demirel (2014b)		Ocegueda <i>et al.</i> (2013)	
	Yurtkan e Demirel (2014a)		Reale <i>et al.</i> (2013)	
	Yurtkan e Demirel (2013)		Vieriu <i>et al.</i> (2015)	
	Yurtkan <i>et al.</i> (2013)		Xue <i>et al.</i> (2015)	
	Zarbaksh e Demirel (2017)		Xue <i>et al.</i> (2014)	
	Zhang <i>et al.</i> (2015)		Xudong Yang <i>et al.</i> (2015)	
	Sheng <i>et al.</i> (2016)		Yao <i>et al.</i> (2018)	
			Zhen <i>et al.</i> (2016)	
			Sheng <i>et al.</i> (2016)	

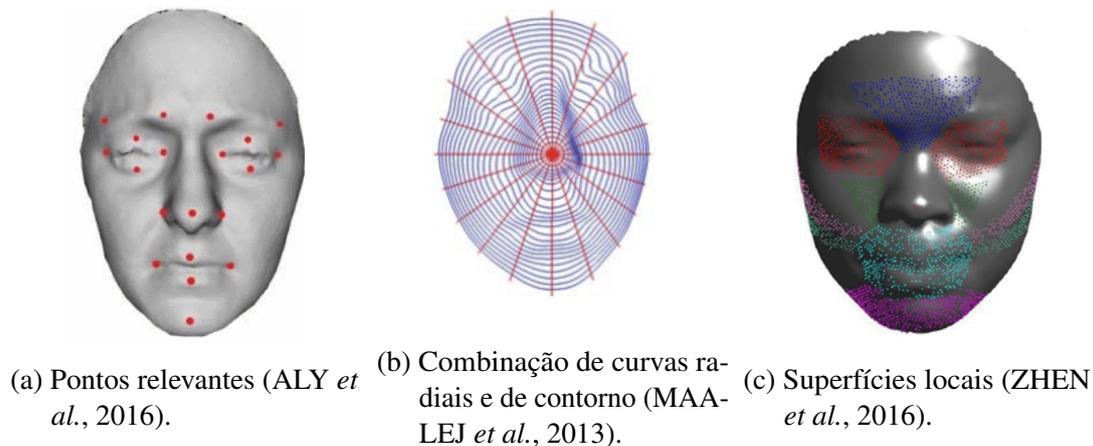
Nota: Sheng *et al.* (2016) aparece duas vezes por combinar dois tipos de RoI locais: *keypoints* e superfícies.
Fonte: elaborado pela autora.

Keypoints são pontos relevantes na superfície da face, cuja relevância pode ser determinada por diferentes métodos. Comumente são considerados *keypoints* pontos característicos de elementos faciais, entretanto, esses pontos também podem ser determinados a partir de critérios geométricos de relevância (RODOLÀ *et al.*, 2015). Métodos locais baseados em *keypoints* utilizam atributos extraídos diretamente desses pontos, como por exemplo, suas coordenadas espaciais (ELHOUI *et al.*, 2016); as distâncias euclidianas entre *keypoints* (ZARBAKSH; DEMIREL, 2017); as distâncias entre pares dos *keypoints* e os ângulos entre esses pontos (SUJA *et al.*, 2016). Tao e Matuszewski (2016) definem um vetor de atributos pelas diferenças entre as posições de *keypoints* nas faces expressivas e seus pontos correspondentes em uma face referencial neutra. Esses trabalhos têm em comum a dependência de uma etapa de detecção *a priori* de pontos relevantes para seu método de representação da face.

Métodos locais baseados em superfícies utilizam estratégias para determinar as ROI baseados em grades regulares ou irregulares ou na vizinhança de *keypoints*. Como um exemplo do primeiro caso, Yao *et al.* (2018) selecionam alguns *frames* de vídeos 3D texturizados, de onde são extraídas regiões locais a partir de uma grade regular aplicada a ambos os tipos de dados: 2D e 3D. Especificamente, descritores LBP são extraídos de seções dos quadros das imagens 2D e coeficientes de dispersão são extraídos de seções de quadros 3D, de maneira a combinar características relevantes de ambos os tipos de dados. Hariri *et al.* (2017), por sua vez, determinam *keypoints* como referências para delimitar superfícies locais de interesse, de tal forma que 30 pontos distribuídos uniformemente sobre os *scans* de faces 3D representam os centros de 30 superfícies locais com sobreposição.

Menos frequentemente, métodos de escopo local baseados em curvas representam a face por um conjunto de curvas sobre sua superfície. Essas curvas são geralmente dos tipos: radial, de um ponto central a uma extremidade da face; ou de contorno, fechadas ao redor de um ponto central.

Figura 8 – Exemplos de regiões de interesse extraídas em métodos locais de representação da face.



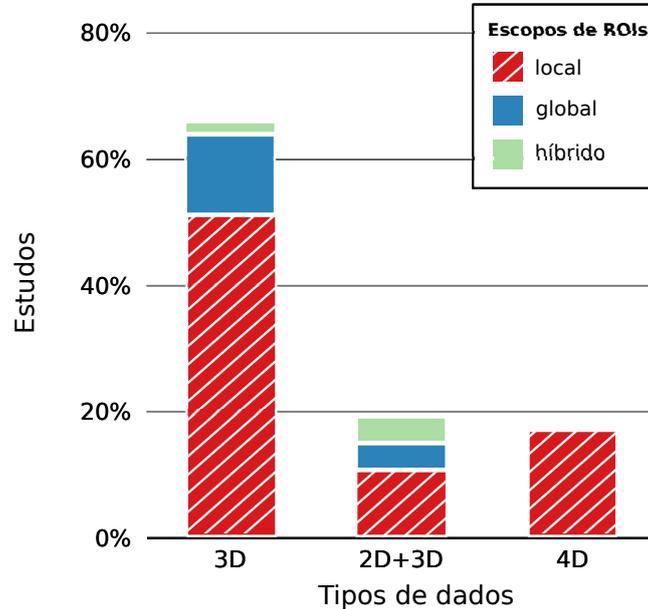
Fonte: elaborado pela autora.

A maioria dos estudos consultados neste trabalho adotam, como forma de representação da face, regiões de escopo local das quais os atributos são extraídos utilizando-se dados puramente 3D. Na Figura 9 são destacadas as proporções de escolha por diferentes escopos de ROI para cada tipo de dados disponível nas imagens empregadas.

Figura 9 – Proporções dos escopos de ROI empregadas por tipo de dados disponíveis nas imagens.

Frequências dos tipos de dados e regiões de interesse para representação da face

Amostra de 49 estudos especializados em REF 3D



Nota: Dentre os estudos categorizados como 4D, os vídeos 3D texturizados também estão incluídos.

Fonte: elaborado pela autora.

4.2 Pré-processamento

Dentre os estudos consultados, as principais técnicas de pré-processamento aplicadas aos *scans* 3D de faces são: (1) correção de pose, (2) corte ou segmentação da face, (3) suavização e (4) preenchimento de buracos. Menos frequentemente, técnicas como re-amostragem e mapeamentos 3D-2D são reportados como pré-processamentos. Múltiplas técnicas de pré-processamento são frequentemente empregadas em um mesmo estudo, como mostrado na Tabela 6, em que todos os estudos consultados que reportam técnicas de pré-processamento são listados e as técnicas, identificadas. A seguir, são apresentados os detalhes da ocorrência dessas abordagens no estudos consultados.

4.2.1 Correção de pose

O algoritmo *Iterative Closest Point* (ICP) é o mais frequentemente mencionado para correção de pose (BESL; MCKAY, 1992). O ICP realiza o registro rígido de objetos 3D ao calcular iterativamente a matriz de rotação e o vetor de translação mais adequados para a

Tabela 6 – Estudos que reportam a aplicação de técnicas para o pré-processamento dos *scans* 3D de faces.

Estudos	Correção de pose	Segmentação da face	Suavização	Preenchimento de buracos	Outros
Li <i>et al.</i> (2015)	✓	✓	✓	✓	✓
Fu <i>et al.</i> (2016)	✓	✓	✓	✓	✗
Amor <i>et al.</i> (2014)	✓	✓	✗	✓	✗
Li <i>et al.</i> (2017a)	✓	✓	✗	✗	✓
Lemaire <i>et al.</i> (2013)	✓	✓	✗	✗	✓
Zhen <i>et al.</i> (2017)	✓	✓	✗	✗	✓
Hayat <i>et al.</i> (2013)	✓	✓	✗	✗	✗
Maalej <i>et al.</i> (2013)	✓	✓	✗	✗	✗
Zhen <i>et al.</i> (2016)	✓	✗	✓	✓	✗
Sheng <i>et al.</i> (2016)	✓	✗	✓	✓	✗
Xue <i>et al.</i> (2014)	✓	✗	✗	✗	✓
Aly <i>et al.</i> (2014)	✓	✗	✗	✗	✓
Zeng <i>et al.</i> (2013)	✓	✗	✗	✗	✓
Reale <i>et al.</i> (2013)	✓	✗	✗	✗	✓
Ocegueda <i>et al.</i> (2013)	✓	✗	✗	✗	✓
Yao <i>et al.</i> (2018)	✓	✗	✗	✗	✗
Savran e Sankur (2017)	✓	✗	✗	✗	✗
Berretti <i>et al.</i> (2013)	✗	✓	✓	✓	✓
Samad e Iftekharuddin (2016)	✗	✓	✓	✓	✓
Hariri <i>et al.</i> (2017)	✗	✓	✓	✓	✗
Gaoyun An <i>et al.</i> (2013)	✗	✓	✓	✗	✓
Xue <i>et al.</i> (2015)	✗	✓	✓	✗	✓
Azazi <i>et al.</i> (2015)	✗	✗	✓	✓	✓
Trimech <i>et al.</i> (2016)	✗	✗	✓	✗	✓
Azazi <i>et al.</i> (2014b)	✗	✗	✗	✗	✓
Xudong Yang <i>et al.</i> (2015)	✗	✗	✗	✗	✓
Derkach e Sukno (2017)	✗	✗	✗	✗	✓
Vieriu <i>et al.</i> (2015)	✗	✗	✗	✗	✓
Azazi <i>et al.</i> (2014a)	✗	✗	✗	✗	✗

Nota: As colunas estão ordenadas, da esquerda para a direita, em ordem decrescente de frequência de uso.

Fonte: elaborado pela autora.

correspondência de dois conjuntos de pontos. A eficácia do ICP é afetada pela posição inicial dos conjuntos de pontos, bem como por parâmetros de ajuste como, por exemplo, o número máximo de iterações, critério de parada, critério de aceitação de correspondência entre pares de pontos. Os pontos utilizados para o cálculo da correspondência podem ser todos os pontos da superfície da face, bem como *keypoints* ou pontos da vizinhança de *keypoints* (ZHEN *et al.*, 2016; XUE *et al.*, 2014; LI *et al.*, 2015; YAO *et al.*, 2018; SAVRAN; SANKUR, 2017; AMOR *et al.*, 2014; FU *et al.*, 2016). Outras técnicas para correção de pose encontradas nos estudos consultados são transformação de Möbius para a normalização de mapas conformes (ZENG *et al.*, 2013), correção de pose baseada em *Principal Component Analysis* (PCA) (HAYAT *et al.*, 2013) e a superposição de Procrustes (ALY *et al.*, 2014).

Apesar de sua identificação entre os estudos consultados, as técnicas de pré-processamento

não são suficientemente descritas. Por exemplo, os estudos que reportam o uso do ICP para registro dos *scans* 3D não fornecem informações sobre os parâmetros utilizados. Além disso, é comum que as técnicas de pré-processamento sejam referidas apenas por termos genéricos como "alinhamento", "normalização" ou "registro das faces" sem que o nome da técnica ou seus detalhes sejam fornecidos.

Ainda, aproximadamente 37% dos estudos consultados neste trabalho não mencionam qualquer pré-processamento aplicado aos dados. Esses estudos, entretanto, não deixam claro se esse tratamento prévio aos dados é de fato dispensável. A exceção é o relato de Xudong Yang *et al.* (2015), em que é explicitamente posto que nenhuma técnica de pré-processamento foi empregada, uma vez que os dados disponíveis na base de dados BU-3DFE já possuíam qualidade satisfatória para seus propósitos.

4.2.2 Corte ou segmentação da face

A segmentação da face, também denominada corte da região central ou *crop*, visa a remoção de pontos periféricos do modelo 3D, que são ruidosos e redundantes para a classificação, como cabelos, orelhas, pescoço e, em alguns casos, o busto. O corte esférico é o mais comumente reportado; consiste na rejeição das regiões que se localizam além de uma vizinhança esférica de um certo ponto de referência, comumente a ponta do nariz. Para o raio de corte, costuma-se empregar medidas entre 70mm e 90mm. A segmentação da face é um procedimento necessário e suficiente para a definição da região de interesse em métodos de escopo global. Entretanto, para os métodos de escopo local, pode corresponder a uma etapa preliminar, precedendo a extração de outras regiões de interesse menores.

A exemplo do que ocorre com a correção de pose, os autores também costumam omitir informações importantes sobre a segmentação da face, empregando apenas denominações genéricas, como 'corte', para se referir a todo o procedimento empregado neste pré-processamento, o que não é suficiente para descrevê-lo. Os estudos que reportam a segmentação da face como pré-processamento são elencados na Tabela 7.

4.2.3 Suavização

A suavização da superfície é aplicada para reduzir o efeito de ruídos que são próprios da aquisição dos dados pelos sensores. Como no caso dos anteriores, esse procedimento não é frequentemente detalhado nos estudos sobre REF 3D. Para descrever esse pré-processamento, os

Tabela 7 – Porção dos estudos consultados que reportam segmentação da face como parte do seu pré-processamento.

Estudos	Corte esférico?	Nariz no centro?	Raio de corte
Hayat <i>et al.</i> (2013)	Sim	Sim	85mm
Berretti <i>et al.</i> (2013)	Sim	Sim	90mm
Li <i>et al.</i> (2017a)	-	-	-
Hariri <i>et al.</i> (2017)	-	-	-
Amor <i>et al.</i> (2014)	Sim	Sim	90mm
Lemaire <i>et al.</i> (2013)	Sim	Sim	80mm
Zhen <i>et al.</i> (2017)	-	-	-
Fu <i>et al.</i> (2016)	Sim	Sim	90mm
Gaoyun An <i>et al.</i> (2013)	Sim	Sim	-
Maalej <i>et al.</i> (2013)	-	-	-
Xue <i>et al.</i> (2015)	Sim	Sim	70mm
Samad e Iftekharuddin (2016)	Não	Não	NA
Li <i>et al.</i> (2015)	Sim	Sim	distância entre ponta do nariz e queixo

Fonte: elaborado pela autora.

estudos consultados mencionam somente o uso de filtros de mediana (LI *et al.*, 2015) (SHENG *et al.*, 2016) (Gaoyun An *et al.*, 2013), filtros gaussianos (BERRETTI *et al.*, 2013) (AZAZI *et al.*, 2015), filtros de média (SAMAD; IFTEKHARUDDIN, 2016) e operador Laplaciano (TRIMECH *et al.*, 2016) ou, de forma ainda mais genérica, utilizam apenas o termo filtros de suavização (ZHEN *et al.*, 2016) (FU *et al.*, 2016) (HARIRI *et al.*, 2017). Dentre os estudos consultados, apenas em (XUE *et al.*, 2015), os autores fornecem algumas informações relevantes para a reprodução do procedimento de suavização. Nesse estudo, aplica-se uma filtragem baseada na média dos pontos de cada face. Os pontos posicionados além de 5 desvios padrões são considerados *outliers* e removidos.

4.2.4 Preenchimento de buracos

O preenchimento de buracos é o quarto pré-processamento mais frequente nos estudos consultados em REF 3D. A interpolação cúbica é a técnica mais mencionada para esse propósito, sendo empregada em Zhen *et al.* (2016), Li *et al.* (2015), Li *et al.* (2015), Berretti *et al.* (2013), Sheng *et al.* (2016), Azazi *et al.* (2015) e Samad e Iftekharuddin (2016). Nenhum dos estudos, entretanto, forneceu maiores informações sobre esse procedimento.

4.2.5 Detecção de pontos relevantes (*keypoints*)

Muito embora a detecção de pontos relevantes nem sempre seja reportada como um procedimento de pré-processamento, nos estudos baseados nesse tipo de técnica, a detecção *a priori* desses pontos é necessária para apoiar duas principais tarefas: extração de atributos e outros procedimentos de pré-processamento, sendo portanto pertinente a esses resultados. No primeiro caso, a dependência ocorre principalmente em métodos de representação da face de escopo local baseados em *keypoints*, nos quais os atributos são extraídos diretamente de coordenadas desses pontos ou suas derivações, conforme detalhado na Seção 4.1.2. Apesar da predominância em métodos locais baseados em *keypoints*, métodos de escopo local baseados em superfícies e curvas também podem depender da detecção de pontos relevantes, como os propostos por Amor *et al.* (2014) e Zhen *et al.* (2017). Mesmo em técnicas de representação da face cuja extração de atributos não depende diretamente da detecção de *keypoints*, esse procedimento ainda pode ser necessário para outras técnicas de pré-processamento, como por exemplo a segmentação da face em que a posição da ponta do nariz se faz necessária (ver Tabela 7).

A dependência da detecção de *keypoints* foi verificada em 78% dos estudos consultados, variando de 1 ponto a 121 pontos. Na Figura 10 essa dependência é quantificada em 12 intervalos. Esse arranjo revela que mais 50% dos métodos consultados dependem da detecção de um número superior a 10 pontos característicos.

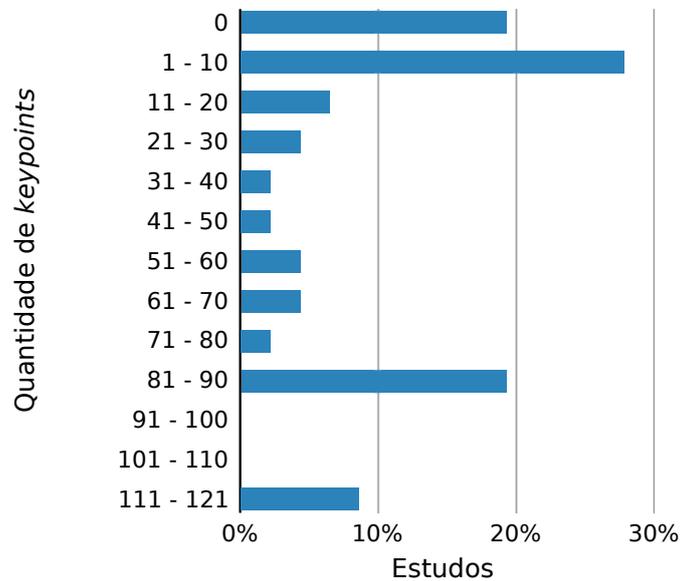
A análise dos dados extraídos ainda revelou que os métodos que requerem a detecção de pontos relevantes, de modo geral, utilizam como critério de relevância características anatômicas da face, apoiando-se em pontos cujas posições possuem um significado conhecido. Exceto por Hariri *et al.* (2017), todos os pontos requeridos nos estudos consultados correspondem a pontos fiduciais faciais, que são aqueles comumente usados para identificar elementos como cantos dos olhos, boca, sobrancelhas, ponta do nariz, dentre outros.

Pontos relevantes podem ser detectados automaticamente ou marcados manualmente sobre modelos de faces. A detecção automática é predominante entre os estudos que requerem *keypoints*, como apresentado na Figura 11. Muitos estudos dispõem, convenientemente, de algoritmos já consolidados no domínio 2D para a detecção de pontos fiduciais faciais, realizando a detecção desses pontos em mapeamentos 3D-2D. Por exemplo, Azazi *et al.* (2014b) utilizam uma representação 2D de *scans* 3D por meio de mapas conformes e aplicam o algoritmo *Structured Output Support Vector Machine* (SO-SVM) com *Deformable Part Model* (DPM) para

Figura 10 – Número de pontos característicos requeridos pelos 49 estudos consultados.

Quantidade de *keypoints* requeridos em REF 3D

Amostra de 49 estudos especializados em REF 3D



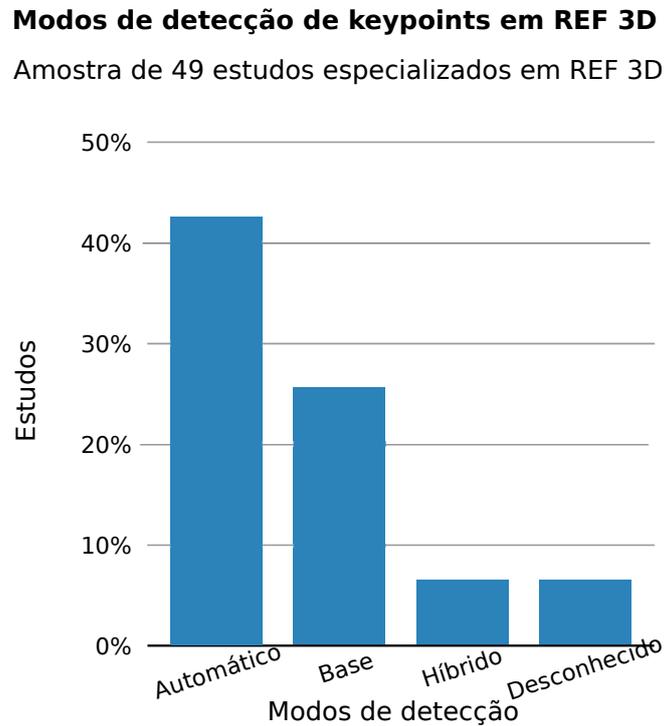
Fonte: elaborado pela autora.

detecção de pontos fiduciais. Xue *et al.* (2014), por sua vez, empregam o algoritmo *Haar-cascade* sobre mapas de profundidade e seus gradientes com o mesmo objetivo. Alternativamente, Li *et al.* (2015), Jan e Hongying Meng (2015), Azazi *et al.* (2015) e Xue *et al.* (2015) realizam a detecção automática de pontos fiduciais faciais nas imagens 2D e os mapeiam para os seus modelos 3D correspondentes.

Métodos dependentes de marcação manual apresentam soluções chamadas semiautomáticas, que, para serem aplicadas em problemas concretos, presumem uma etapa prévia e independente de detecção desses pontos. Nenhum dos estudos consultados apoia-se exclusivamente na marcação manual de pontos. Por outro lado, uma porção significativa de estudos utilizam a marcação fornecida pelas bases de dados 3D que escolheram para validação. Esse tipo de dependência também torna o método semiautomático.

Métodos híbridos de identificação de pontos relevantes combinam marcação manual com detecção automática. Métodos com essa característica foram identificados em alguns poucos estudos. Por exemplo, Zhen *et al.* (2016) realizam a marcação manual de 71 pontos fiduciais faciais, usados posteriormente para a delimitação de 11 RoI sobre um único *scan* modelo. Os segmentos definidos na face-modelo são usados para segmentar regiões equivalentes em todas

Figura 11 – Ocorrência dos modos de detecção de pontos relevantes nos estudos consultados.



Fonte: elaborado pela autora.

as demais faces usando o algoritmo *Iterative Closest Normal Point* (ICNP). No mesmo estudo, para normalização da pose, a ponta do nariz de cada face é identificada automaticamente como o ponto de maior magnitude da coordenada z.

Zhen *et al.* (2017), Fu *et al.* (2016) e Gaoyun An *et al.* (2013) reportam a dependência de apenas um ponto relevante: a ponta do nariz, mas não mencionam seu método de detecção, portanto classificado como desconhecido.

4.3 Experimentos de classificação

Além do algoritmo de classificação automática, um experimento de classificação em REF 3D se caracteriza pela combinação de aspectos relacionados à amostra de dados e às estratégias de classificação. Os principais resultados da investigação desses aspectos nos estudos consultados são apresentados a seguir.

4.3.1 Amostra de dados

Os *scans* faciais empregados nos experimentos de classificação analisados se restringem àqueles disponíveis em quatro bases de dados públicas e duas bases de dados privadas.

Cada uma delas apresenta particularidades em sua composição que influenciam no projeto experimental. Na Tabela 8 são listados alguns elementos relevantes da composição dessas bases. A seguir, são apresentados os resultados relacionados a amostra de dados discriminados pela base de dados de origem.

Tabela 8 – Bases de dados utilizadas para validação nos estudos consultados.

Bases de dados	Tipo de dados	Composição	Total de amostras
BU-3DFE (YIN <i>et al.</i> , 2006)	RGB + 3D	100S × (6E + 1N)	2.500 <i>scans</i>
Bosphorus (SAVRAN <i>et al.</i> , 2008)	RGB + 3D	105S × (6E + 4N)	4.666 <i>scans</i>
BU-4DFE (YIN <i>et al.</i> , 2008)	RGB + 4D	101S × 6E	606 sequências (60.600 <i>frames</i>)
VT-KFER (ALY <i>et al.</i> , 2015)	RGB + 4D	32S × (6E + 1N)	32 sequências (61.374 <i>frames</i>)
Elhoufi <i>et al.</i> (ELHOULFI <i>et al.</i> , 2016)	3D	10S × 6E	60 <i>scans</i>
Aly <i>et al.</i> (ALY <i>et al.</i> , 2014)	3D	10S × 6E	17.000 <i>scans</i>

Nota: A letra S acompanha a quantidade de indivíduos, enquanto E e N acompanham, respectivamente, a quantidade de expressões prototípicas e *scans* neutros disponíveis na base.

Fonte: elaborado pela autora.

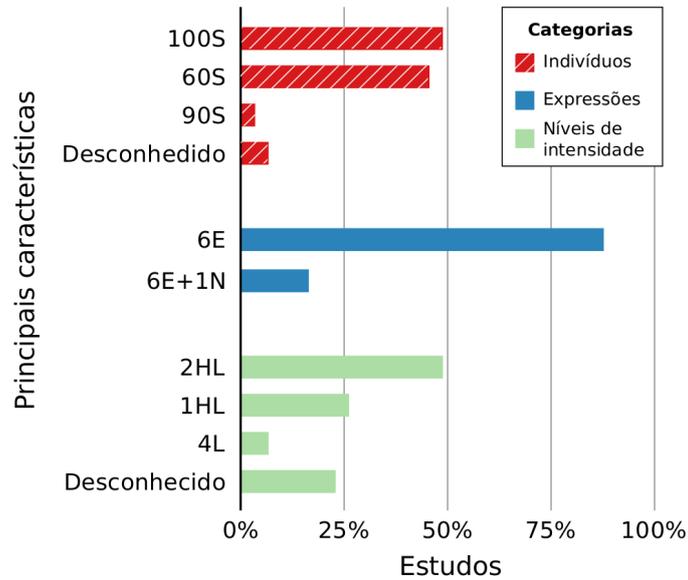
4.3.1.1 BU-3DFE

Os *scans* faciais provenientes da BU-3DFE foram utilizados em 63% dos estudos consultados, sendo essa a base de dados mais empregada. Apesar de dispor de 2500 *scans* faciais balanceados entre seis classes de emoções e graduados em quatro níveis de intensidade, em geral, apenas subconjuntos desses dados são empregados. Esses subconjuntos variam quanto ao número de indivíduos, expressões e intensidades. A análise desses fatores isoladamente revela que há predominância de amostras de dados compostas por todos os 100 indivíduos da base, 6 classes de emoções e 2 níveis mais intensos de expressividade. Na Figura 12 são apresentadas as ocorrências desses aspectos. Essas estatísticas, no entanto, consideram esses fatores descorrelacionados. Na Figura 13, combinadas, essas características descrevem completamente os perfis de amostras de dados mais frequentemente utilizadas a partir dessa base. O perfil mais frequente de amostras de dados da BU-3DFE é aquele composto pelos *scans* dos dois níveis mais expressivos (2HL) das seis expressões prototípicas (6E) de 60 indivíduos (60S). Mesmo assim, além da notória heterogeneidade dos dados empregados, aproximadamente 21% dos estudos consultados não mencionam a amostra de dados empregada para a validação dos resultados de classificação.

Figura 12 – A frequência das principais características de amostras de dados empregadas em experimentos com a BU-3DFE.

Principais características das amostras de dados em experimentos com BU-3DFE

Amostras de 32 estudos especializados em REF 3D

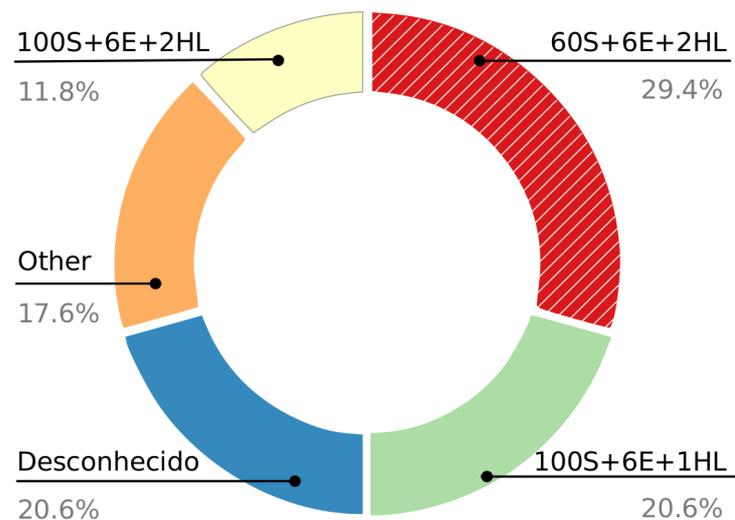


Fonte: elaborado pela autora.

Figura 13 – Amostras de dados mais frequentes da BU-3DFE

Perfis das amostras de dados empregadas em experimentos com BU-3DFE

Amostra de 32 estudos especializados em REF 3D



Fonte: elaborado pela autora.

4.3.1.2 *Bosphorus*

A base de dados Bosphorus é a segunda mais frequente para validação em REF 3D. Diferentemente da BU-3DFE, que em muitos estudos é a única base adotada, os *scans* da Bosphorus são frequentemente empregados para formar um segundo conjunto de dados para validação, de tal maneira que os experimentos com essa base são realizados complementarmente aos experimentos com alguma outra base, como por exemplo em Li *et al.* (2015), Li *et al.* (2017a), Hariri *et al.* (2017) e Jan e Hongying Meng (2015). Dentre os estudos que utilizam a Bosphorus, apenas em Zhang *et al.* (2015), Azazi *et al.* (2014b) e Trimech *et al.* (2016), essa base foi a única empregada.

Uma vez que a Bosphorus não oferece indicação de intensidade das expressões faciais, apenas o total de indivíduos e de expressões são considerados elementos suficientes para descrever completamente a amostra de dados. Apesar dessa base incluir os *scans* de 105 indivíduos, apenas 65 deles possuem amostras de todas as expressões prototípicas. Por essa razão, os estudos em REF 3D que empregam seus dados, usualmente, consideram ou todas as amostras dos 65 indivíduos, como Xue *et al.* (2014), Hariri *et al.* (2017) e Trimech *et al.* (2016); ou um subconjunto delas, como Li *et al.* (2015), Li *et al.* (2017a) e Azazi *et al.* (2015). Em 30% dos estudos, a composição da amostra de dados não foi mencionada. Nos estudos consultados, o uso de *scans* de expressão neutra além dos expressivos (6E+1N) é tão frequente quanto o uso de unicamente *scans* expressivos (6E).

4.3.1.3 *BU-4DFE*

A BU-4DFE é a base de dados empregada em apenas 9 dos estudos consultados. Conforme sumarizado na Tabela 8, a BU-4DFE é uma base de vídeos 3D e, portanto, consiste de sequências de *frames* 3D. A amostra de dados empregada em experimentos com essa base é caracterizada pelo número de indivíduos, expressões e conjunto de *frames* selecionados. Apenas experimentos empregando 100 indivíduos (7 estudos) e 60 indivíduos (2 estudos) foram identificados. Somente Zhen *et al.* (2016) e Berretti *et al.* (2013) utilizam as sequências completas providas pela base. Todos os demais empregam *frames*-chave ou subsequências. A amostra de dados mais frequente, portanto, é a formada por *frames*-chave, selecionados a partir de amostras das 6 expressões prototípicas de 60 indivíduos selecionados aleatoriamente, a assim chamada amostra 60S+6E+KF.

4.3.1.4 Outras bases de dados

Outras bases de dados foram identificadas nos experimentos de classificação de apenas 3 estudos consultados. A base de dados de vídeos 3D VT-KFER é empregada por Aly *et al.* (2016). Nesse estudo, apenas o número de expressões (6E) e suas intensidades (2HL) são reportadas. Elhoufi *et al.* (2016) e Aly *et al.* (2014), realizam a aquisição das duas próprias bases de dados, ambas com composição 10S+6E.

4.3.2 Classificadores e estratégia de classificação

O emprego de SVM como algoritmo classificador foi verificado em experimentos de 57% dos estudos consultados. O seu uso é normalmente justificado pelo bom desempenho em situações de dados limitados e espaço de atributos de altas dimensões (KHASHMAN; CONKBAYIR, 2013) (HARIRI *et al.*, 2017) (YURTKAN; DEMIREL, 2014b). De fato, o SVM tem registrado taxas de classificação superiores às taxas obtidas em condições equivalentes por outros métodos tradicionais de classificação, como demonstrado por Azazi *et al.* (2014a) e An e Ruan (2016). No pior caso, o SVM alcança taxas inferiores, mas ainda suficientemente próximas àquelas alcançadas, em experimentos similares, com *Hidden Markov Model* (HMM) e *Fisher's Linear Discriminant Analysis* (FLDA), utilizados, respectivamente, em Zhen *et al.* (2017) e Derkach e Sukno (2017). Outros algoritmos para classificação automática foram menos frequentemente encontrados nos estudos consultados, tais como *Random Forest* (RF), empregado em Amor *et al.* (2014) e Tao e Matuszewski (2016), *Nearest neighbors* (NN), aplicado em Fu *et al.* (2016), Jazouli *et al.* (2017) e Xue *et al.* (2015), *K-nearest neighbors* (K-NN), empregado em Azazi *et al.* (2014a), An e Ruan (2016), Han e Ming (2014) e, finalmente redes neurais artificiais, empregadas em Elhoufi *et al.* (2016) e Suja *et al.* (2016).

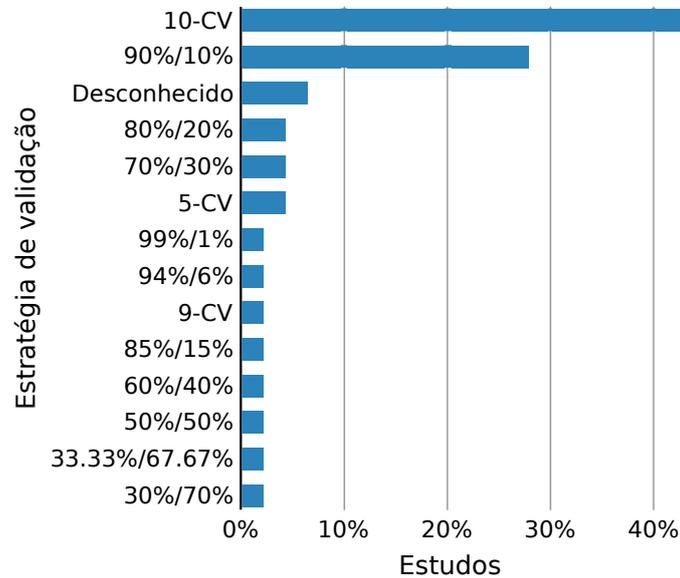
No que se refere à estratégia de classificação, 43% dos estudos consultados reportaram a realização de validação cruzada *10-fold* para estimar a efetividade do modelo preditivo, sendo a técnica mais frequentemente adotada. No gráfico da Figura 14 são apresentadas todas as estratégias de validação identificadas nos experimentos de classificação, no que diz respeito às proporções de treinamento e teste.

Adicionalmente, o particionamento de dados seguido da classificação automática pode ser repetido para análise estatística do classificador em que a média das taxas de reconhecimento ao longo de múltiplas rodadas de classificação é reportada. O número de rodadas de

Figura 14 – Proporções de treinamento e teste mais frequentes em REF 3D.

Estratégias de validação empregadas em REF 3D

Amostra de 49 estudos especializados em REF 3D



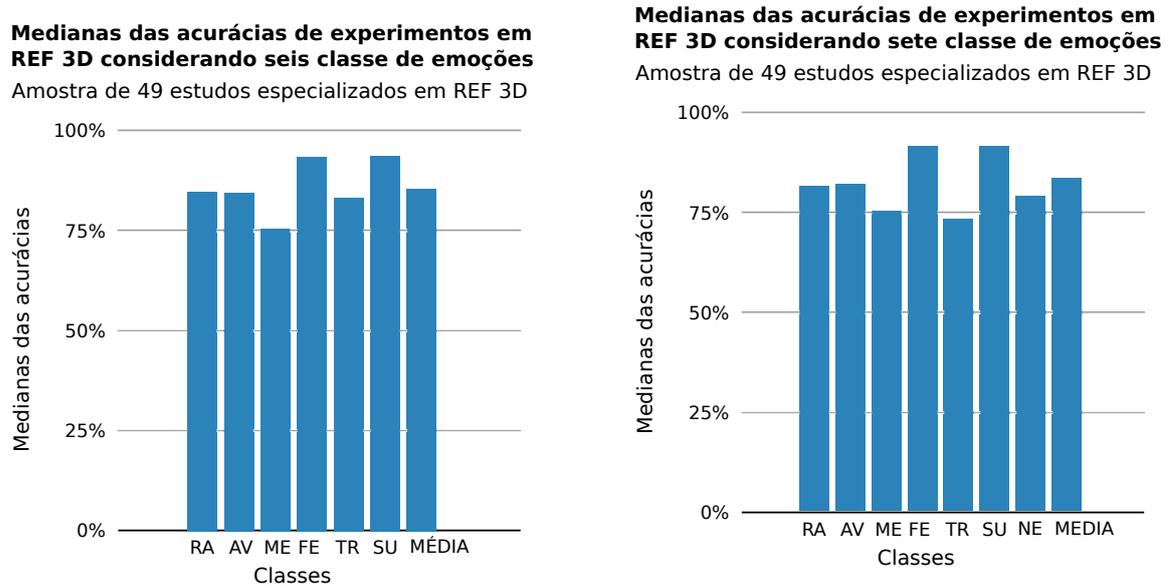
Fonte: elaborado pela autora.

classificação adotadas nos experimentos variam de uma única execução a 1000 repetições. A única execução da estratégia de validação cruzada *10-fold* é a configuração mais frequente e, mesmo assim, presente em apenas 17% dos estudos, dada a heterogeneidade das estratégias empregadas. Ainda 25% dos trabalhos omitem essa informação.

A avaliação do modelo preditivo é normalmente associada a avaliação do próprio método proposto. A acurácia média da classificação automática é a métrica mais amplamente utilizada para reportar o desempenho de classificação e é frequentemente mencionada como taxa de reconhecimento ou *recognition rate* (RR). Os valores das medianas das múltiplas RR obtidas por classe de emoções evidenciam as expressões de felicidade e surpresa como as mais facilmente distinguíveis das demais, bem como as expressões de medo e tristeza como as mais desafiadoras. Além disso, as acurácias médias alcançadas para felicidade e surpresa são altamente consistentes, apresentando baixos valores de desvio padrão, enquanto as demais expressões apresentam maiores flutuações. Na Figura 15 são apresentadas as medianas das acurácias reportadas em todos os estudos consultados, discriminadas por emoção.

Finalmente, na Tabela 9 os estudos são categorizados quanto aos aspectos que descrevem completamente seus experimentos de classificação. Por conveniência de apresentação,

Figura 15 – Medianas das acurácias reportadas em REF 3D, discriminadas por emoção.



(a) Universo de experimentos que consideram seis classes de emoções.

(b) Universo de experimentos que consideram sete classes de emoções: RA, AV, ME, FE, TR, SU e NE

Fonte: elaborado pela autora.

apenas os quatro experimentos de maiores RR são listados por categoria. A lista completa de experimentos, bem como todos os demais dados de experimentos extraídos podem ser consultados no Apêndice B.

4.4 Considerações finais

Os resultados obtidos a partir da avaliação dos dados extraídos dos estudos consultados permitem a caracterização dos estudos em REF 3D quanto às decisões para representação da face, pré-processamento e experimentos de classificação. Na Tabela 10 é apresentada a síntese dos principais resultados, que, aliada aos gráficos e tabelas indicados ao longo deste capítulo, compõem o conjunto de informações necessárias para responder objetivamente às questões de pesquisa definidas na Seção 3.1.1. Essas respostas são apresentadas a seguir:

QP1 - Quais estratégias para representação da face são mais frequentemente utilizadas em REF 3D?

Mais frequentemente, os *scans* de faces 3D são representados por meio de descritores extraídos de dados puramente espaciais, desconsiderando informações de cor, textura ou relação temporal que tenham, eventualmente, sido adquiridas adicionalmente. Ainda, são

majoritariamente consideradas regiões de escopo local para a efetiva extração desses atributos.

QP2 - Quais pré-processamentos têm sido requeridos para o tratamento de dados 3D com o propósito de REF?

Diversos procedimentos são empregados com o objetivo de preparar os *scans* 3D de faces para representação e experimentos de classificação. Dentre eles, destacam-se correção de pose, frequentemente realizada por meio do algoritmo ICP, segmentação da face, por meio do corte esférico baseado na ponta do nariz, suavização com filtros espaciais e preenchimento de buracos com interpolação cúbica. A detecção automática de pontos relevantes também é requisito em grande parte dos estudos consultados, mas em geral, é realizada sobre representações 2D dos *scans* 3D.

QP3 - Quais protocolos têm sido seguidos para a realização dos experimentos de classificação em REF 3D?

Apesar do número limitado de *scans* 3D de faces disponíveis para REF, os experimentos de classificação nesse domínio são muito heterogêneos, mesmo quando sobre os dados da mesma base. A principal amostra de dados empregada é a 60S+6E+2HL da base BU-3DFE. Uma única execução da validação cruzada *10-fold* é a estratégia mais popular para estimativa de desempenho do modelo preditivo.

Tabela 9 – Bases de dados utilizadas para validação nos estudos consultados.

Estudo	Amostra de dados	Classificador	Estratégia de classificação	RR
6 emoções básicas				
BU-4DFE				
4D				
Zhen <i>et al.</i> (2017)	60S+6E+KF	HMM	10-CV	95,13 %
Maalej <i>et al.</i> (2013)	100S+6E+KF	HMM	10-CV	94,45 %
Amor <i>et al.</i> (2014)	60S+6E+WIN	HMM	10-CV	93,83 %
Li <i>et al.</i> (2018)	60S+6E+Full seq.	DGIN	10-CV	92,22 %
BU-3DFE				
2D+3D				
Li <i>et al.</i> (2015)	60S+6E+2L	RBF-SVM	100 × 90%/10%	94,56 %
Jan e Hongying Meng (2015)	60S+6E+2L	SVM	100 × 10-CV	90,04 %
Li <i>et al.</i> (2017a)	100S+6E+2L	DF-CNN (svm)	100 × 10-CV	86,86 %
Li <i>et al.</i> (2015)	60S+6E+2L	RBF-SVM	100 × 90%/10%	86,32 %
3D				
Hariri <i>et al.</i> (2017)	100S+6E+2L	SVM	10 × 90%/10%	92,62 %
Sheng <i>et al.</i> (2016)	60S+6E+2L	SVM	20 × 94%/6%	92,1 %
Gaoyun An <i>et al.</i> (2013)	60S+6E+NR	SVM	1000 × 90%/10%	91,3 %
Yurtkan e Demirel (2014a)	100S+6E+1L	SVM	2 × 90%/10%	90,8 %
Bosphorus				
2D+3D				
Li <i>et al.</i> (2015)	60S+6E	RBF-SVM	100 × 90%/10%	84,33 %
Li <i>et al.</i> (2017a)	60S+6E	DF-CNN (svm)	100 × 10-CV	80,28 %
Moeini <i>et al.</i> (2016)	6E*	SVM	10-CV	97,3 %
3D				
Zhang <i>et al.</i> (2015)	56S+6E	CMTNN	5-CV	92,2 %
Trimech <i>et al.</i> (2016)	65S+6E	SVM	80%/20%	73,07 %
6 emoções básicas e 1 neutra				
BU-3DFE				
2D+3D				
Jan e Hongying Meng (2015)	60S+7E+2L	SVM	100 × 10-CV	88,32 %
3D				
Khashman e Conkbayir (2013)	60S+7E+1L	MSVM	80%/20%	95,2 %
Savran e Sankur (2017)	100S+7E+2L	Adaboost with NBC as "weak learner"	10-CV	83,2 %
Han e Ming (2014)	100S+7E *	k-NN	50%/50%	77,3 %
Vieriu <i>et al.</i> (2015)	100S+7E+2L	Random Forest	5-CV	73,71 %
Bosphorus				
2D+3D				
Moeini <i>et al.</i> (2016)	7E **	SVM	10-CV	96 %
Azazi <i>et al.</i> (2015)	60S+7E	SVM with EPE	10-CV	84 %
Jan e Hongying Meng (2015)	7E **	SVM	100 × 10-CV	79,46 %
3D				
Azazi <i>et al.</i> (2014b)	65S+7E	RBF-SVM	10-CV	79 %
Hariri <i>et al.</i> (2017)	65S+7E	SVM	10 × 90%/10%	86,17 %

Nota: (*) indica que o nível de intensidade das expressões dos *scans* utilizados no experimentos não foi reportado.

(**) indica que o número de indivíduos considerados nos experimentos não foi reportado.

Fonte: elaborado pela autora.

Tabela 10 – Quadro resumo dos resultados.

Representação da face	
	As decisões de representação da face são, em geral, o cerne dos métodos propostos em REF 3D.
	A determinação de atributos apropriados para a representação da face para um sistema automático de classificação está relacionada a escolha de descritores robustos, bem como à determinação de regiões de interesse relevantes e aos tipos de dados a disposição da técnica.
	A maior parte dos estudos em REF 3D utiliza regiões de escopo local das quais atributos são extraídos utilizando-se dados puramente espaciais para representação da face.
Pré-processamento	
	37% dos estudos não mencionam quaisquer informações sobre pré-processamento, não deixando claro se os empregam ou se os mesmos são, de fato, dispensados.
	Dentre os estudos que relatam o emprego de pré-processamento, 59% reportam correção de pose, 45% reportam segmentação da face, 38%, suavização e 31%, preenchimento de buracos. É comum o emprego de múltiplos procedimentos de pré-processamento em um mesmo estudo.
	Comumente, são utilizados apenas termos genéricos para descrever a técnica de pré-processamento empregada, como por exemplo "alinhamento" ou "registro" para correção de pose. Essa prática foi observada em todas as demais técnicas de pré-processamento reportadas. Mesmo em estudos em que são mencionadas as técnicas específicas, faltam informações fundamentais para garantir sua replicabilidade.
	A detecção de pontos relevantes é requisito para aproximadamente 80% dos estudos em REF 3D.
	Mais de 50% de todos os estudos consultados propõem métodos que requerem a identificação de uma grande quantidade de pontos relevantes. Para isso, cerca de 42% realizam detecção automática.
Experimentos de classificação	
	Dois principais aspectos caracterizam os experimentos de classificação: a amostra de dados empregada e a estratégia de classificação.
	A formação da amostra de dados depende da composição das bases de dados empregadas, cujo número é limitado em REF 3D. Os <i>scans</i> das bases BU-3DFE, Bosphorus e BU-4DFE são os mais frequentemente usados.
	Sobre os dados da BU-3DFE, o perfil de amostra mais empregado é o 60S+6E+2HL.
	Sobre os dados da Bosphorus, os perfis de amostra mais frequentes são o 65S+6E e 65S+6E+1N.
	Sobre os dados da BU-4DFE, o perfil de amostra mais comum é o 60S+6E+KF.
	O SVM é o algoritmo de classificação mais utilizado em REF 3D, identificado em 57% dos estudos consultados.
	A única execução da estratégia de validação cruzada 10- <i>fold</i> é a configuração para estimativa da efetividade do modelo preditivo.
	25% dos estudos consultados não reportam adequadamente sua estratégia de classificação.

Fonte: elaborado pela autora.

5 DISCUSSÕES

Neste capítulo são apresentadas as discussões sobre a caracterização dos métodos 3D empregados no reconhecimento de expressões faciais sob os pontos de vista técnico e metodológico. Essas discussões são organizadas de acordo com os aspectos: representação da face, pré-processamento e experimentos de classificação.

5.1 Representação da face

A literatura sobre REF 3D aponta consistentemente para a eficácia de métodos que combinam diferentes tipos de dados. Por exemplo, Li *et al.* (2015) e Jan e Hongying Meng (2015) demonstram que, em experimentos equivalentes, métodos que combinam descritores 3D e de cor e textura apresentam taxas de reconhecimento superiores às apresentadas por métodos que empregam apenas atributos espaciais ou apenas atributos de cor. De fato, a apresentação de variedade de dados, de maneira geral, enriquece o modelo e a capacidade de generalização dos classificadores. Similarmente, Sandbach *et al.* (2011) demonstram que a informação extraída a partir de amostras sequenciais, que permitem a captura da variação de expressões ao longo do tempo, também melhora os resultados em REF 3D.

Ainda que essas evidências e outras similares tenham tido seu papel de estimular investigações nessa direção, constata-se por meio dos estudos consultados nesta revisão que, em grande parte, a literatura especializada propõe soluções baseadas em representações puramente 3D. Alguns fatores contribuem para justificar essas estatísticas: a significativa simplicidade de métodos puramente 3D, requerendo a extração de atributos a partir de apenas um tipo de dados, e a manipulação de um volume menor de dados em comparação com aquele requerido por métodos de dados combinados. Essas características contribuem para o baixo custo computacional dos métodos propostos, aumentando sua aplicabilidade em ambientes com recursos computacionais limitados.

Em métodos tradicionais, a preferência por ROI de escopo local, que consideram *keypoints*, curvas e superfícies, sugere a busca pelo equilíbrio entre simplicidade e robustez nas técnicas reportadas recentemente. Esses métodos descrevem a face em termos de regiões notadamente mais relevantes, em vez do uso de toda a superfície da face, e podem intencionalmente eliminar a influência de regiões redundantes entre as classes de emoções. De fato, como formalizado pelo FACS, a adoção dessas técnicas de representação da face é baseada na

percepção de que há regiões cujas deformações causadas por expressões emocionais são mais evidentes que outras. Não por acaso, os pontos considerados relevantes nos estudos em REF 3D geralmente coincidem com pontos fiduciais faciais, que marcam regiões características sobre a face, como cantos dos olhos, cantos da boca, sobrancelhas e ponta do nariz. Similarmente, trabalhos que consideram regiões locais, em geral, cobrem essas mesmas regiões.

5.2 Pré-processamento

O pré-processamento é inegavelmente um aspecto relevante na solução de problemas envolvendo aprendizagem de máquinas. Apesar da relativamente fácil identificação da aplicação desses procedimentos na literatura especializada em REF 3D, sua descrição frequentemente falha ao fornecer detalhes relevantes. Isso afeta a capacidade de outros pesquisadores replicarem os métodos, uma vez que, dessa forma, não há garantias de que os procedimentos serão conduzidos nas mesmas condições do original e, portanto, compromete a comparação justa entre técnicas, uma importante premissa em produções científicas.

Por exemplo, diversos métodos que dependem da detecção da ponta do nariz para a representação da face, o fazem pela identificação do ponto de maior magnitude da coordenada z . Esse procedimento, apesar de simples, é efetivo apenas em cenários "bem comportados", uma vez que assume que os *scans* têm pose corrigida e não possuem pontos ruidosos ou atributos faciais, como queixo protuberante, que potencialmente atrapalhem essa detecção. Como essa não é a situação da aquisição da maioria dos *scans* 3D disponíveis nas bases, procedimentos de correção de pose e remoção de ruído se tornam necessários para uma adequada representação da face. Um potencial problema surge quando, por exemplo, o algoritmo mais popular para correção de pose, o ICP, é sensível aos parâmetros e, naturalmente, produz resultados diferentes de registro para valores diferentes desses parâmetros, mesmo quando produzidos sobre a mesma base de dados. Dessa forma, a falha ao reportar adequadamente o procedimento de correção de pose tem impacto na identificação precisa de um ponto relevante, que, por sua vez, determina as ROI usadas na representação da face. Consequências similares são aplicáveis a outros procedimentos como suavização ou preenchimento de buracos.

Outro aspecto relevante a ser notado é que, por estar relacionada à qualidade dos dados a serem submetidos ao método proposto, a não aplicação de pré-processamento pode indicar a robustez da técnica a condições adversas, sendo portanto, informação valorizada. Apesar disso, os resultados obtidos por meio da RSL mostraram que é prática em estudos

especializados em REF 3D a omissão de informação sobre pré-processamento. Tal omissão não é explicitamente justificada no texto, deixando dúvidas sobre a possível dispensabilidade desses procedimentos. Essa prática foi observada em 43% dos artigos de conferência e 33% dos artigos de periódicos, o que reforça a evidência da omissão generalizada dessas informações.

Esse é um problema que se apresenta como técnico e metodológico, uma vez que surge da necessidade de pré-processamento para a adequada representação da face e se aprofunda com o negligenciamento de informação relevante sobre esse procedimento. Como oportunidade nessa direção está o desenvolvimento de técnicas robustas a condições adversas, como ruído ou rotação. Avanços têm sido reportados recentemente com métodos baseados em aprendizagem profunda e evidenciam potencial para melhorias (CHEN *et al.*, 2018).

Ainda, a detecção de pontos relevantes tem sido identificada como um requisito tradicional para apoiar extração de atributos e outros procedimentos de pré-processamento. Nesse contexto, no qual a dependência de detecção *a priori* de pontos implica ou num custo computacional tão alto quanto o número de pontos requeridos ou em severas restrições na aplicabilidade do método, justifica-se o interesse na proposição de métodos completamente independentes da detecção de pontos relevantes. Muito embora apenas 20% dos estudos consultados sejam completamente independentes de *keypoints*, demonstra-se uma tendência ao abandono de técnicas com esse tipo de dependência nos últimos anos, conforme mostrado na Figura 16.

5.3 Experimentos de classificação

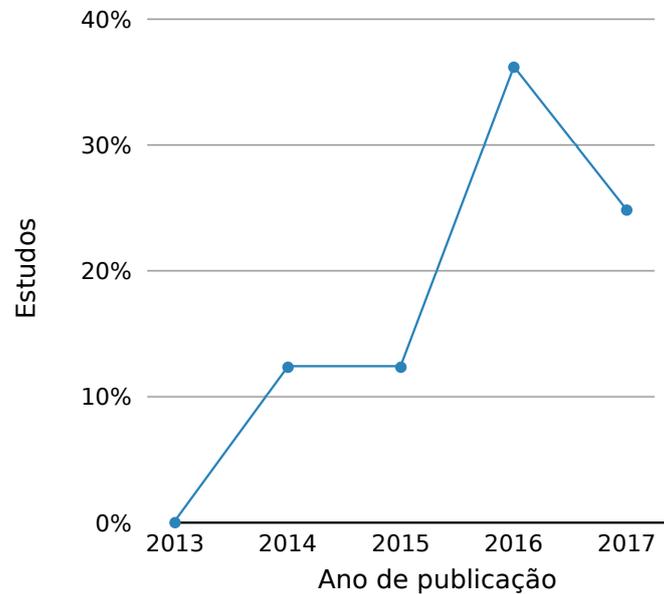
O detalhamento adequado do experimento de classificação é elemento crítico para a comparação entre métodos. Isso porque, além de métricas de qualidade usadas para avaliação quantitativa da efetividade do modelo preditivo (em geral, RR), é imprescindível que se mantenha a capacidade de realizar a comparação entre técnicas sob condições equivalentes, o que naturalmente inclui amostra de dados e estratégia de classificação.

A base BU-3DFE se destaca como a principal fonte de dados para os experimentos neste domínio. Sua alta utilização justifica-se, em parte, por ter sido a primeira base de dados 3D criada para o estudo de expressões faciais, como revela a cronologia apresentada em Corneanu *et al.* (2016). Mesmo com o advento de outras bases de dados, a BU-3DFE se consolidou como referência nos experimentos na área. Apesar dessa predominância, não há protocolos bem definidos sobre a composição da amostra de dados utilizada nos experimentos. Ao contrário, registra-se uma significativa variabilidade na escolha do número de indivíduos, expressões e

Figura 16 – Proporção de estudos independentes de *key-points* por ano de publicação.

Proporção de estudos independentes de key-points nos últimos anos

Amostra de 49 estudos especializados em REF 3D



Fonte: elaborado pela autora.

graus de intensidade considerados, sendo o mais frequente perfil, 60S+6E+2HL, presente em apenas 29% dos estudos que usam a base. Além disso, a exemplo do que acontece com o relato dos procedimentos de pré-processamento, a descrição apropriada do experimento de classificação é frequentemente negligenciada, o que compromete a credibilidade dos resultados e dificulta a comparação entre métodos distintos.

A estratégia de classificação é outro aspecto importante para a replicabilidade e comparação justa, uma vez que está diretamente relacionada à estabilidade das estimativas de desempenho de classificação. Dessa forma, é fator preocupante que a principal estratégia de classificação, execução única da validação cruzada *10-fold*, seja também aquela que resulta em estimativas mais instáveis.

Em classificação de padrões, a estabilidade de classificadores é um conceito-chave associado a generalidade de um modelo e depende da quantidade de dados (n) disponíveis para o aprendizado (DUDA *et al.*, 2012). Isso significa que, quanto maior a amostra de dados ($n \rightarrow \infty$), melhor. Na prática, no entanto, não há dados infinitos à disposição. De fato, como já exposto, importantes limitações em n são frequentes, particularmente em REF 3D. Assim, o resultado da apresentação de dados infinitos ao classificador é estimado por meio de técnicas de re-

amostragem. Essa estimativa é utilizada como principal medida para avaliar o modelo preditivo e, portanto é o principal critério de comparação entre técnicas em cenários similares. Como todas as estimativas, no entanto, essas medidas possuem incerteza intrínseca e sua confiança aumenta com o número de repetições que lhes dá origem. O Apêndice C apresenta uma exposição mais aprofundada da relação entre estabilidade das estimativas e confiança nos resultados.

Apesar de esse se revelar um problema mais metodológico do que técnico, essa prática representa um desafio que não pode ser negligenciado, pois impacta negativamente na proposição de novas técnicas, levando a comunidade a, inadvertidamente, perseguir resultados que são, na realidade, inatingíveis sob determinadas circunstâncias ou incomparáveis sob outras.

6 CONCLUSÕES E TRABALHOS FUTUROS

A análise crítica sobre as técnicas 3D empregadas no reconhecimento de expressões faciais revelou que a caracterização técnica dos métodos recentemente propostos nesse domínio passa, necessariamente, pela avaliação metodológica. Constatou-se que, enquanto as decisões para representação da face são consideradas o cerne dos relatos científicos nessa área, apresentando, portanto, riqueza de detalhes na maioria dos casos e pouca dificuldade em identificar o detalhamento do método, por outro lado, os demais aspectos que compõem uma solução para REF 3D não recebem a mesma atenção. O relato do pré-processamento é frequentemente insuficiente quando não inexistente. Na mesma direção, aspectos fundamentais para caracterizar completamente um experimento de classificação são sistematicamente omitidos. Ainda mais grave, a maioria dos experimentos de classificação nesse domínio empregam configurações que geram estimativas de desempenho instáveis, claramente desconsiderando o impacto do projeto do experimento na confiabilidade dos resultados apresentados.

Apresenta-se, portanto, como oportunidade de trabalhos futuros, a proposição de métodos 3D para reconhecimento de expressões faciais cujo relato esteja em conformidade com os requisitos técnicos e metodológicos para ser de efetiva utilidade para a comunidade científica. Dentre as oportunidades técnicas estão a proposição de métodos robustos a condições adversas de aquisição e independentes de detecção *a priori* de pontos relevantes ou com dependência mínima; e a proposição de técnicas de representação da face que capturem adequadamente características de expressões desafiadoras, como medo e tristeza, elevando suas taxas de reconhecimento. Como oportunidade de contribuição para a qualidade metodológica dos estudos na área está a adequada configuração dos experimentos de classificação, levando em consideração a estabilidade dos resultados reportados, assim como a descrição rica da configuração do experimento e dos eventuais procedimentos de pré-processamento.

Finalmente, entende-se que não é viável estabelecer cenários de experimentação absolutamente rígidos. De fato, uma exigência como essa seria contraditória à atividade de pesquisa científica que é por natureza inventiva e dinâmica. Apesar disso, há de se chamar atenção para as fortes evidências de que esta comunidade se beneficiaria de abordagens mais cuidadosas na condução e elaboração de relatos de seus experimentos.

REFERÊNCIAS

- ALEXANDRE, G. R.; THÉ, G. A. P.; SOARES, J. M. Reliability of Results and Fairness in the Comparison of Rates Among 3D Facial Expression Recognition Works. In: M., P. G. V. (Ed.). **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. Cham: Springer, 2019. v. 11678 LNCS, cap. 31, p. 391–401. ISBN 9783030298876. Disponível em: <http://link.springer.com/10.1007/978-3-030-29888-3_31>.
- ALY, S.; ABBOTT, A. L.; TORKI, M. A multi-modal feature fusion framework for kinect-based facial expression recognition using Dual Kernel Discriminant Analysis (DKDA). In: **2016 IEEE Winter Conference on Applications of Computer Vision (WACV)**. [S.l.]: IEEE, 2016. p. 1–10. ISBN 978-1-5090-0641-0.
- ALY, S.; TRUBANOVA, A.; ABBOTT, L.; WHITE, S.; YOUSSEF, A. VT-KFER: A Kinect-based RGBD+time dataset for spontaneous and non-spontaneous facial expression recognition. In: **2015 International Conference on Biometrics (ICB)**. [S.l.: s.n.], 2015. p. 90–97. ISSN 2376-4201.
- ALY, S.; YOUSSEF, A.; ABBOTT, L. Adaptive feature selection and data pruning for 3D facial expression recognition using the Kinect. In: **2014 IEEE International Conference on Image Processing (ICIP)**. Paris: IEEE, 2014. p. 1361–1365. ISBN 978-1-4799-5751-4. ISSN 1522-4880.
- AMOR, B. B.; DRIRA, H.; BERRETTI, S.; DAOUDI, M.; SRIVASTAVA, A. 4-D Facial Expression Recognition by Learning Geometric Deformations. **IEEE Transactions on Cybernetics**, v. 44, n. 12, p. 2443–2457, 12 2014. ISSN 2168-2267.
- AN, S.; RUAN, Q. 3D facial expression recognition algorithm using local threshold binary pattern and histogram of oriented gradient. In: **2016 IEEE 13th International Conference on Signal Processing (ICSP)**. [S.l.]: IEEE, 2016. p. 265–270. ISBN 978-1-5090-1344-9. ISSN 2164-5221.
- AZAZI, A.; LUTFI, S. L.; VENKAT, I. Analysis and evaluation of SURF descriptors for automatic 3D facial expression recognition using different classifiers. In: **2014 4th World Congress on Information and Communication Technologies, WICT 2014**. [S.l.]: IEEE, 2014. p. 23–28. ISBN 9781479981151.
- AZAZI, A.; LUTFI, S. L.; VENKAT, I. Identifying universal facial emotion markers for automatic 3D facial expression recognition. In: **2014 International Conference on Computer and Information Sciences (ICCOINS)**. [S.l.]: IEEE, 2014. p. 1–6. ISBN 978-1-4799-4390-6.
- AZAZI, A.; LUTFI, S. L.; VENKAT, I.; FERNÁNDEZ-MARTÍNEZ, F. Towards a robust affect recognition: Automatic facial expression recognition in 3D faces. **Expert Systems with Applications**, v. 42, n. 6, p. 3056–3066, 4 2015. ISSN 09574174.
- BERRETTI, S.; BIMBO, A. D.; PALA, P. Automatic Facial Expression Recognition in Real-time from Dynamic Sequences of 3D Face Scans. **Vis. Comput.**, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 29, n. 12, p. 1333–1350, 2013. ISSN 0178-2789.
- BESL, P. J.; MCKAY, N. D. A method for registration of 3-D shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 2, p. 239–256, 1992. ISSN 1939-3539.

BIOLCHINI, J. C. de A.; MIAN, P. G.; NATALI, A. C. C.; CONTE, T. U.; TRAVASSOS, G. H. Scientific research ontology to support systematic review in software engineering. **Advanced Engineering Informatics**, v. 21, n. 2, p. 133–151, 2007. ISSN 14740346.

BOUSQUET, O.; ELISSEEFF, A. Stability and Generalization. **Journal of Machine Learning Research**, v. 2, n. 3, p. 499–526, 2002. ISSN 15324435.

CHEN, Z.; HUANG, D.; WANG, Y.; CHEN, L. Fast and Light Manifold CNN based 3D Facial Expression Recognition across Pose Variations. In: **2018 ACM Multimedia Conference on Multimedia Conference - MM '18**. New York, New York, USA: ACM Press, 2018. p. 229–238. ISBN 9781450356657.

CORNEANU, C. A.; SIMON, M. O.; COHN, J. F.; GUERRERO, S. E. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 38, n. 8, p. 1548–1568, 8 2016. ISSN 0162-8828.

DANELAKIS, A.; THEOHARIS, T.; PRATIKAKIS, I. A Survey on Facial Expression Recognition in 3D Video Sequences. **Multimedia Tools Appl.**, Kluwer Academic Publishers, Hingham, MA, USA, v. 74, n. 15, p. 5577–5615, 2015. ISSN 1380-7501.

DARWIN, C. The expression of the emotions in man and animals. **New York: Philosophical Library**, 1872.

DERKACH, D.; SUKNO, F. M. Local Shape Spectrum Analysis for 3D Facial Expression Recognition. In: **2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)**. [S.l.]: IEEE, 2017. p. 41–47. ISBN 978-1-5090-4023-0.

DUDA, R. O.; HART, P. E.; STORK, D. G. Algorithm-independent machine learning. In: **Pattern Classification**. New York, NY: John Wiley & Sons, 2012. cap. 9, p. 453–516. ISBN 0-471-05669-3.

EKMAN, P. Universals and cultural differences in facial expressions of emotion. **Nebraska Symposium on Motivation**, University of Nebraska Press, US, v. 19, p. 207–283, 1971. ISSN 0070-2099(Print).

EKMAN, P.; FRIESEN, W. **Facial Action Coding System: A Technique for the Measurement of Facial Movement**. Palo Alto: Consulting Psychologists Press, 1978.

EKMAN, P.; FRIESEN, W. **Unmasking the Face: A Guide to Recognizing Emotions From Facial Clues**. [S.l.: s.n.], 2003. ISBN 9781883536367.

ELHOULFI, S.; JAZOULI, M.; MAJDA, A.; ZARGHILI, A.; AALOUANE, R. Automatic recognition of facial expressions using microsoft kinect with artificial neural network. In: **2016 International Conference on Engineering & MIS (ICEMIS)**. [S.l.]: IEEE, 2016. p. 1–5. ISBN 978-1-5090-5579-1.

EMAMBAKSHH, M.; EVANS, A. Nasal Patches and Curves for Expression-Robust 3D Face Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 5, p. 995–1007, 5 2017. ISSN 01628828.

FELIZARDO, K. R.; NAKAGAWA, E. Y.; FABBRI, S. C. P. F.; FERRARI, F. C. **Revisão Sistemática da Literatura em Engenharia de Software: teoria e prática**. 1. ed. Rio de Janeiro: Elsevier Editora Ltda., 2017. 144 p. ISBN 978-85-352-8641-0.

FU, Y.; RUAN, Q.; AN, G.; JIN, Y. Fast nonnegative tensor factorization based on graph-preserving for 3D facial expression recognition. In: **2016 IEEE 13th International Conference on Signal Processing (ICSP)**. [S.l.]: IEEE, 2016. (International Conference on Signal Processing), p. 292–297. ISBN 978-1-5090-1344-9.

Gaoyun An; Qiuqi Ruan; Xiaoli Li. 3D facial expression recognition using delta faces. In: **5th IET International Conference on Wireless, Mobile and Multimedia Networks (ICWMMN 2013)**. [S.l.]: Institution of Engineering and Technology, 2013. v. 2013, n. 641 CP, p. 10–4. ISBN 978-1-84919-726-7. ISSN 0954-8963.

GOLEMAN, D. **Inteligência emocional: a teoria revolucionária que redefine o que é ser inteligente**. 2. ed. Rio de Janeiro: Objetiva, 2012. 384 p. ISBN 9788573020809.

GONG, B.; WANG, Y.; LIU, J.; TANG, X. Automatic Facial Expression Recognition on a Single 3D Face by Exploring Shape Deformation. In: **ACM International Conference on Multimedia**. [S.l.: s.n.], 2009. p. 569–572.

HAN, D.; MING, Y. Facial expression recognition with LBP and SLPP combined method. In: **2014 12th International Conference on Signal Processing (ICSP)**. [S.l.]: IEEE, 2014. v. 2015-Janua, n. October, p. 1418–1422. ISBN 978-1-4799-2186-7.

HARIRI, W.; TABIA, H.; FARAH, N.; BENOURETH, A.; DECLERCQ, D. 3D facial expression recognition using kernel methods on Riemannian manifold. **Engineering Applications of Artificial Intelligence**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 64, n. C, p. 25–32, 9 2017. ISSN 09521976.

HASANI, B.; MAHOOR, M. H. Facial Expression Recognition Using Enhanced Deep 3D Convolutional Neural Networks. In: **2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)**. [S.l.]: IEEE, 2017. v. 2017-July, p. 2278–2288. ISBN 978-1-5386-0733-6.

HAYAT, M.; BENNAMOUN, M.; EL-SALLAM, A. A. Clustering of Video-patches on Grassmannian Manifold for Facial Expression Recognition from 3D Videos. In: **Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV)**. Washington, DC, USA: IEEE Computer Society, 2013. (WACV '13), p. 83–88. ISBN 978-1-4673-5053-2.

JAN, A.; Hongying Meng. Automatic 3D facial expression recognition using geometric and textured feature fusion. In: **2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. [S.l.]: IEEE, 2015. v. 05, p. 1–6. ISBN 978-1-4799-6026-2.

JAZOULI, M.; MAJDA, A.; ZARGHILI, A. A \$P recognizer for automatic facial emotion recognition using Kinect sensor. In: **2017 Intelligent Systems and Computer Vision (ISCV)**. [S.l.]: IEEE, 2017. p. 1–5. ISBN 978-1-5090-4062-9.

KAHNEMAN, D. **Rápido e devagar: duas formas de pensar**. 1. ed. Rio de Janeiro: Objetiva, 2012. 608 p. ISBN 9788539003839.

KHASHMAN, A.; CONKBAYIR, F. O. Intelligent recognition of emotional expressions in 3D face images. In: **2013 21st Signal Processing and Communications Applications Conference (SIU)**. [S.l.]: IEEE, 2013. p. 1–4. ISBN 978-1-4673-5563-6.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. TR/SE-0401, p. 28, 3 2004. ISSN 13537776.

KONAR, A.; CHAKRABORTY, A. **Emotion recognition: A pattern analysis approach**. [S.l.]: John Wiley & Sons, 2015.

LAKATOS, E. M.; MARCONI, M. d. A. **Eva_Maria_Lakatos_&_Marina_de_Andrade_Marconi-_Metodologia_do_trabalho_científico.pdf**. 2008. 138 p.

LEKDIOUI, K.; RUICHEK, Y.; MESSOUSSI, R.; CHAABI, Y.; TOUAHNI, R. Facial expression recognition using face-regions. **Proceedings - 3rd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2017**, p. 1–6, 2017.

LEMAIRE, P.; ARDABILIAN, M.; CHEN, L.; DAOUDI, M. Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients. In: **2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. IEEE, 2013. p. 1–7. ISBN 978-1-4673-5546-9. Disponível em: <<http://ieeexplore.ieee.org/document/6553821/>>.

LI, H.; DING, H.; HUANG, D.; WANG, Y.; ZHAO, X.; MORVAN, J.-M.; CHEN, L. An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition. **Computer Vision and Image Understanding**, Elsevier Science Inc., New York, NY, USA, v. 140, n. C, p. 83–92, 11 2015. ISSN 10773142.

LI, H.; SUN, J.; XU, Z.; CHEN, L. Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network. **IEEE Transactions on Multimedia**, v. 19, n. 12, p. 2816–2831, 2017. ISSN 1520-9210.

LI, Q.; AN, G.; RUAN, Q. 3D Facial expression recognition using orthogonal tensor marginal fisher analysis on geometric maps. In: **2017 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)**. [S.l.]: IEEE, 2017. v. 35, p. 65–71. ISBN 978-1-5386-0410-6.

LI, S.; DENG, W. Deep Facial Expression Recognition: A Survey. p. 1–25, 2018. Disponível em: <<http://arxiv.org/abs/1804.08348>>.

LI, W.; HUANG, D.; LI, H.; WANG, Y. Automatic 4D Facial Expression Recognition Using Dynamic Geometrical Image Network. In: **2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)**. IEEE, 2018. p. 24–30. ISBN 978-1-5386-2335-0. Disponível em: <<https://ieeexplore.ieee.org/document/8373807/>>.

LI, X.; RUAN, Q.; AN, G.; JIN, Y.; ZHAO, R. Multiple strategies to enhance automatic 3D facial expression recognition. **Neurocomputing**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 161, n. C, p. 89–98, 8 2015. ISSN 09252312.

LUNESKI, A.; KONSTANTINIDIS, E.; BAMIDIS, P. D. Affective Medicine. **Methods of Information in Medicine**, v. 49, n. 03, p. 207–218, 1 2010. ISSN 0026-1270.

MAALEJ, A.; TABIA, H.; BENHABILES, H. Dynamic 3D Facial Expression Recognition Using Robust Shape Features. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2013. v. 7944 LNCS, p. 309–318.

- MARCONI, M.; LAKATOS, E. **Fundamentos de metodologia científica**. [S.l.: s.n.], 2003. 310 p. ISSN 9788522457588. ISBN 8522433976.
- MOEINI, A.; FAEZ, K.; SADEGHI, H.; MOEINI, H. 2D Facial Expression Recognition via 3D Reconstruction and Feature Fusion. **JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION**, Academic Press, Inc., Orlando, FL, USA, v. 35, n. C, p. 1–14, 2016. ISSN 1047-3203.
- OCEGUEDA, O.; Tianhong Fang; SHAH, S. K.; KAKADIARIS, I. A. 3D Face Discriminant Analysis Using Gauss-Markov Posterior Marginals. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 3, p. 728–739, 3 2013. ISSN 0162-8828.
- REALE, M.; ZHANG, X.; YIN, L. Nebula feature: A space-time feature for posed and spontaneous 4D facial behavior analysis. In: **2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. [S.l.]: IEEE, 2013. p. 1–8. ISBN 978-1-4673-5546-9.
- RODOLÀ, E.; ALBARELLI, A.; CREMERS, D.; TORSELLO, A. A simple and effective relevance-based point sampling for 3D shapes. **Pattern Recognition Letters**, v. 59, n. 2015, p. 41–47, 7 2015. ISSN 01678655.
- SALVADOR, A. D. **Metodos e tecnicas de pesquisa bibliografica: elaboração de trabalhos científicos**. [S.l.]: Sulina, 1986.
- SAMAD, M. D.; IFTEKHARUDDIN, K. M. Frenet Frame-Based Generalized Space Curve Representation for Pose-Invariant Classification and Recognition of 3-D Face. **IEEE Transactions on Human-Machine Systems**, v. 46, n. 4, p. 522–533, 8 2016. ISSN 2168-2291.
- SANDBACH, G.; ZAFEIRIOU, S.; PANTIC, M.; RUECKERT, D. A dynamic approach to the recognition of 3D facial expressions and their temporal models. In: **Face and Gesture 2011**. IEEE, 2011. p. 406–413. ISBN 978-1-4244-9140-7. ISSN 01628828. Disponível em: <<http://ieeexplore.ieee.org/document/5771434/>>.
- SANDBACH, G.; ZAFEIRIOU, S.; PANTIC, M.; YIN, L. Static and dynamic 3D facial expression recognition: A comprehensive survey. **Image and Vision Computing**, Elsevier B.V., v. 30, n. 10, p. 683–697, 2012. ISSN 02628856.
- SAVRAN, A.; ALYÜZ, N.; DIBEKLIOĞLU, H.; ÇELIKTUTAN, O.; GÖKBERK, B.; SANKUR, B.; AKARUN, L. Bosphorus database for 3D face analysis. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 5372 LNCS, p. 47–56, 2008. ISSN 03029743.
- SAVRAN, A.; SANKUR, B. Non-rigid Registration Based Model-free 3D Facial Expression Recognition. **Comput. Vis. Image Underst.**, Elsevier Science Inc., New York, NY, USA, v. 162, n. C, p. 146–165, 2017. ISSN 1077-3142.
- SCHULLER, B.; WIMMER, M.; ARSIC, D.; MOOSMAYR, T.; RIGOLL, G. Detection of security related affect and behaviour in passenger transport. In: **Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH**. Brisbane, AUSTRALIA: ISCA-INT SPEECH COMMUNICATION ASSOC, 2008. p. 265–268. ISBN 978-1-61567-378-0. ISSN 19909772.

SHENG, N.; CAI, Y.; ZHAN, C.; Changyan Qiu; Yize Cui; Xurong Gao. 3D facial expression recognition using distance features and LBP features based on automatically detected keypoints. In: **2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)**. Datong: IEEE, 2016. p. 396–401. ISBN 978-1-5090-3710-0.

SOLTANPOUR, S.; BOUFAMA, B.; WU, Q. M. J. A Survey of Local Feature Methods for 3D Face Recognition. **Pattern Recognition**, Elsevier Science Inc., New York, NY, USA, v. 72, n. C, p. 391–406, 2017. ISSN 0031-3203.

SUJA, P.; KRISHNASRI, D.; TRIPATHI, S. Pose invariant method for emotion recognition from 3D images. In: **12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015**. [S.l.]: IEEE, 2016. p. 1–5. ISBN 9781467373999.

TAO, L.; MATUSZEWSKI, B. J. Is 2D Unlabeled Data Adequate for Recognizing Facial Expressions? **IEEE Intelligent Systems**, v. 31, n. 3, p. 19–29, 5 2016. ISSN 15411672.

TRIMECH, I. H.; MAALEJ, A.; AMARA, N. E. B. 3D facial expression recognition using nonrigid CPD registration method. In: **2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)**. [S.l.]: IEEE, 2016. p. 478–481. ISBN 978-1-5090-4712-3.

VALSTAR, M. F.; JIANG, B.; MEHU, M.; PANTIC, M.; SCHERER, K. The first facial expression recognition and analysis challenge. **2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011**, IEEE, p. 921–926, 2011.

VIERIU, R.-L.; TULYAKOV, S.; SEMENIUTA, S.; SANGINETO, E.; SEBE, N. Facial expression recognition under a wide range of head poses. In: **2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. [S.l.]: IEEE, 2015. v. 1, p. 1–7. ISBN 978-1-4799-6026-2.

Xudong Yang; Di Huang; Yunhong Wang; Liming Chen. Automatic 3D facial expression recognition using geometric scattering representation. In: **2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. [S.l.]: IEEE, 2015. p. 1–6. ISBN 978-1-4799-6026-2.

XUE, M.; MIAN, A.; LIU, W.; Ling Li. Fully automatic 3D facial expression recognition using local depth features. In: **IEEE Winter Conference on Applications of Computer Vision**. [S.l.]: IEEE, 2014. p. 1096–1103. ISBN 978-1-4799-4985-4. ISSN 1550-5790.

XUE, M.; MIAN, A.; LIU, W.; LI, L. Automatic 4D Facial Expression Recognition Using DCT Features. In: **2015 IEEE Winter Conference on Applications of Computer Vision**. [S.l.]: IEEE, 2015. p. 199–206. ISBN 978-1-4799-6683-7. ISSN 1550-5790.

YAO, Y.; HUANG, D.; YANG, X.; WANG, Y.; CHEN, L. Texture and Geometry Scattering Representation-Based Facial Expression Recognition in 2D+3D Videos. **ACM Transactions on Multimedia Computing, Communications, and Applications**, ACM, New York, NY, USA, v. 14, n. 1s, p. 1–23, 3 2018. ISSN 15516857.

YIN, L.; CHEN, X.; SUN, Y.; WORM, T.; REALE, M. A High-Resolution 3D Dynamic Facial Expression Database. **Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on**, n. 1, p. 1–6, 2008.

YIN, L.; WEI, X.; SUN, Y.; WANG, J.; ROSATO, M. J. A 3D facial expression database for facial behavior research. **FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition**, v. 2006, p. 211–216, 2006. ISSN 0769525032.

YURTKAN, K.; DEMIREL, H. Person independent facial expression recognition using 3D facial feature positions. In: **Computer and Information Sciences III - 27th International Symposium on Computer and Information Sciences, ISCIS 2012**. [S.l.: s.n.], 2013. p. 321–329.

YURTKAN, K.; DEMIREL, H. Entropy-based feature selection for improved 3D facial expression recognition. **Signal, Image and Video Processing**, v. 8, n. 2, p. 267–277, 2 2014. ISSN 1863-1703.

YURTKAN, K.; DEMIREL, H. Feature selection for improved 3D facial expression recognition. **Pattern Recognition Letters**, Elsevier Science Inc., New York, NY, USA, v. 38, n. 1, p. 26–33, 3 2014. ISSN 01678655.

YURTKAN, K.; SOYEL, H.; DEMIREL, H. Feature Selection for Enhanced 3D Facial Expression Recognition Based on Varying Feature Point Distances. In: GELENBE, E.; LENT, R. (Ed.). **Information Sciences and Systems 2013**. [S.l.: s.n.], 2013, (Lecture Notes in Electrical Engineering, v. 264). p. 209–217. ISBN 978-3-319-01604-7; 978-3-319-01603-0.

ZARBAKSH, P.; DEMIREL, H. Fuzzy SVM for 3D facial expression classification using sequential forward feature selection. In: **2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)**. [S.l.]: IEEE, 2017. p. 131–134. ISBN 978-1-5090-5001-7.

ZENG, W.; LI, H.; CHEN, L.; MORVAN, J.-M.; GU, X. D. An automatic 3D expression recognition framework based on sparse representation of conformal images. In: **2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)**. [S.l.]: IEEE, 2013. p. 1–8. ISBN 978-1-4673-5546-9.

ZHANG, T. Facial Expression Recognition Based on Deep Learning: A Survey. In: **Advances in Intelligent Systems and Computing**. [S.l.: s.n.], 2018. v. 686, p. 345–352. ISBN 9783319690957.

ZHANG, Y.; ZHANG, L.; HOSSAIN, M. Adaptive 3D facial action intensity estimation and emotion recognition. **Expert Systems with Applications**, v. 42, n. 3, p. 1446–1464, 2 2015. ISSN 09574174.

ZHEN, Q.; HUANG, D.; DRIRA, H.; AMOR, B. B.; WANG, Y.; DAOUDI, M. Magnifying Subtle Facial Motions for Effective 4D Expression Recognition. **IEEE Transactions on Affective Computing**, PP, n. 99, p. 1–1, 2017. ISSN 1949-3045.

ZHEN, Q.; HUANG, D.; WANG, Y.; CHEN, L. Muscular Movement Model-Based Automatic 3D/4D Facial Expression Recognition. **IEEE Transactions on Multimedia**, IEEE Press, Piscataway, NJ, USA, v. 18, n. 7, p. 1438–1450, 7 2016. ISSN 1520-9210.

APÊNDICE A – STRINGS DE BUSCA GENÉRICA E ADAPTADAS

A.1 *String de busca genérica*

(3D OR “Three dimension” OR “Three dimensional”) AND (face OR facial) AND (recognition OR recognizer OR detection OR detector OR detecting OR identification OR identifier OR identifying) AND (expression OR affect OR affection OR emotion OR emotional OR sentiment OR feeling) AND (“keypoint extraction” OR “key point extraction” OR “keypoints extraction” OR “key points extraction” OR “feature selection” OR “features selection” OR “attribute selection” OR “attributes selection” OR descriptor OR signature OR classification OR classifying OR classifier OR application OR system)

A.2 *String de busca adaptada para base bibliográfica Scopus*

((KEY (("keypoint* extraction"OR "key point* extraction"OR "feature* selection"OR "attribute* selection"OR descriptor* OR signature OR classif* OR system OR application))) OR (ABS (("keypoint* extraction"OR "key point* extraction"OR "feature* selection"OR "attribute* selection"OR descriptor* OR signature OR classif* OR system OR application))) OR (TITLE (("keypoint* extraction"OR "key point* extraction"OR "feature* selection"OR "attribute* selection"OR descriptor* OR signature OR classif* OR system OR application)))) AND (ABS ((3*d OR "Three dimension*") AND (face OR facial) AND (expression OR affect* OR emotion* OR sentiment OR feeling*) AND (recognition OR recognizer OR detect* OR identif*))) AND (TITLE ((3*d OR "Three dimension*") AND (face OR facial) AND (expression OR affect* OR emotion* OR sentiment OR feeling*) AND (recognition OR recognizer OR detect* OR identif*))) AND (LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013)) AND (LIMIT-TO (LANGUAGE , "English"))

A.3 *String de busca adaptada para a base bibliográfica Web of Science*

Combinação das seguintes pesquisas na ferramenta de busca:

Search #1: TOPIC: (("keypoint extraction"OR “keypoints extraction” OR "key point extraction"OR “key points extraction” OR "feature selection"OR “features selection” OR "attribute selection"OR “attributes selection” OR descriptor OR signature OR classif* OR system OR

application)) Timespan=All years Search language=Auto

Search #2: TITLE: ((3D OR “Three dimension” OR “Three dimensions” OR “Three dimensional”) AND (face OR facial) AND (recogni* OR detect* OR identif*) AND (expression OR affect* OR emotion* OR sentiment OR feeling)) Timespan=All years Search language=Auto

Search #3: Search #2 AND Search #1 Timespan=All years Search language=Auto

Search #4: Search #3 Refined by: PUBLICATION YEARS: (2013 OR 2014 OR 2015 OR 2016 OR 2017 OR 2018) Timespan=All years Search language=Auto

Search #5: Search #4 AND LANGUAGES: (ENGLISH) Timespan=All years Search language=Auto

A.4 String de busca adaptada para a base bibliográfica *IEEE Xplore*

((((((((((("3D") OR "Three dimension") OR "Three dimensions") OR "Three dimensional")))) AND (((("face") OR "facial")))) AND (((((((((((("recognition") OR "recognizer") OR "detection") OR "detector") OR "detecting") OR "identification") OR "identifier") OR "identifying")))) AND (((((((((((("expression") OR "affect") OR "affective") OR "affection") OR "emotion") OR "emotional") OR "sentiment") OR "feeling")))) AND (((((((((((("keypoint extraction") OR "keypoints extraction") OR "key point extraction") OR "feature selection") OR "features selection") OR "attribute selection") OR "attributes selection") OR "descriptor") OR "signature") OR classif* OR system OR application)) Date filter: 2013 to 2018

A.5 String de busca adaptada para a base bibliográfica *ACM Digital Library*

((acmdlTitle:(3*D “Three dimension” “Three dimensions” “Three dimensional”)) AND (acmdlTitle:(face facial)) AND (acmdlTitle:(recognition recognizer detection detector detecting identification identifier identifying))) AND ((recordAbstract:(3*D “Three dimension” “Three dimensions” “Three dimensional”)) AND (recordAbstract:(face facial)) AND (recordAbstract:(recognition recognizer detection detector detecting identification identifier identifying)))) AND (((acmdlTitle: ("keypoint extraction""keypoints extraction" "key point extraction""key points extraction" "feature selection""features selection" "attribute selection""attributes selection" descriptor* signature classif* system application))) OR ((recordAbstract: ("keypoint extraction""keypoints extraction" "key point extraction""key points extraction" "feature selection""features selection" "attribute selection""attributes selection" descriptor* signature classif*

system application))) OR ((keywords.author.keyword: ("keypoint extraction""keypoints extraction" "key point extraction""key points extraction" "feature selection""features selection" "attribute selection""attributes selection" descriptor* signature classif* system application))))

Date filter: 2013 to 2018

APÊNDICE B – FORMULÁRIO DE EXTRAÇÃO DE DADOS PREENCHIDO

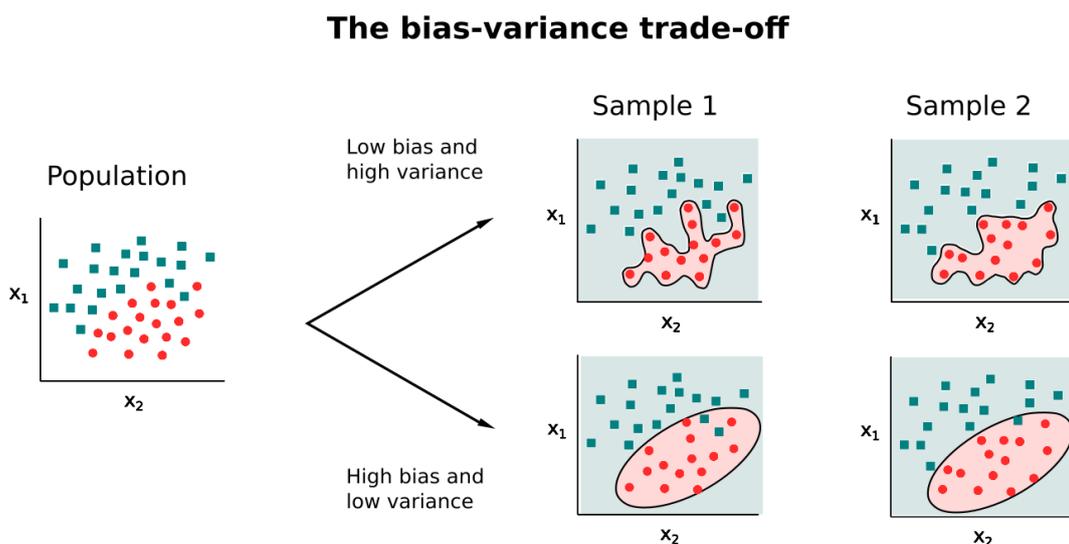
O formulário de extração de dados preenchido para os 49 estudos consultados nesta revisão encontra-se disponível pelo *link*: <https://bit.ly/33AJKub>.

APÊNDICE C – ESTABILIDADE DAS ESTIMATIVAS EM CLASSIFICAÇÃO DE PADRÕES

O conteúdo deste apêndice é um trecho extraído do artigo autoral Alexandre *et al.* (2019) e trata dos fundamentos matemáticos que sustentam o conceito de estabilidade das estimativas de classificação de padrões e como esses conceitos se relacionam com os achados da RSL.

A estabilidade de algoritmos de aprendizagem tem sido abordadas na literatura especializada por décadas (BOUSQUET; ELISSEEFF, 2002). Classificadores instáveis são sensíveis a mudanças nos dados de treinamento e, portanto, levam a modelos de baixa generalidade. As medidas de viés e variância quantificam o quão bom é um dado modelo para os dados de treinamento, levando em consideração sua acurácia e precisão, respectivamente. Um classificador pouco enviesado se adequa bem às particularidade dos dados de treinamento, enquanto um classificador de baixa variância é consistente por todas as variações dos dados de treinamento. Idealmente, os classificadores devem possuir baixo viés e baixa variância. Entretanto, essas medidas não são independentes e aumentar a flexibilidade para adaptar o classificador aos dados de treinamento geralmente leva a baixo viés, mas alta variância (DUDA *et al.*, 2012). A Figura 17 ilustra esse fenômeno. Essa é, na realidade, outra forma de se manifestar o já bem conhecido *trade-off* entre simplificar o modelo e se adequar aos dados.

Figura 17 – O *Trade-off* viés-variância.



Fonte: (ALEXANDRE *et al.*, 2019)

Adicionalmente, o tamanho do conjunto de treinamento desempenha um papel

importante no ajuste das variâncias, uma vez que, para um dado viés, o aumento no número de exemplos leva a um erro de generalização mais baixo e estável. Frequentemente, problemas de aprendizado de máquinas não dispõem de um conjunto de treinamento muito grande e, portanto, recorrem a métodos de re-amostragem para obter estimativas informativas para avaliar o desempenho do classificador (acurácia, por exemplo).

Especificamente, no domínio de REF 3D, há um número limitado de bases de *scans* faciais 3D, dos quais a BU-3DFE e a Bosphorus são as mais amplamente utilizadas. A primeira é composta por 2500 *scans* de 100 indivíduos, cada um reproduzindo um mesmo conjunto de condições: uma pose neutra e expressões de seis emoções básicas (raiva, aversão, medo, felicidade, tristeza e surpresa), distribuídas em quatro níveis de intensidade. Apesar disso, frequentemente apenas os *scans* dos níveis mais altos de intensidade de um subconjunto de indivíduos são efetivamente considerados nos experimentos.

A Bosphorus, por sua vez, possui 4666 *scans* de 105 indivíduos em variadas condições de pose, oclusão e expressão. Nessa base de dados, nem todos os indivíduos reproduzem todos os cenários. Assim, apenas um subconjunto da base de dados completa é composta por indivíduos realizando expressões faciais das seis emoções básicas e, portanto, empregadas em REF 3D.

Nesse cenário de dados limitados, múltiplas execuções dos experimentos de classificação com métodos de re-amostragem são realizados e a medida média de acurácia de todas as execuções (RR) é frequentemente utilizada como estimativa de desempenho do classificador. Há, entretanto, uma relação conhecida entre o número de repetições da estratégia de classificação e a estabilidade da medida de avaliação. Essa relação foi empiricamente demonstrada por Gong *et al.* (2009), em que experimentos de classificação realizados apenas algumas poucas vezes apresentaram resultados altamente instáveis. De fato, esse é um comportamento esperado e que encontra base teórica na conexão entre os intervalos de confiança para estimativa da taxa de reconhecimento e o número de repetições do experimento de classificação.

Partindo dessa perspectiva, quando consideramos k como o resultado de cada re-amostragem seguida do procedimento de classificação e N o total de repetições, a taxa de reconhecimento média pode ser tomada como a frequência relativa

$$\hat{p} = \frac{k}{N}, \tag{C.1}$$

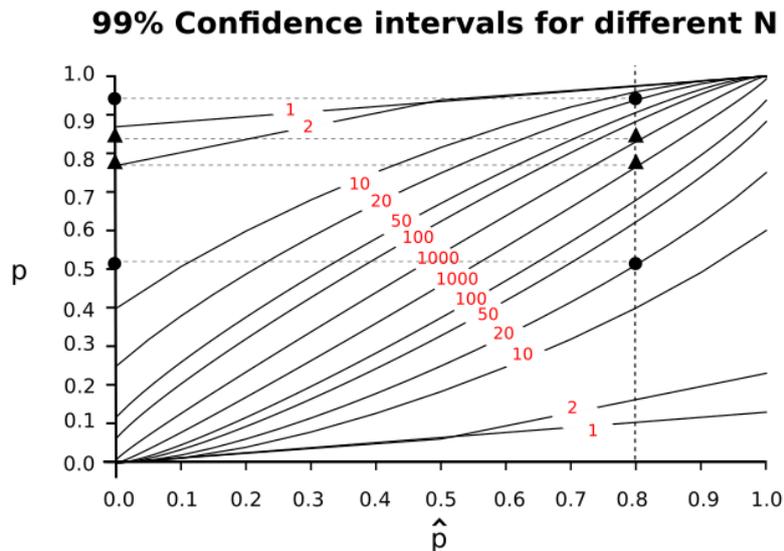
do evento "classificação correta", que tem distribuição binomial sobre um total de N repetições,

de acordo com a Equação C.2.

$$P(k) = \binom{N}{k} p^k (1-p)^{N-k}. \quad (\text{C.2})$$

O conceito de intervalo de confiança (IC) fornece os meios para descrever a confiabilidade de uma dada medida, uma vez que ele evidencia os limites de potenciais valores verdadeiros de uma estimativa em um dado nível de confiança. Ao fixar o nível de confiança em 99% e variar N , é possível obter uma série de possíveis valores verdadeiros p para uma dada estimativa \hat{p} . É possível ainda notar que a incerteza associada a um dado \hat{p} é menor quanto maiores os valores de N , conforme Figura 18.

Figura 18 – Os intervalos de confiança de 99% para diferentes valores de N .



Fonte: (ALEXANDRE *et al.*, 2019)

Dessa forma, evidencia-se que a configuração experimental tem influencia significativa no resultado final e, muitas vezes, de maneira pouco aparente, afetando por exemplo a certeza associada a ele. As RR derivadas da média de poucas re-amostragens carregam, potencialmente, uma medida pouco representativa da performance de um dado método, uma vez que ela é também instável, representando um desafio para as comparações justas entre técnicas. Para superar essa dificuldade, há de se adotar como prática especialmente nesse domínio a realização de experimentos que promovam a estabilidade dos resultados.