



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

IZAÍAS EMÍDIO MACHADO JÚNIOR

UM ESTUDO SOBRE RECONHECIMENTO FACIAL 3D TOLERANTE A OCLUSÃO
E BASEADO EM HISTOGRAMAS DE ATRIBUTOS

FORTALEZA

2020

IZAÍAS EMÍDIO MACHADO JÚNIOR

UM ESTUDO SOBRE RECONHECIMENTO FACIAL 3D TOLERANTE A OCLUSÃO E
BASEADO EM HISTOGRAMAS DE ATRIBUTOS

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: Engenharia de Teleinformática

Orientador: Prof. Dr. George André Pereira Thé

Coorientador: Prof. Dr. José Marques Soares

FORTALEZA

2020

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)

J1e Júnior, Izaias Emídio Machado.
Um estudo sobre reconhecimento facial 3D tolerante a oclusão e baseado em histogramas de atributos /
Izaias Emídio Machado Júnior. – 2020.
72 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-
Graduação em Engenharia de Teleinformática, Fortaleza, 2020.
Orientação: Prof. Dr. George André Pereira Thé.
Coorientação: Prof. Dr. José Marques Soares.

1. Imagens. 2. Face. 3. Oclusão. 4. 3D. 5. Descritores. I. Título.

CDD 621.38

IZAÍAS EMÍDIO MACHADO JÚNIOR

UM ESTUDO SOBRE RECONHECIMENTO FACIAL 3D TOLERANTE A OCLUSÃO E
BASEADO EM HISTOGRAMAS DE ATRIBUTOS

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: Engenharia de Teleinformática

Aprovada em:

BANCA EXAMINADORA

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

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

Prof. Dr. Jarbas Joaci de Mesquita Sá Junior
Universidade Federal do Ceará (UFC Sobral)

Prof. Dr. Paulo Cesar Cortez
Universidade Federal do Ceará (UFC)

Prof. Dr. João Paulo Silva do Monte Lima
Universidade Federal Rural de Pernambuco (UFRPE)

Dedico aos meu pais, Izaias e Ariene, às minhas irmãs Iza e Camila e ao meu cunhado Júnior, que são a minha "base vetorial" na vida.

AGRADECIMENTOS

A Deus, meu eterno pai que sempre me sustentou e me dá forças para continuar caminhando.

Aos meus professores Dr. George André Pereira Thé e Dr. José Marques Soares que, além de não desistirem do meu trabalho, me motivaram a continuar persistindo e seguindo em frente.

Aos membros do grupo de pesquisa Interfaces. Em especial, aos membros: Jamile, Gilderlane, Polycarpo, Arthur, Luiz e Marcus pelas discussões, parcerias em trabalhos, debates e convivências no "café da química".

À minha irmã, Tácia Camila, pela ajuda na revisão e correção desta dissertação.

Ao Programa de Pós-Graduação em Engenharia de Teleinformática que me deu todo o suporte que necessitei durante o período do mestrado.

À Universidade Federal do Ceará, nas pessoas dos professores do Departamento de Engenharia de Teleinformática do Centro de Tecnologia em Fortaleza que me proporcionaram suporte de conhecimento científico para conclusão do trabalho.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo apoio financeiro por meio da bolsa de estudos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 pelo auxílio financeiro que compete ao Programa de Pós-graduação em Engenharia de Teleinformática (PPGETI-UFC).

“Não vos inquieteis, pois, pelo dia de amanhã,
porque o dia de amanhã cuidará de si mesmo.
Basta a cada dia o seu mal.”

(Mateus 6:34)

RESUMO

O presente trabalho se insere no contexto de identificação de indivíduos por meio de suas imagens faciais 3D. Para identificar os indivíduos, métodos de descrição local de atributos e técnicas de localização de regiões relevantes são exploradas a fim de permitir a classificação mesmo imerso num cenário onde há diferentes tipos de oclusão. Essa circunstância é bastante desafiadora, pois os locais ocluídos implicam em regiões que não possuem correspondência coerente com a amostra de face neutra do indivíduo. As estratégias utilizadas nesse trabalho apresentam resultados comparáveis ao estado da arte, sendo aplicadas em base de dados frequentemente utilizada para estudo de reconhecimento facial com oclusões parciais e obtendo taxa de 100% para os experimentos Neutro-Neutro e taxas de 98.1%, 95.2%, 88.6% e 86.6% nos experimentos utilizando amostras com oclusão do olho pela mão, oclusão dos olhos por óculos, oclusão da boca pela mão e oclusão da face pelo cabelo, respectivamente. Todos esses resultados são de experimentos no rank-1 e utilizando apenas uma única amostra de treino na base (galeria) para cada indivíduo.

Palavras-chave: Imagens. 3D. Face. Oclusão Reconhecimento. Descritores. Relevantes. Classificação.

ABSTRACT

This study revolves around the context of individual recognition based on their 3D facial images. In order to identify individuals, local attribute description methods and relevant region location techniques are explored to allow classification even if immersed in a scenario where there are different types of occlusion. This is a challenging environment, since the occluded regions do not present a consistent match with the individual's neutral face sample. The strategies applied to this study present the state-of-the-art results and have been tested on a commonly used database for facial recognition studies with partial occlusions. The experiments have achieved a recognition rate of 100% for neutral-neutral tests and 98.1%, 95.2%, 88.6% and 86.6% for tests utilizing samples with eye occlusion by hand, eye occlusion by glasses, mouth occlusion by hand and face occlusion by hair, respectively. All results correspond to rank-1 experiments using only a single training sample from the dataset (gallery) for each individual.

Keywords: Images. 3D. Face. Recognition. Occlusion. Attribute. Classification

LISTA DE FIGURAS

Figura 1 – Métodos usados para lidar com oclusões faciais.	17
Figura 2 – Representações de faces 3D	19
Figura 3 – Nuvem de pontos com informação de cor	20
Figura 4 – Tipos de oclusão da Bosphorus.	21
Figura 5 – Espaço de escalas e diferenças de gaussianas	26
Figura 6 – Seleção de ponto relevante no espaço <i>Difference of Gaussian</i> (DoG).	27
Figura 7 – Configurações possíveis do <i>frame</i> intrínseco.	28
Figura 8 – Representação geométrica do Darboux <i>frame</i>	30
Figura 9 – Resultado da aplicação do método em diferentes locais geométricos.	31
Figura 10 – Experimento com diferentes amostragens (cores) e aplicação de ruído (imagem à direita).	32
Figura 11 – Seleção modificada dos pares para o cálculo do histograma.	33
Figura 12 – Diagrama de blocos representativo do <i>pipeline</i>	35
Figura 13 – Redução da densidade com diferentes valores de <i>leaf size</i>	36
Figura 14 – Remoção de <i>outliers</i>	37
Figura 15 – Aplicação do filtro MLS para geração da face modelo.	38
Figura 16 – Regiões utilizadas para o registro.	39
Figura 17 – Registro das amostras de um mesmo indivíduo em relação à sua primeira amostra neutra. Em cada imagem, do lado esquerdo são as faces antes da sobreposição e à direita é o resultado da aplicação do método.	39
Figura 18 – Resumo da aplicação dos métodos percorridos nessa seção.	40
Figura 19 – Vizinhanças utilizadas para o cálculo do descritor na região do nariz.	41
Figura 20 – Descrição dos pontos fiduciais marcados manualmente.	42
Figura 21 – Parâmetros para gerar os pontos radiais.	44
Figura 22 – Projeção dos pontos radiais.	44
Figura 23 – Correspondências encontradas após a detecção.	49
Figura 24 – Correspondência entre os pontos destacando as 20 melhores de verde.	49
Figura 25 – Taxas de reconhecimento variando raios da normal e do descritor na região em torno do nariz (Neutral vs Neutral)	53
Figura 26 – Taxas de reconhecimento variando raios da normal nas regiões da face	53

Figura 27 – Taxas de reconhecimento variando a combinação de raios no experimento de fusão dos descritores.	54
Figura 28 – Resultados dos experimentos em função de k usando os pontos radiais. . . .	57
Figura 29 – Representação dos pontos radiais gerados em um indivíduo da base.	57
Figura 30 – Histograma das escolhas dos pontos na etapa de correspondência e seleção dos k descritores.	58
Figura 31 – Associação dos pontos com o histograma.	59
Figura 32 – Resultados dos experimentos em função de k usando malha radial com pontos esparsos na região nasal.	59
Figura 33 – Resultados dos experimentos em função de k no <i>rank-1</i>	60
Figura 34 – Curvas <i>Cumulative Match Curve</i> (CMC)	63

LISTA DE TABELAS

Tabela 1 – Comparativo melhor raio global vs melhores raios locais.	54
Tabela 2 – Taxas de reconhecimento fazendo a seleção manual das regiões.	55
Tabela 3 – Comparativo de tempo dos dois métodos utilizados no trabalho.	61
Tabela 4 – Comparativo de experimentos <i>rank-1</i> na base Bosphorus.	62

LISTA DE ABREVIATURAS E SIGLAS

AnFM	<i>Annotated Face Model</i>
AvFM	<i>Average Face Model</i>
CMC	<i>Cumulative Match Curve</i>
CP-ICP	<i>Cloud Partitioning - Iterative Closest Point</i>
DoG	<i>Difference of Gaussian</i>
FNTFGP	<i>Fast Nonnegative Tensor Factorization Based On Graph-Preserving</i>
FPFH	<i>Fast Persistent Feature Histogram</i>
ICP	<i>Iterative Closest Point</i>
ISS	<i>Intrinsic Shape Signatures</i>
KNN	<i>K-Nearest Neighbors</i>
LoG	<i>Laplacian of Gaussian</i>
MLS	<i>Moving Least Squares</i>
NARF	<i>Normal Aligned Radial Feature</i>
NMS	<i>Non-Maxima Suppression</i>
PCA	<i>Principal Component Analysis</i>
PCL	<i>Point Cloud Library</i>
PFH	<i>Persistent Feature Histogram</i>
RMSE	<i>Root Mean Square Error</i>
RR	<i>Recognition Rate</i>
SIFT	<i>Scale Invariant Feature Transform</i>
SPFH	<i>Simplified Point Feature Histogram</i>
SVD	<i>Singular Value Decomposition</i>
SVM	<i>Support Vector Machine</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização e problemática	15
1.2	Trabalhos Relacionados	16
1.3	Objetivos Gerais e específicos	18
1.4	Estrutura da dissertação	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Dados 3D	19
2.1.1	<i>Bosphorus</i>	20
2.2	Registro de Nuvem de Pontos	21
2.2.1	<i>ICP</i>	22
2.2.2	<i>CP-ICP</i>	24
2.3	Pontos Relevantes	24
2.4	SIFT	25
2.5	ISS	27
2.6	Descrição Local	29
2.6.1	<i>PFH</i>	29
2.6.2	<i>FPFH</i>	32
2.6.3	<i>Considerações Finais</i>	34
3	METODOLOGIA	35
3.1	Pré-processamento	35
3.1.1	<i>Pré-seleção de pontos</i>	36
3.1.2	<i>Correção de pose</i>	37
3.2	Seleção de regiões de interesse	40
3.2.1	<i>Pontos Fiduciais</i>	41
3.2.2	<i>Pontos Radiais</i>	43
3.2.3	<i>SIFT</i>	45
3.2.4	<i>ISS</i>	45
3.3	Classificação	46
3.3.1	<i>Correspondência Indexada</i>	46
3.3.1.1	<i>Classificadores</i>	46

3.3.2	<i>Filtragem Dinâmica</i>	47
3.4	Considerações Finais	50
4	RESULTADOS	51
4.1	Descrição introdutória aos experimentos	51
4.2	Resultados da classificação	52
4.2.1	<i>Regiões Fiduciais</i>	52
4.2.2	<i>Pontos Radiais</i>	56
4.2.3	<i>Detectores SIFT e ISS</i>	60
5	CONCLUSÕES E TRABALHOS FUTUROS	65
	REFERÊNCIAS	67

1 INTRODUÇÃO

1.1 Contextualização e problemática

Representar dados reais no espaço virtual é algo que tem ganhado bastante atenção nos dias atuais, sobretudo com o avanço das tecnologias de realidade virtual e aumentada que têm crescido bastante nos últimos anos. Uma crescente demanda por aplicações envolvendo reconhecimento facial tem acompanhado essa evolução, pois aplicações em diversas áreas como controle de acesso, investigação forense, sistemas de vigilância, entretenimento e serviços de autoatendimento têm sido cada vez mais requisitadas.

Nas últimas décadas o reconhecimento facial com imagens 2D tem sido amplamente estudado e muitos métodos foram desenvolvidos. Entretanto ainda existem muitas limitações desse domínio que são superadas no campo tridimensional, pois neste último é possível prover informações geométricas mais confiáveis em relação ao domínio 2D. Além disso, as técnicas aplicadas no 3D têm uma menor sensibilidade na extração de certos atributos sob condições adversas de iluminação, escalas e rotação (SOLTANPOUR *et al.*, 2017).

Devido à facilidade de ataques a sistemas de reconhecimento facial 2D, também chamados de ataques de *spoofing*, muitas soluções utilizando dados tridimensionais têm surgido (GALBALLY *et al.*, 2014). A própria natureza da aquisição em si é uma vantagem sobre algumas técnicas aplicadas somente no campo 2D. Em complemento, a independência das condições de iluminação do ambiente para a aquisição dos dados é uma característica importante desses dados, um exemplo de aplicação recente disso são os *smartphones* capazes de desbloquear a tela com reconhecimento facial, mesmo na ausência de luz. Zhou e Xiao (2018) ressaltam que, em muitas aplicações no contexto de reconhecimento facial, as técnicas que utilizam imagens 2D têm dificuldade de reconhecimento em condições adversas de expressões, diferentes poses e cenários de oclusões parciais.

Apesar das técnicas que utilizam dados 3D terem certas vantagens sobre os métodos restritos a imagens 2D, reconhecer um indivíduo sob condições de oclusões parciais em dados tridimensionais ainda é um desafio, pois tais condições produzem uma perda de informação que, dependendo da área ocluída, pode comprometer o método de reconhecimento. Sistemas de autoatendimento em bancos, hospitais ou estabelecimentos de vendas têm sido bastante demandados, sobretudo no contexto atual de pandemia no qual vários países, inclusive o Brasil, estão inseridos. Neste cenário, o reconhecimento do indivíduo mesmo utilizando máscara, por

exemplo, evitaria a remoção desnecessária deste utensílio em certos ambientes em que seu uso é obrigatório. Portanto, este trabalho se insere no contexto em que há uma necessidade de lidar com oclusões parciais 3D e superar problemas de fraudes encontradas nas técnicas aplicadas a imagens 2D.

1.2 Trabalhos Relacionados

Soltanpour *et al.* (2017) classificam o reconhecimento facial em três grupos: holísticos, baseados em atributos e os híbridos. O primeiro se baseia em similaridades globais e utiliza descritores representando a face inteira. Como exemplo, têm-se técnicas como a *Principal Component Analysis* (PCA) que é aplicada considerando todos os pontos da nuvem (RUSS *et al.*, 2006), o *delta faces* que é aplicado diretamente na malha 3D (LI *et al.*, 2013) e o *Fast Nonnegative Tensor Factorization Based On Graph-Preserving* (FNTFGP), desenvolvido por Fu *et al.* (2016), que usa um tensor 4D para classificar expressões. Os métodos baseados em atributos são técnicas que buscam encontrar características locais na face ou focam em regiões específicas como nariz, boca, olhos, etc. Nessa categoria, existem métodos como o desenvolvido por Li e Da (2012) e Lei *et al.* (2013) que fazem combinações de regiões na face. Já (GUO *et al.*, 2016) e (DENG *et al.*, 2020) utilizam pontos chave como forma de descrever as regiões dinamicamente. Os algoritmos híbridos correspondem a uma combinação entre as duas categorias ou a utilização dos dados 2D e 3D (SOLTANPOUR *et al.*, 2017).

Alexandre *et al.* (2020) abordaram aspectos relacionados às regiões de interesse na face, as quais podem ser divididas em três grupos: pontos chave, curvas e superfícies locais; ou combinações destas abordagens. Técnicas baseadas em pontos chave têm o intuito de analisar as vizinhanças desses pontos a fim de descrever as características e atributos dessas adjacências. Zhong (2009) e Gilani *et al.* (2018) utilizam os autovalores da matriz de covariância extraída de uma vizinhança local como forma de selecionar pontos de alta relevância. Darom e Keller (2012) e Li *et al.* (2015) selecionam os pontos utilizando uma adaptação de um algoritmo de detecção projetado inicialmente para imagens 2D.

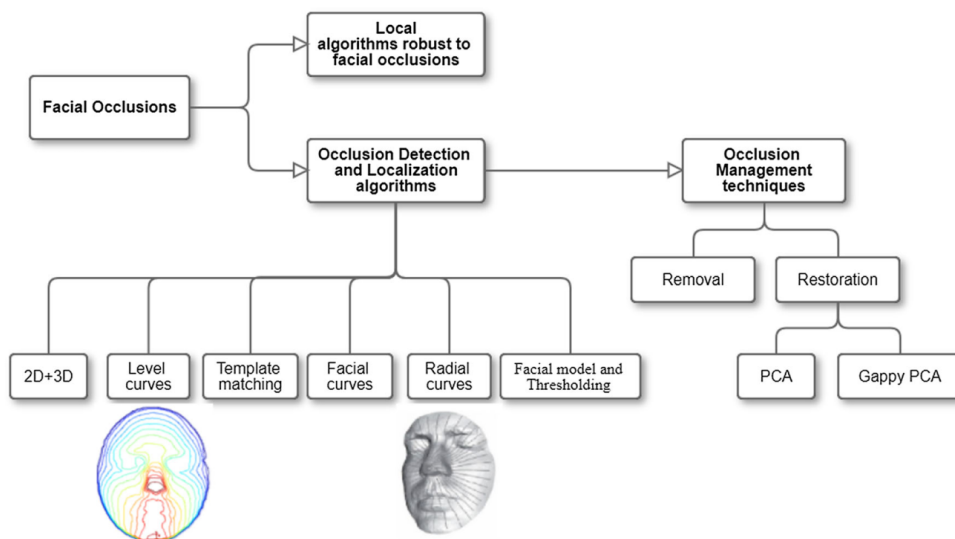
Alguns métodos propõem uma aprendizagem profunda por meio de redes neurais convolucionais. O método de Gilani e Mian (2018) alcança 100% de acurácia em um determinado cenário, entretanto necessita de uma quantidade grande de amostras para treinar a rede neural incluindo amostras com oclusões geradas de maneira artificial. Já os métodos convencionais, em geral, utilizam uma amostra por indivíduo para formar o conjunto de treino e, além disso,

trabalham com uma quantidade menor de atributos. Xu *et al.* (2019) mostram um ganho significativo nas taxas de acurácia ao triplicar a quantidade de amostras no seu conjunto de treino usando redes neurais, entretanto os resultados usando a abordagem de atributos locais mostram taxas próximas de 100% utilizando apenas uma amostra de cada indivíduo para formar a base de treino.

Assim como as faces que possuem expressões, as oclusões podem ser entendidas como uma deformação da face neutra. O uso de estratégias locais pode ser bastante eficaz no contexto em que se têm regiões de deformação, pois permite os algoritmos se adaptarem a essas variações de superfície. Ekenel e Stiefelhagen (2009) relatam que, quando a face é parcialmente ocluída, a performance dos algoritmos de reconhecimento pode cair drasticamente devido a perda de informação discriminante. Segundo Dagnes *et al.* (2019), em cenários não controlados, as oclusões faciais podem comprometer métodos tradicionais de reconhecimento, os quais podem ser explorados por pessoas má intencionadas com o objetivo de enganar sistemas de segurança.

Dada a importância de lidar com oclusões em sistemas de reconhecimento facial, alguns métodos propõem soluções para enfrentar esse problema, seja removendo-as ou construindo técnicas que sejam robustas à deficiência de informação provocada por tais oclusões. A Figura 1 mostra as estratégias de estudos voltados para a oclusão de faces no contexto de reconhecimento de indivíduos.

Figura 1 – Métodos usados para lidar com oclusões faciais.



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

Esta dissertação se enquadra no grupo de técnicas robustas a oclusão utilizando-se estratégias de descrição local sobre pontos específicos da face (pontos chave). A inspiração

vem de métodos que conseguem chegar a taxas próximas ou igual a 100% e são constantemente utilizados como referência na literatura (LI *et al.*, 2015; GILANI *et al.*, 2018; GUO *et al.*, 2016).

1.3 Objetivos Gerais e específicos

De maneira geral, este trabalho se propõem a identificar indivíduos por meio da análise de imagens 3D de suas faces mesmo em condições de oclusão. Embasado nos métodos conhecidos na literatura, supõem-se a possibilidade de realizar a classificação mesmo não utilizando todas as regiões da face. De maneira mais específica, alguns objetivos específicos podem ser apresentados:

1. explorar meios de processar os dados a fim de permitir o reconhecimento de indivíduos sob condições de oclusões parciais;
2. avaliar o desempenho da classificação de faces utilizando a estratégia de descrição local;
3. localizar regiões de interesse propícias para a extração de atributos; e
4. propor soluções de flexibilização do processo de classificação a fim de permitir a identificação do indivíduo mesmo com regiões oclusas.

1.4 Estrutura da dissertação

Esta dissertação está estruturada da seguinte forma, no Capítulo 2 são abordados os aspectos técnicos dos principais métodos utilizados no trabalho e da base de dados na qual foram realizados os experimentos. O modo como foram conduzidos os experimentos e o detalhamento dos processos empregados estão descritos de maneira detalhada no Capítulo 3. No Capítulo 4 são mostrados os resultados provenientes da aplicação da metodologia empregada neste trabalho e também feitas algumas considerações e discussões ao mesmo tempo em que os resultados vão sendo mostrados. Ao final, são catalogadas as conclusões e possibilidades de melhorias do trabalho proposto, apresentadas no Capítulo 5.

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo são apresentados rapidamente alguns conceitos do tipo de dados utilizados no trabalho e apresentada a base de dados na qual a metodologia foi testada. Alguns métodos são explorados com mais detalhes, especificamente uma técnica de registro, um método de descrição local e duas técnicas de seleção de pontos relevantes. A explanação desses algoritmos é importante para fundamentar o uso dessas técnicas na metodologia apresentada no próximo Capítulo e alcançar os objetivos apresentados no Capítulo anterior.

2.1 Dados 3D

Em geral, as superfícies do mundo real podem ser reconstruídas por meio de um conjunto de câmeras 2D com posições distintas (fotogrametria) ou por amostragens feitas por *scanners* 3D. Isso gera o que se chama de nuvem de pontos, que é um conjunto de pontos com coordenadas x,y,z com ou sem informações de cor (r,g,b), dependendo da tecnologia de captura. Algumas técnicas como *mobile laser scanning* e *terrestrial laser scanning* são utilizadas para amostragens de ambientes controlados e geralmente com espaços fechados voltados para escaneamento da estrutura ou de objetos dentro desses locais (LEHTOLA *et al.*, 2017).

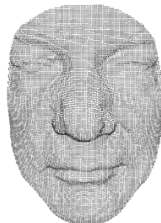
A densidade da amostragem da superfície pode variar dependendo do equipamento e os dados podem ser armazenados de maneira estruturada ou não. A estruturação permite a correta associação com as imagens 2D e isso possibilita o uso da informação de cor nas nuvens e também viabiliza a aplicação de técnicas envolvendo os domínios 2D e 3D.

As imagens faciais 3D podem ser representadas de diversas maneiras: como malhas, mapas de profundidade ou somente os pontos com ou sem informações de cor.

Figura 2 – Representações de faces 3D



(a) Mapa de profundidade.



(b) Nuvem de pontos.



(c) Malha.

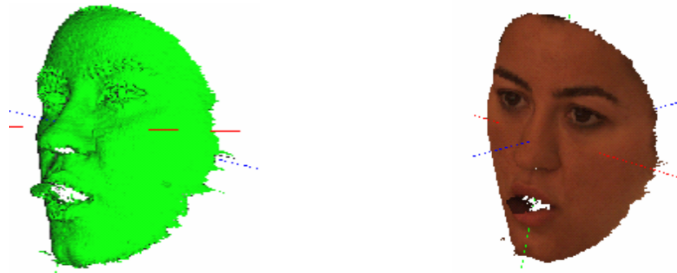
Fonte: (BENNAMOUN *et al.*, 2015).

A imagem da Figura 2a contém informação de profundidade. Esta pode ser definida

como $I = D(i, j)$, i e j são os índices dos pontos na imagem e a função D define o valor dos *pixels*, quantificados em função de suas distâncias ao ponto de aquisição. Na Figura 2b, tem-se um conjunto de pontos com coordenadas x, y, z tendo o ponto de aquisição como origem do sistema de coordenadas. Os dados podem apresentar indicações de interligação entre os pontos de modo a formar uma estrutura de polígonos. Geralmente, é organizada num formato de quadriláteros ou triângulos após a renderização. Um exemplo do resultado final desse processo pode ser visto na Figura 2c.

As nuvens também podem ter informações de cor como na Figura 3. Dependendo do sensor, as informações de cor de cada ponto podem ser capturadas ou a correspondência entre os *pixels* da imagem 2D e os pontos 3D podem ser encontrados.

Figura 3 – Nuvem de pontos com informação de cor



(a) Colorida artificialmente

(b) Com informação de cor 2D

Fonte: (SAVRAN *et al.*, 2008)

2.1.1 Bosphorus

A base, proposta por Savran *et al.* (2008), é uma das mais abrangentes em relação aos tipos de amostras que disponibiliza por sujeito. Ao todo são 4652 amostras faciais de 105 indivíduos com diversas expressões, diferentes poses, e quatro tipos de oclusão. São 60 homens e 45 mulheres, a maioria dos sujeitos são caucasianos e têm entre 25 e 35 anos de idade. Cada face tem seus respectivos pontos fiduciais, marcados manualmente, somando um total de 24. Em algumas amostras não foi possível a marcação de alguns pontos, por conta da ausência de dados (rotação da cabeça de mais de 45 graus) ou no caso das oclusões. A base tem quatro tipos de oclusões (Figura 4) rotuladas como:

1. O_EYE - oclusão do olho com a mão (105 amostras);
2. O_MOUTH - oclusão da boca com a mão (105 amostras);
3. O_GLASSES - oclusão dos olhos pelos óculos (104 amostras); e

4. O_HAIR - oclusão da face pelo cabelo (67 amosras).

Os sujeitos que têm boca e olhos ocluídos ficaram livres para posar da forma como quisessem, desde que ocluindo essas regiões de alguma forma com a mão ou parte dela. Na oclusão por óculos, os sujeitos selecionaram de forma aleatória um conjunto de óculos disponibilizados. Por fim, os indivíduos com cabelos compridos suficiente foram escaneados deixando o seu cabelo ocluir parte da face.

Figura 4 – Tipos de oclusão da Bosphorus.



Fonte: (SAVRAN *et al.*, 2008)

As faces foram capturadas utilizando um sensor comercial, o Inspeck Mega Capturor II 3D, que tem uma resolução de 0,3 mm, 0,3 mm e 0,4 mm em x, y e z , respectivamente. As imagens 2D coloridas tem dimensão de 1600x1200 *pixels*. Após a aquisição, alguns filtros como média e filtros gaussianos são utilizados a fim de remover ruídos. Ao final, cada nuvem de pontos é sub-amostrada e tem aproximadamente 35 mil pontos.

2.2 Registro de Nuvem de Pontos

Basicamente, o registro de nuvens de pontos consiste em posicionar coerentemente, em um mesmo referencial, duas nuvens com origens distintas. Isso pode ocorrer em um cenário em que o ponto de aquisição (câmera) está se movimentando, com o referencial estático e o objeto se movimentando, ou com ambas as duas descrições anteriores. No caso das aquisições faciais, em geral, a câmera é estática e as posições do indivíduo não variam fortemente, pois, em muitas aplicações, existe a necessidade de controlar o ambiente de aquisição para evitar distorções no processo de captura. Nesta dissertação é necessário o registro das nuvens para a etapa de localização dos pontos em distribuição radial. A técnica também melhora a estabilidade dos dados para a extração de atributos, pois minimiza mudanças bruscas de posição entre as aquisições.

Os algoritmos de registro em geral podem ser classificados em dois grupos: métodos com transformação rígida e não rígida. O primeiro assume que as transformações nesse espaço

podem ser modeladas apenas usando os seis graus de liberdade, enquanto que no outro contexto os objetos podem mudar sua forma a cada tempo (BELLEKENS *et al.*, 2014). Na literatura há o método *Iterative Closest Point* (ICP), bastante utilizado para registro de nuvens de pontos. Uma variante deste método, o *Cloud Partitioning - Iterative Closest Point* (CP-ICP), foi utilizada para a correção de pose com a finalidade de registrar as nuvens, mesmo sob condições de oclusão parcial da face.

2.2.1 ICP

Proposto originalmente por Besl e McKay (1992), é um algoritmo iterativo que busca a otimização da matriz de transformação que melhor sobrepõe uma nuvem fonte a uma nuvem de referência. Essa otimização é feita estimando a matriz de transformação a cada iteração. No algoritmo original, a matriz é estimada usando *Singular Value Decomposition* (SVD), método que busca otimizar a soma das distâncias euclidianas entre os pontos correspondentes nas nuvens. Com base nessas correspondências, é possível encontrar a matriz M_c de correlação entre os dois conjuntos. Essa matriz pode ser representada em função da matriz de autovetores e autovalores da seguinte maneira,

$$M_c = USV^T, \quad (2.1)$$

em que U e V são matrizes ortonormais compostas pelos autovetores extraídos de $M_c M_c^T$ e $M_c^T M_c$ respectivamente. S é uma matriz diagonal na qual a diagonal principal contém os autovalores de U e S (ambos têm os mesmos autovalores), que são os valores singulares. Segundo Bellekens *et al.* (2014), a solução do problema de mínimos quadrados é definida pela matriz de rotação R_q^p que transforma o conjunto de pontos \mathcal{P} para o espaço de referência em que se encontra o conjunto de pontos \mathcal{Q} . A matrix é dada por

$$R_q^p = UV^T. \quad (2.2)$$

O vetor de translação é o vetor entre os centroides, dado por $v_t = c_q - R_q^p c_p$. Em essência, o algoritmo pode ser resumido nos passos descritos a seguir.

1. **Correspondências:** dado um critério de distância, para cada ponto $p_i \in \mathcal{P}$ é encontrado um ponto $q_j \in \mathcal{Q}$ de tal forma que a distância entre p_i e q_j seja menor do que a distância de p_i para qualquer outro ponto em \mathcal{Q} ;

2. **Calculo da Matriz:** é estimada a matriz por meio da decomposição de valores singulares (equação 2.2);
3. **Transformação:** a matriz encontrada é aplicada à nuvem fonte (\mathcal{P}); e
4. **Convergência:** se um critério de convergência não for atendido, as etapas anteriores são repetidas. Em geral, esse critério se baseia no *Root Mean Square Error* (RMSE) entre a iteração atual e a anterior. O parâmetro (ϵ) indica o grau de mudança que é tolerado, pois, se o erro não varia muito, isso pode indicar que as transformações não estão tendo um efeito significativo. Ele pode ser calculado com base no determinante da matriz de transformação ou pode ser a raiz da diferença entre o RMSE da iteração atual e o da iteração anterior.

A partir do trabalho de Besl e McKay (1992), muitas variantes surgiram. Entretanto é importante salientar que no ano anterior Chen e Medioni (1991) propuseram um método também iterativo para registro de *range images*, que são imagens 2D que possuem informação de profundidade, ou seja, os dados também podem ser representados de certa forma no domínio 3D, alguns autores chamam de domínio 2.5D.

Algumas abordagens diferentes desse algoritmo surgiram ao longo do tempo, Turk e Levoy (1994) e Masuda *et al.* (1996) fazem uma amostragem dos pontos a cada iteração, o primeiro de maneira uniforme e o segundo de maneira aleatória. Weik (1997) seleciona os pontos com alta intensidade do gradiente e Rusinkiewicz e Levoy (2001) usam uma estratégia de seleção das correspondências com base na normal.

Segal *et al.* (2009) propuseram uma generalização do ICP, adicionando um modelo probabilístico na etapa de minimização da distância. Yang *et al.* (2013) utiliza o ICP no núcleo de sua técnica como forma de encontrar uma solução global ótima e Pereira *et al.* (2015), com o método CP-ICP, realiza um pré-processamento nos dados e registra as nuvens com um tempo de processamento ao menos vinte vezes menor.

Recentemente, Lu *et al.* (2019) utilizou pontos relevantes como forma de ajudar na etapa de seleção dos pontos correspondentes. Alguns trabalhos utilizam redes neurais e descritores locais como suporte na etapa de correspondência (ZHANG *et al.*, 2019; DENG *et al.*, 2019).

2.2.2 CP-ICP

Proposto por Pereira *et al.* (2015), essa técnica tem como um dos objetivos reduzir o tempo de processamento, mantendo o mesmo nível de qualidade do ICP. Para isso, o método utiliza uma estratégia de particionamento das nuvens. Inicialmente, particiona-se, ao longo de um eixo principal, as duas nuvens a serem registradas, então cada par de sub-nuvem é submetido ao ICP. O método permite obter tantas matrizes de transformação quantas forem as partições definidas no início do algoritmo. Tais matrizes são então avaliadas e uma é escolhida com a finalidade de registrar a nuvem fonte em sua integridade.

Considerando que o número de pontos em cada nuvem é N_p e N_q , o custo para computar as correspondências, considerando a força bruta, é da ordem de $O(N_p N_q)$. Entretanto, com uso do particionamento a complexidade passa a ser $k_{sub} O(N_p N_q / k^2)$. Dependendo da quantidade de pontos da nuvem e da quantidade de sub-nuvs (k_{sub}) a diferença pode ser considerável.

Adaptou-se o particionamento para os três eixos principais: frontal, sagital e transversal. Ao final, escolhe-se a matriz de transformação associada à melhor correção de pose. O CP-ICP minimiza, portanto, uma função de custo equivalente a do ICP clássico na proporção do número de partições, com a particularidade de que os pontos a serem associados serão aqueles presentes nas sub-nuvs. A função de custo é definida por

$$F_j(\vec{q}) = \frac{1}{N} \sum_{i \in N} \| \vec{m}_i - (R\vec{s}_i + T) \| . \quad (2.3)$$

Nesta equação, $F_j(\vec{q})$ é a função a ser minimizada, N a quantidade de pontos da nuvem, \vec{m}_i é o i -ésimo ponto da nuvem de referência associado ao i -ésimo ponto da nuvem fonte, \vec{s}_i , transformado pela matriz de rotação R e transladado pelo vetor T . O termo j é o índice da sub-nuvem.

A próxima seção aborda sobre regiões discriminantes na face, pontos comumente conhecidos como pontos chave ou pontos relevantes. Tais regiões devem se sobressair de sua vizinhança local e guiar os algoritmos de extração de atributos a fim de gerar atributos com alta discriminância para o processo de classificação.

2.3 Pontos Relevantes

Quando se aborda sobre relevância em imagens é possível se deparar com a seguintes questões: Existem pontos específicos que se destacam mais do que outros aos olhos humanos? Quais atributos e características possuem?

Para responder essas perguntas, David Marr, neurocientista britânico, foi pioneiro ao propor um estudo aprofundado comparando o sistema visual humano com sistemas computacionais. Em seu livro (MARR, 1982), bastante conhecido no contexto de visão computacional, ele discute características, restrições e o processo de aquisição da informação visual do sistema biológico humano. Um série de análises comparativas são feitas em diferentes níveis, desde transistores até complexas teorias da computação e desde o processo de captação visual da retina até processamentos da informação em regiões complexas do cérebro como neocórtex e o arquicórtex.

Encontrar pontos de alta relevância na face significa detectar pontos que possuem certas propriedades que os dão alta distintividade de sua vizinhança local e alta estabilidade em várias instâncias do modelo (CHIAM *et al.*, 2017). Também conhecidos como pontos salientes, pontos característicos ou pontos de interesse, são principalmente utilizados como suporte ao registro de nuvens (QUAN; MA, 2019) ou na construção de descritores locais num contexto de reconhecimento de objetos (KIM *et al.*, 2017). Alguns métodos detectam pontos de interesse em imagens de profundidade como o *Normal Aligned Radial Feature* (NARF), proposto por Steder *et al.* (2011). Nessa técnica, a nuvem de pontos é transformada para uma imagem de profundidade e, após isso, os pontos de interesse são calculados sobre a imagem.

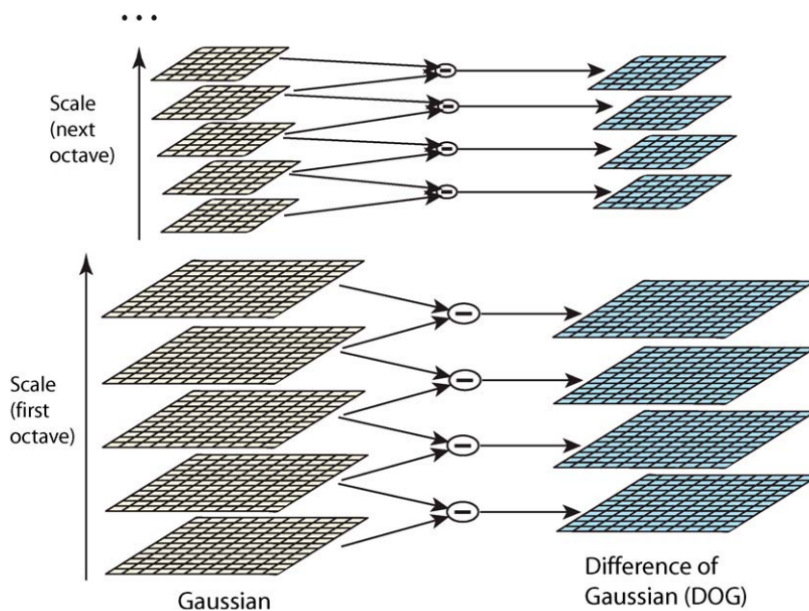
Em princípio, pontos salientes devem ser persistentes e detectáveis mesmo em diferentes condições de escala, rotação ou posição no espaço. A seguir, duas técnicas de estimação de pontos característicos bastante conhecidas na literatura são exploradas. Ambas são disponibilizadas pela *Point Cloud Library* (PCL) (RUSU; COUSINS, 2011) e utilizadas na etapa de detecção de pontos relevantes sobre os quais é definida a vizinhança de pontos necessária para a extração de atributos. A seguir, dois métodos de detecção, o *Scale Invariant Feature Transform* (SIFT) e o *Intrinsic Shape Signatures* (ISS), são descritos.

2.4 SIFT

Proposto inicialmente em (LOWE, 1999), o SIFT é uma técnica direcionada essencialmente para imagens 2D e propõe um conjunto de características invariantes à escala, translação, rotação e parcialmente invariantes às mudanças de iluminação e projeção afim. Posteriormente, o mesmo autor fez uma análise mais profunda de seu método e aplicou algumas melhorias na estabilidade e invariância dos descritores (LOWE, 2004). Apesar desta técnica ter duas etapas, detecção de pontos relevantes e descrição local, apenas a etapa de detecção é abordada.

Em síntese, o SIFT detecta pontos de alta distintividade dentre vários espaços de escala. Utiliza-se a *Difference of Gaussians* (DoG) como uma aproximação do *Laplacian of Gaussian* (LoG) para detectar os pontos característicos. Inicialmente, são geradas várias imagens em escalas diferentes, no caso do trabalho de Lowe (2004), estas são geradas convolvendo a imagem em gaussianas com diferentes valores de escalas, representadas pela variável σ . Para cada imagem, chamada também de *octave*, é gerado um conjunto de imagens convolvidas novamente por gaussianas para gerar um subespaço de escalas dentro de cada *octave*. Após isso, a diferença de gaussianas é computada para cada *octave* (imagem à direita da Figura 5).

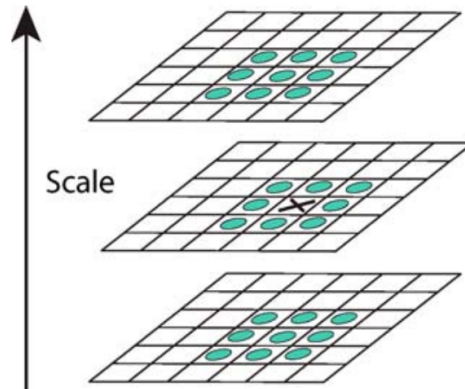
Figura 5 – Espaço de escalas e diferenças de gaussianas .



Fonte: (LOWE, 2004)

Os pontos relevantes são escolhidos dentro do espaço das diferenças de gaussianas calculadas. Em cada espaço, cada ponto é comparado com seus oito vizinhos dentro do mesmo espaço DoG e com os nove vizinhos no espaço superior e inferior como mostra na Figura 6. Se o valor do pixel em questão for maior que todos os vizinhos ou menor que todos eles, o mesmo é selecionado como ponto relevante.

Figura 6 – Seleção de ponto relevante no espaço DoG.



Fonte: (LOWE, 2004)

Esse método foi adaptado para nuvem de pontos e disponibilizado pela biblioteca PCL. Na adaptação, as escalas são geradas usando o método *voxel grid*. A vizinhança é definida pelos pontos que estão dentro de uma esfera centrada no ponto em questão. A normal de cada ponto é calculada e utilizada no 3D como sendo o equivalente à intensidade do pixel para imagens 2D. A implementação padrão utiliza os valores de RGB de cada ponto para calcular a intensidade, entretanto é escolhido a normal, pois durante todo o processo do trabalho apenas as informações das coordenadas x , y , z são utilizadas. O espaço de diferença de gaussianas é mantido, entretanto estendido para o domínio 3D. Ainda nessa adaptação desenvolvida pela PCL, adiciona-se uma etapa posterior de *Non-Maxima Suppression* (NMS) para filtrar os pontos selecionados. Basicamente, toma-se como entrada os pontos detectados como relevantes e, para cada ponto, é verificado se existe, dentro de uma vizinhança definida por uma esfera de raio R_{nms} , algum outro ponto detectado com um valor da medida de relevância maior que a do ponto em questão. Se existir, o ponto analisado é removido do conjunto de pontos relevantes. Este valor de relevância depende das métricas usadas no algoritmo de detecção.

2.5 ISS

Assim como SIFT, o método proposto por Zhong (2009) consiste em um *pipeline* completo de extração de características, detecção e descrição. Entretanto, apenas a parte de detecção será explorada.

No ISS primeiramente é definido um intrínseco *frame* de referência F_i para cada ponto p_i , considerando os k_{sup} pontos dentro de uma esfera de raio R_{sup} (raio de suporte). Então, com esses pontos é computada a matriz de covariância ponderada,

$$M_{cov}(p_i) = \frac{1}{W} \sum_{j=1}^{k_{sup}} w_j (p_i - p_j)(p_i - p_j)^T,$$

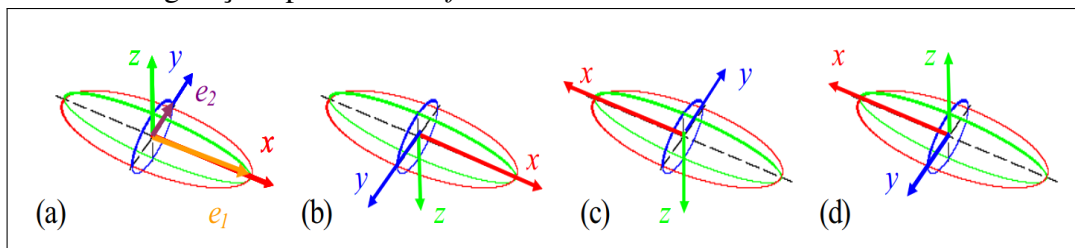
$$W = \sum_{j=1}^{k_{sup}} w_j$$

Em que w_j é o fator de ponderação de cada ponto vizinho a p_i . Esse fator é inversamente proporcional à quantidade de pontos na vizinhança esférica de p_j delimitada por um raio $r_{densidade}$. Após isso, os autovalores $\lambda_i^1, \lambda_i^2, \lambda_i^3$ e os autovetores v_i^1, v_i^2, v_i^3 são calculados de tal forma que $\lambda_i^1 > \lambda_i^2 > \lambda_i^3$. Os expoentes referem-se aos três possíveis autovalores e autovetores que provêm da matriz de covariância, pois os dados possuem apenas três atributos (x, y, z) . Então um *frame* é definido como sendo $F_i = \{v_i^1, v_i^2, (v_i^1 \times v_i^2)\}$, com p_i na origem.

Para eliminar pontos com *frames* ambíguos em locais de simetria são usados os limiares γ_{21} e γ_{32} baseados nos autovalores. Se $\frac{\lambda_i^2}{\lambda_i^1} < \gamma_{21}$ e $\frac{\lambda_i^3}{\lambda_i^2} < \gamma_{32}$ o ponto é selecionado. Dessa forma, é possível regular os parâmetros para encontrar pontos em locais de alta curvatura e forte variação nos eixos principais.

Pelo fato de a matriz computar a direção de variância dos dados no espaço 3D e ser possível ter mais de uma solução para o cálculo dos autovetores, a orientação dos eixos pode apresentar ambiguidades de 180° (Figura 7), ou seja, podem ser obtidas quatro configurações possíveis do *frame* para a mesma solução.

Figura 7 – Configurações possíveis do *frame* intrínseco.



Fonte: (ZHONG, 2009)

Algumas medidas podem ser tomadas para remover algumas ambiguidades, como verificar a posição do sensor para resolver uma das direções, por exemplo.

Na implementação do método, a primeira solução de *frame* encontrada é a utilizada e o fator de ponderação w_i é desconsiderado. É necessário definir também uma quantidade mínima de pontos para o cálculo de F_i . Ao final, uma etapa de NMS é adicionada assim como na implementação do SIFT.

2.6 Descrição Local

De maneira geral, um descritor é uma representação simplificada de imagens na forma de um vetor contendo valores numéricos que descrevem aspectos geométricos ou topológicos dos dados. Em resumo, pode-se elencar um conjunto de elementos que devem ser observados na extração de atributos como:

1. os atributos dos pontos locais devem significativamente se sobressair de sua vizinhança;
2. invariância à rotação, escala, translação e iluminação;
3. robustez ao ruído; e
4. distintividade. Num contexto de classificação multi-classe, os descritores devem ter bastante coesão intra-classe e alta diferenciação inter-classe (KAZMI *et al.*, 2013).

Descritores locais são bastante utilizados na identificação de correspondências entre pontos de nuvens distintas no contexto de registro (ZHANG *et al.*, 2019). Alguns autores utilizam esses descritores para o reconhecimento de objetos ou em recuperação de dados 3D nos algoritmos de busca (ZHONG, 2009). No contexto de reconhecimento facial, os trabalhos de Deng *et al.* (2020) e Li *et al.* (2015) constroem vetores robustos e representativos das faces com o intuito de identificar indivíduos. Ambos utilizam a estratégia de extrair pontos relevantes e montar o vetor de características com base na vizinhança de cada ponto. Lei *et al.* (2013) descreve que um bom descritor representativo da face deve ser não ambíguo, isso significa que faces 3D de diferentes indivíduos devem produzir descritores bastante diferentes entre si. Na classificação, deve haver uma alta diferenciação entres os descritores de faces de indivíduos diferentes, mas uma forte similaridade e coesão entre os descritores de faces do mesmo indivíduo.

Há ainda descritores baseados em histograma de atributos que são amplamente utilizados no contexto de registro de cenas e objetos, o *Persistent Feature Histogram* (PFH) e o *Fast Persistent Feature Histogram* (FPFH). Neste estudo, somente o FPFH foi utilizado como descritor de atributos, porém é necessário descrever inicialmente o método que o deu origem.

2.6.1 PFH

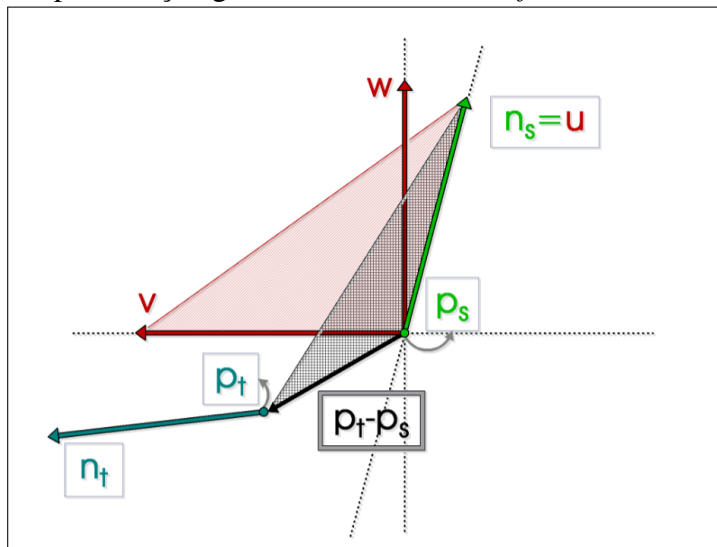
Proposto por Rusu *et al.* (2008), o método consiste na construção de um histograma de atributos com base na vizinhança de um determinado ponto na nuvem. No seu trabalho posterior (RUSU *et al.*, 2009), o autor fez algumas melhorias no método quanto à discriminância e também à complexidade do processamento.

Para cada ponto p_i são selecionados todos os pontos dentro de uma esfera de raio r_{feat} . Seja p_s e p_t um par de pontos dentro desse conjunto, tal que $s \neq t$, $s < t$ e suas respectivas normais sejam n_s e n_t . Se $\langle n_s, p_t - p_s \rangle \leq \langle n_t, p_t - p_s \rangle$, então o ponto p_s será a origem. Assumindo um par de pontos em que p_s é a origem, um Darboux *frame* (Figura 8) é definido pelos eixos u, v, w sendo $u = n_s, v = u \times (p_t - p_s), w = u \times v$. A partir disso, os quatro atributos do algoritmo são computados como propôs Wahl *et al.* (2003):

$$\begin{aligned}
 f1 &= \langle v, n_t \rangle; \\
 f2 &= \|p_t - p_s\|; \\
 f3 &= \langle u, p_t - p_s \rangle / f2; \\
 f4 &= atan(\langle w, n_t \rangle, \langle u, n_t \rangle),
 \end{aligned}
 \tag{2.4}$$

Em que f_1 é o cosseno do ângulo entre a componente v (de vermelho na Figura 8) e a normal ao ponto p_t . A componente v tem uma dependência da normal do ponto p_s e a distância entre os dois pontos, pois v é um vetor na direção ortogonal a n_s e o vetor $p_t - p_s$. Note que f_2 é basicamente a distância entre o par de pontos e f_3 é o cosseno do ângulo entre o vetor $p_t - p_s$ e n_s . Já o atributo f_4 é o único ângulo e depende das normais dos pontos e o eixo w do *Darboux frame*. Todos os atributos dependem essencialmente da distância entre pontos e suas normais, portanto as características tendem a ser robustas à rotação e translação.

Figura 8 – Representação geométrica do Darboux *frame*



Fonte: (RUSU *et al.*, 2008)

Pra cada atributo f_i de cada par (p_s, p_t) é feita a seguinte quantização,

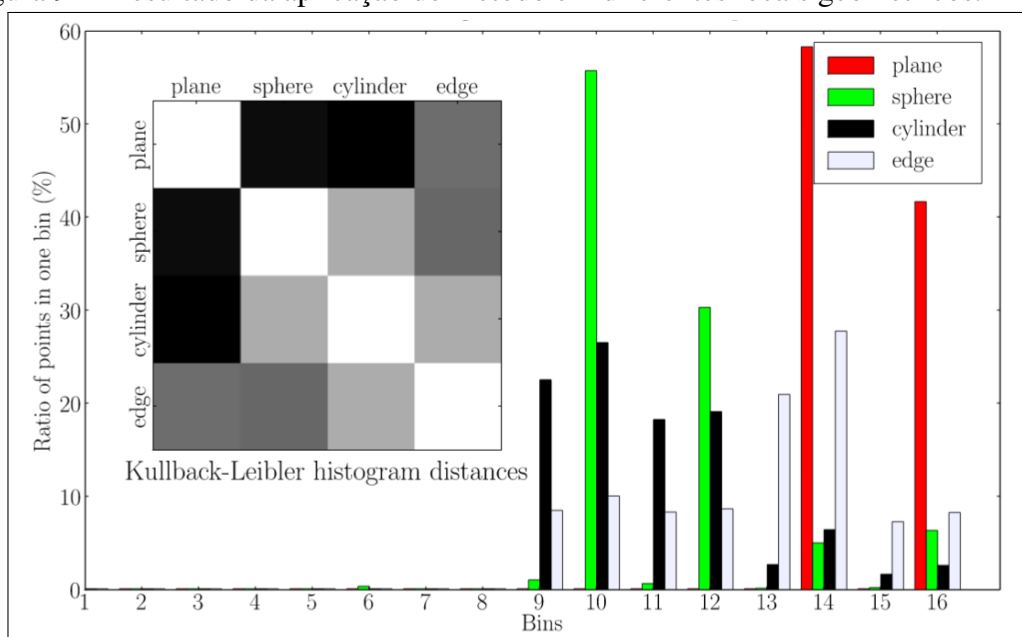
$$idx = \sum_i^{i \leq 4} step(t_i, f_i) \cdot sub^{i-1}.
 \tag{2.5}$$

Verifique que a quantização é binária, a função *step* é 1 no caso de $f_i > t_i$ e 0 caso contrário. Então, cada um dos 4 atributos pode assumir apenas valores 0 ou 1 dependendo do limiar t . Os intervalos pra cada atributo são

$$\begin{aligned} f1 &: [-1, +1], \\ f2 &: [0, 2R], \\ f3 &: [-1, +1], e \\ f4 &: [-\pi, \pi]. \end{aligned} \tag{2.6}$$

Os limiares foram então escolhidos no centro de cada intervalo, 0 para f_1, f_3, f_4 e r para f_2 . Segundo a equação 2.5 o número de índices do histograma é sub^4 e sub é o número de subdivisões geradas pelos limiares em cada atributo f_i . No caso das Figuras 9 e 10 sub é 2 e o número total de combinações é $2^4 = 16$. Ou seja, existem 16 possíveis combinações de f que podem surgir para cada par (p_s, p_t) dentro da vizinhança. Um histograma é gerado contando o número de aparição de cada combinação. Por exemplo, se em um determinado par (p_1, p_2) os atributos tiverem valor $\{1, 1, 1, 0\}$, o idx na equação 2.5 terá valor final igual a 7 e o índice de número 7 do histograma será incrementado. A Figura 9 mostra um exemplo do histograma calculado para diferentes formas geométricas (definido pelas cores). Como na equação 2.5, os valores podem variar entre 0 e 15, a figura a seguir está transladada, mostrando os valores entre 1 e 16.

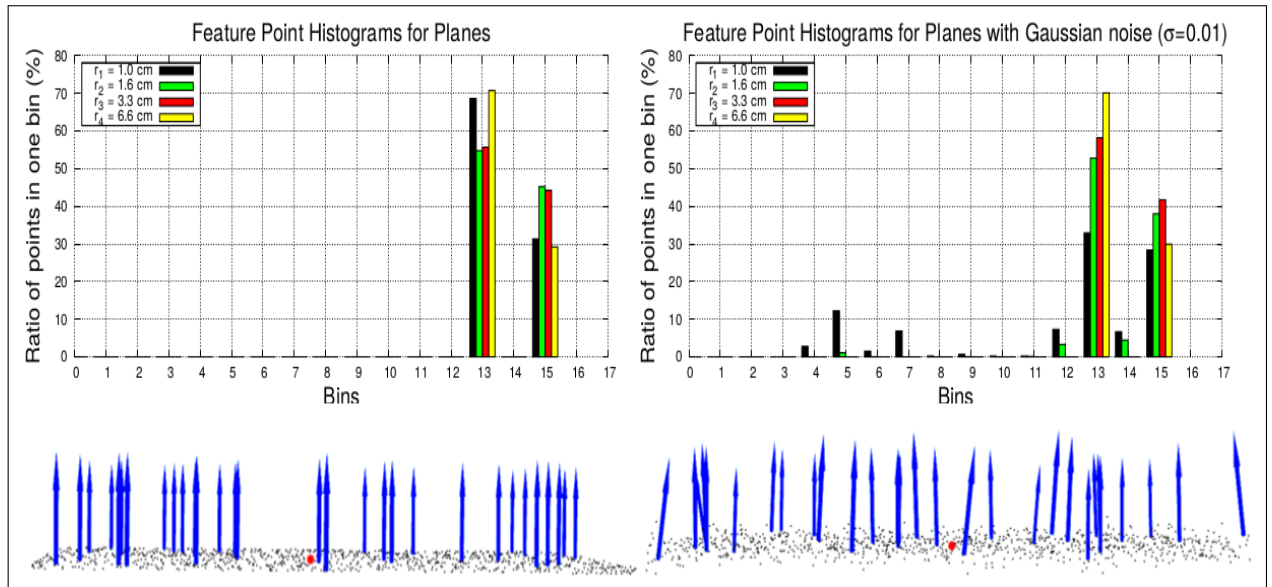
Figura 9 – Resultado da aplicação do método em diferentes locais geométricos.



Fonte: (RUSU *et al.*, 2008)

Algumas características desse descritor, como a sua tolerância ao ruído e a invariância à densidade, são fundamentais para aplicação em dados de baixa resolução adquiridos por sensores de baixo custo. Tais propriedades são reveladas nos testes em que é adicionado um ruído nas nuvens e feita a análise do método em diferentes subamostragens. Na Figura 10, do lado esquerdo, tem-se o método sendo aplicado sobre o ponto vermelho indicado na Figura inferior, em diferentes condições de amostragens (definidas pelas cores). Do lado direito, verifica-se o mesmo experimento sobre o ponto vermelho com as diferentes amostragens, entretanto a nuvem de entrada é submetida a um ruído gaussiano.

Figura 10 – Experimento com diferentes amostragens (cores) e aplicação de ruído (imagem à direita).



Fonte: (RUSU *et al.*, 2008)

2.6.2 FPFH

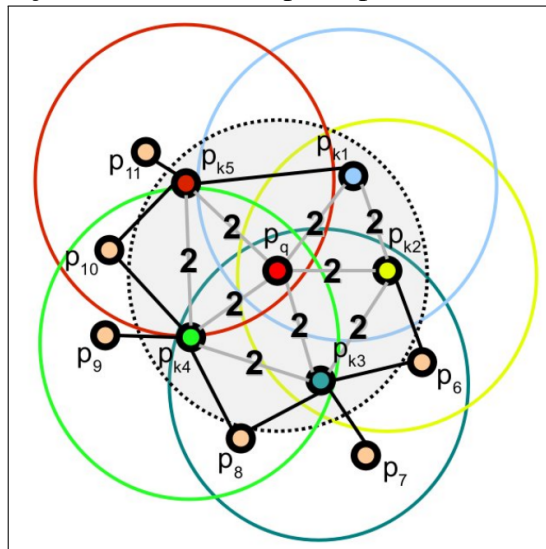
Neste método, Rusu *et al.* (2009) sugere algumas mudanças em relação ao seu trabalho anterior. A primeira diz respeito à quantidade de atributos, pois o atributo f_2 é dispensado devido os seus testes mostrarem que sua exclusão não provoca decaimento da robustez do método. Desta forma, a quantidade de combinações fica $2^3 = 8$. Esse conjunto reduzido de 3 atributos é chamado no trabalho de *Simplified Point Feature Histogram* (SPFH). A outra modificação é na seleção dos pares de pontos dentro da vizinhança. Ao invés de selecionar todos os pares, são selecionados somente alguns pontos em duas etapas. Na primeira, para cada ponto p_i o SPFH é calculado somente entre o ponto em questão e seus k_n vizinhos mais próximos. A Figura 11

ilustra essa mudança mostrando a ligação dos pontos que possuem uma relação (não são todos os pares). Já na segunda etapa, para cada ponto vizinho p_k , são encontrados os seus k_p vizinhos mais próximos e calculado o SPFH da mesma forma, mas com um fator de ponderação, dado por

$$FPFH(p) = SPFH(p) + \frac{1}{k_p} \sum_{i=1}^{k_p} \frac{1}{w_k} \cdot SPFH(p_k). \quad (2.7)$$

Neste fator de ponderação, w_k é a distância (em uma determinada métrica) entre o ponto p_k e o ponto central. O raio utilizado para a busca de k_n e k_p é o mesmo. A Figura a seguir ilustra a seleção dos pares de pontos.

Figura 11 – Seleção modificada dos pares para o cálculo do histograma.



Fonte: (RUSU *et al.*, 2009)

Os experimentos no trabalho de Rusu *et al.* (2009) mostraram que mesmo com essas alterações e redução da quantidade de atributos para três, o método FPFH consegue ser tão discriminante quanto o PFH. Porém, alcançando os resultados de maneira bem mais rápida devido a redução no espaço de busca dos pares.

Na implementação do FPFH o número de subdivisões usado é igual a 11, ou seja, cada atributo poderia ter 11 valores diferentes. Assim, as combinações possíveis seriam de 1331 (11^3). Para evitar a enorme quantidade de zeros que o histograma total poderia conter, computou-se o histograma de cada característica individualmente (11 atributos) e após isso os 3 histogramas são concatenados. Dessa maneira, tem-se um vetor de atributos final com 33 valores ao invés de 1331.

2.6.3 Considerações Finais

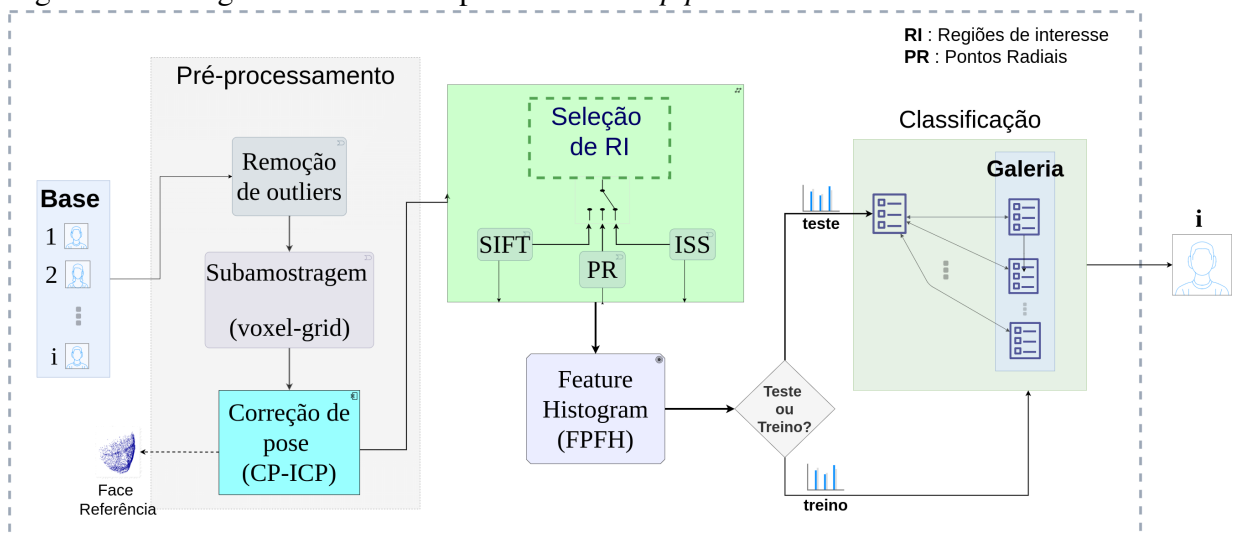
Em síntese, a base de dados utilizada foi apresentada e as principais técnicas usadas no trabalho foram detalhadas. O método do CP-ICP, duas técnicas de detecção de pontos relevantes e um descritor foram apresentados com o intuito de dar um suporte técnico para a leitura dos próximos capítulos em que essas técnicas são utilizadas. Esses métodos fazem parte do núcleo de processamento da metodologia que será apresentada. Entretanto, antes de aplicá-los, os dados são processados a fim de prepará-los para as etapas seguintes. Além disso, é introduzida uma maneira mais simples de localizar pontos na face para a extração de atributos.

3 METODOLOGIA

Neste Capítulo a metodologia de investigação do sistema de reconhecimento facial é apresentada detalhadamente, descrevendo questões e aspectos essenciais para cada uma das etapas sequenciadas em *pipeline*. Uma questão relevante é a correção de pose das amostras e, diante da diversidade metodológica existente na literatura relacionada, é conveniente a apresentação e discussão da abordagem adotada.

A organização do Capítulo, portanto, tem uma conexão com o *pipeline*, o qual é mostrado na ilustração da Figura 12 e representa, em síntese, um processo de classificação supervisionado, com suas etapas de pré-processamento, seleção de regiões de interesse, extração de características e classificação. Em algumas etapas, mais de uma abordagem foi utilizada, como no caso da seleção de regiões de interesse, nesse momento são apresentados dois métodos da literatura e também uma estratégia de seleção desenvolvida pelo autor.

Figura 12 – Diagrama de blocos representativo do *pipeline*.



Fonte: Elaborado pelo autor

3.1 Pré-processamento

Antes de submeter os dados ao método de localização das regiões e extração das características faciais para a classificação, um pré-processamento foi aplicado na base de dados que é utilizada. Este se divide em duas etapas, a aplicação de técnicas de filtragem e correção de pose.

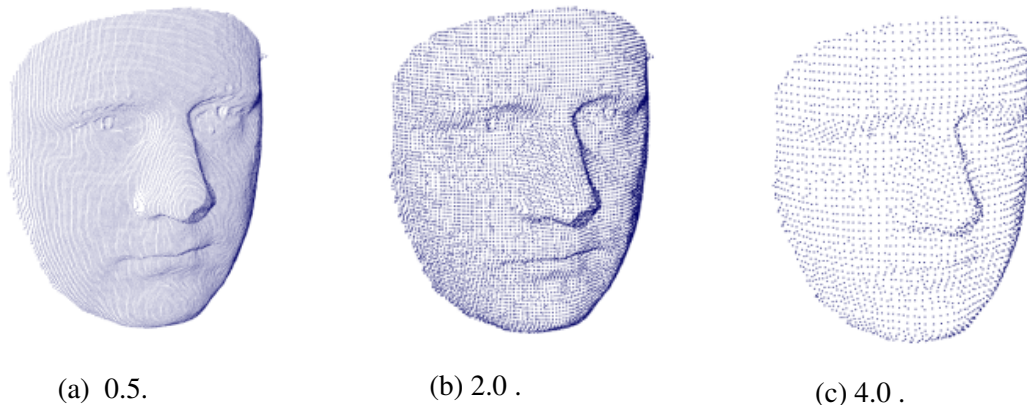
3.1.1 Pré-seleção de pontos

Inicialmente, foram aplicados dois métodos de pré-processamento. O primeiro consiste na aplicação de um filtro a fim de remover possíveis *outliers* que possam estar nas nuvens. Para cada ponto, é verificado se possui um mínimo de pontos m_p dentro de uma esfera de raio r_p , centrada no ponto em questão. Se a quantidade de pontos pré estabelecida for menor que m_p , o ponto é considerado *outlier* e é removido da nuvem.

O segundo processo é uma redução na densidade de pontos utilizando o *voxel grid*. Essa técnica consiste na construção de uma estrutura de paralelepípedos organizados de forma hierárquica no espaço que envolve a nuvem. O método precisa da definição das dimensões do menor paralelepípedo que a hierarquia pode possuir, são os valores L_x , L_y e L_z . Isso é necessário para definir o critério de parada do algoritmo. Ao final, a técnica retorna os centroides de cada paralelepípedo criado na hierarquia ou pode ser o ponto mais próximo do centroide, dependendo da aplicação desejada.

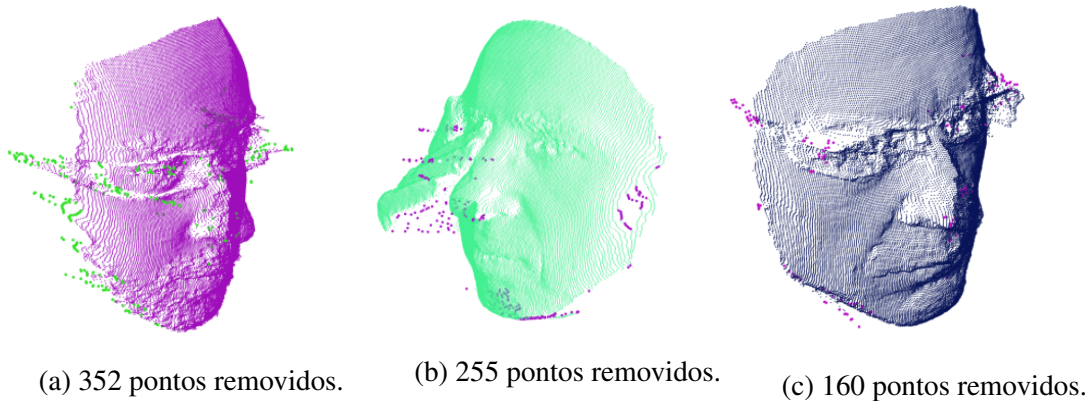
As nuvens foram sub-amostradas a fim de reduzir o tempo de processamento e também verificar o desempenho do método em dados com baixa resolução. Para a filtragem, foi utilizado o valor de 2.0 para todas as dimensões L_x , L_y e L_z . A média de tamanho das nuvens caiu de 35 mil para 10 mil pontos aproximadamente e o tempo médio da aplicação dos dois métodos em uma nuvem foi 0.33 segundos. A Figura a seguir demonstra os dois processos aplicados em indivíduos da base.

Figura 13 – Redução da densidade com diferentes valores de *leaf size*.



Fonte: Elaborado pelo autor.

Figura 14 – Remoção de *outliers*.



Fonte: Elaborado pelo autor.

3.1.2 Correção de pose

Antes de submeter as nuvens ao processo de extração de características e classificação, é realizada a correção de pose da base. Uma nuvem é utilizada como referência para registrar toda a base.

Alguns autores (ALYUZ *et al.*, 2012a; ALYUZ *et al.*, 2012b) constroem uma face de referência, chamada *Average Face Model* (AvFM), gerada a partir de uma média aplicada sobre as faces de treinamento no espaço 2.5D, também conhecido como mapas de profundidade. Dessa forma, aplica-se a seguinte equação: $M = \frac{1}{N} \sum_{n=1}^N D_i$, sendo N o número de amostras e D_i o i -ésimo mapa (DAGNES *et al.*, 2019).

Outros pesquisadores, como Gilani e Mian (2016), Lei *et al.* (2016) utilizam o *Annotated Face Model* (AnFM). Desenvolvida por Passalis *et al.* (2005), esta técnica é baseada em uma malha 3D facial média construída a partir de dados estatísticos.

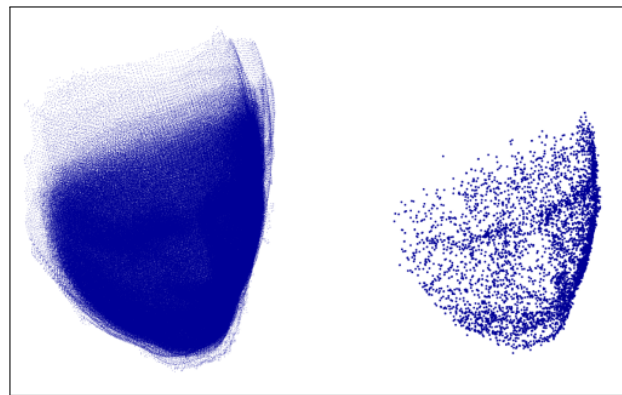
A face de referência foi criada utilizando diretamente os pontos x, y, z . Para isso, foi aplicado um método, desenvolvido por (ALEXA *et al.*, 2003), que aproxima a superfície com polinômios usando o *Moving Least Squares* (MLS). Essa técnica foi escolhida sobretudo por não necessitar das imagens de profundidade ou da reconstrução de malha para sua execução. Além disso, o algoritmo se enquadra bem a esse trabalho, pois as informações de entrada conhecidas são somente as coordenadas x, y, z da face.

Nessa estratégia, com base nos pontos de entrada, os dados são aproximados a uma superfície polinomial que, depois disso, sofre uma reamostragem. Para cada ponto, são encontrados os k_n vizinhos com distância menor que um valor de limiar r_n e, a partir desses pontos, um polinômio é aproximado. A subamostragem é feita de forma iterativa e o critério

para calcular os pontos que vão pertencer a nuvem final é construído com base no raio r_n e em uma estrutura do tipo *octree* montada dentro dessa vizinhança.

Para gerar a face de referência, são concatenadas todas as amostras da galeria, formada pela primeira amostra da face neutra (sem expressões, oclusões ou grandes rotações) de cada indivíduo. Após a junção dessas amostras em uma única nuvem de pontos, é realizada a subamostragem usando o MLS com $r_n = 0.2mm$ (obtido empiricamente). A Figura a seguir ilustra o resultado da aplicação do filtro.

Figura 15 – Aplicação do filtro MLS para geração da face modelo.

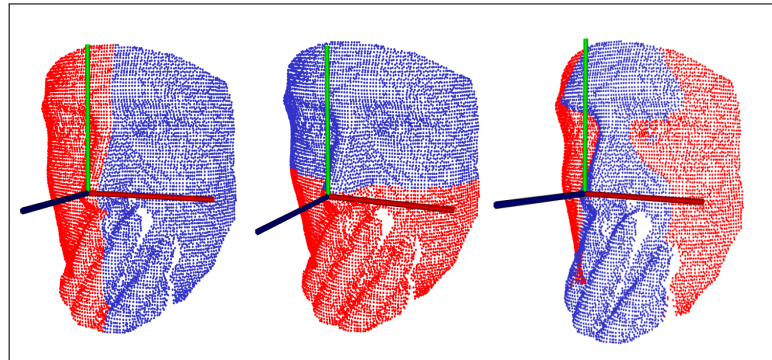


Fonte: Elaborado pelo autor.

Após a definição da face modelo, todas as amostras são submetidas ao registro usando o CP-ICP (Seção 2.2.2). Este método é muito útil no caso de amostras com regiões oclusas devido a sua capacidade de trabalhar com partes da nuvem, pois o algoritmo pode excluir partes ocluídas da nuvem do processo de registro. No caso de um registro global, submetendo a nuvem inteira ao ICP, as regiões de deformação poderiam comprometer o desempenho do algoritmo, pois tais locais não possuem uma correspondência coerente de pontos. Isso poderia fazer com que as partes que possuem correta correspondência (regiões não ocluídas) não tivessem uma boa sobreposição dos pontos.

Para um par de nuvens (fonte e referência) são feitos cortes em cada direção (\hat{x} , \hat{y} e \hat{z}) e realizado o registro com o ICP para cada par de sub-nuvem gerada a partir dos cortes. Foi adotado $k_{sub} = 2$, parâmetro que define a quantidade de sub-nuvs por eixo, pois não está no escopo desse trabalho tratar regiões oclusas que tenham extensão superior à metade da superfície facial. Sendo assim, têm-se 6 nuvens particionadas, 2 por eixo (Figura 16) e, portanto, 6 matrizes de transformação. Então, é escolhida a matriz de transformação que possuir o menor RMSE para registrar a nuvem.

Figura 16 – Regiões utilizadas para o registro.

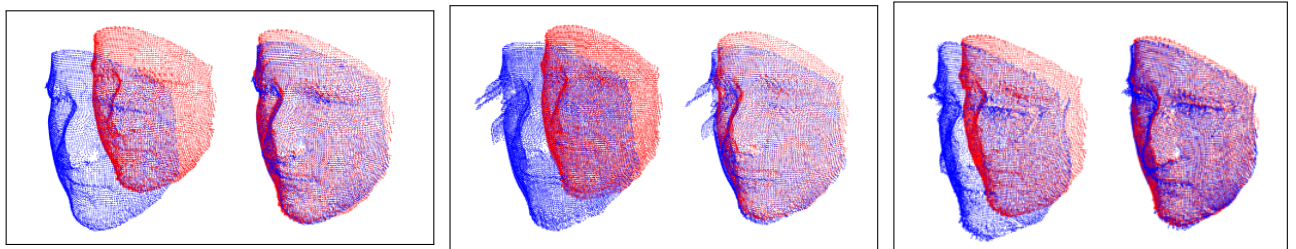


(a) Cortes usando o CP-ICP.

Fonte: Elaborado pelo autor.

Para o ICP processado em cada sub-nuvem, adotou-se o erro como $\varepsilon=0.0001$ (ver Seção 2.2.1) e o número máximo de iterações igual a 200. A Figura 17 é um exemplo do resultado do registro na amostra neutra e nas amostras oclusas de um indivíduo da base.

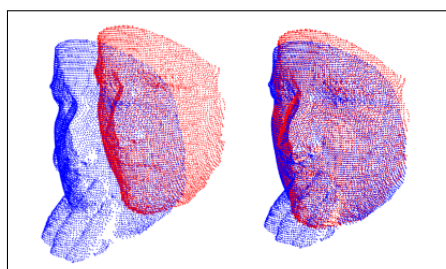
Figura 17 – Registro das amostras de um mesmo indivíduo em relação à sua primeira amostra neutra. Em cada imagem, do lado esquerdo são as faces antes da sobreposição e à direita é o resultado da aplicação do método.



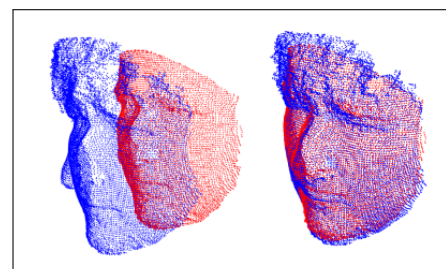
(a) Neutral.

(b) Eye.

(c) Glasses.



(d) Mouth.



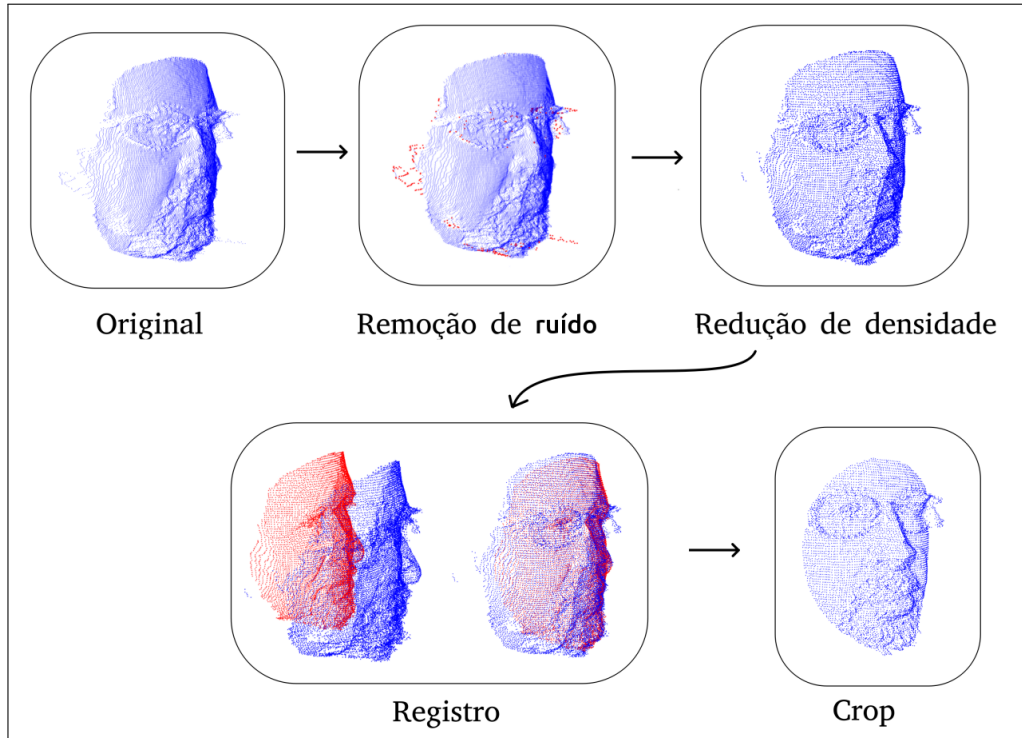
(e) Hair.

Fonte: Elaborado pelo autor.

Após o registro, todas as nuvens passam por mais um processo de filtragem a fim de focalizar nas partes mais centrais e discriminantes da face e eliminar possíveis partes ruidosas nas extremidades das amostras. A filtragem consiste em eliminar os pontos fora de uma esfera de raio $80mm$ com o centro na extremidade do nariz. Essa etapa é chamada de *crop* e pode ser

visualizada na Figura 18, a qual mostra a ordem dos processos realizados nessa seção. A partir daqui, a base com todos esse processos aplicados é utilizada ao longo das próximas etapas, que consiste na localização das regiões, extração de características e classificação.

Figura 18 – Resumo da aplicação dos métodos percorridos nessa seção.



Nota – A nuvem com cor vermelha na imagem do registro é a nuvem neutra do mesmo indivíduo. Apesar dela não ter sido a nuvem usada como referência para o alinhamento, optou-se por mostrá-la para se ter uma noção do resultado final da transformação feita nas nuvens.

Fonte: Elaborado pelo autor.

3.2 Seleção de regiões de interesse

As etapas de pré-processamento descritas anteriormente produzem um conjunto ainda extenso de dados e portanto passíveis de uma redução que favoreça a classificação. Cabe-se, então, uma estratégia de seleção de regiões de interesse, a partir das quais será calculado o vetor de atributos que descreve as amostras de face para o classificador. Nessa etapa também é descrito detalhadamente o conjunto de parâmetros necessários para reproduzir as estratégias de localização implementadas.

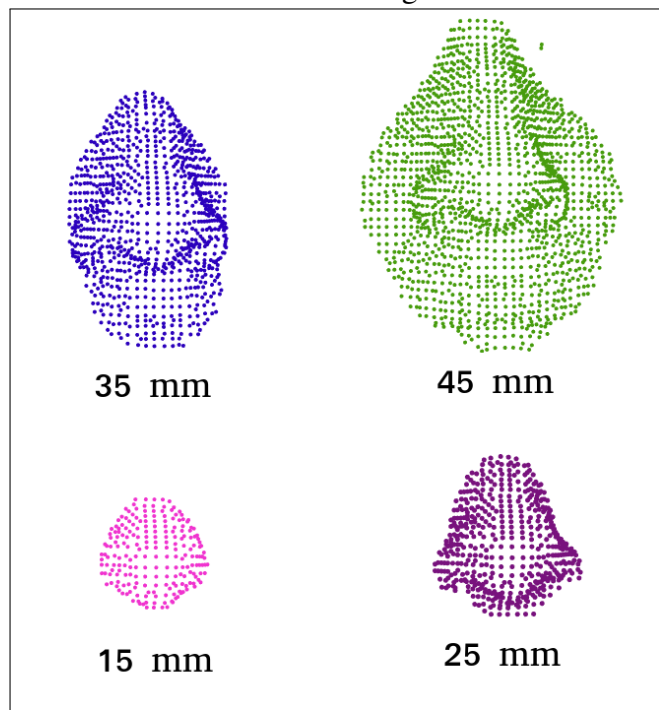
3.2.1 Pontos Fiduciais

Nas primeiras investigações, os pontos fiduciais marcados na base Bosphorus foram utilizados com a finalidade de verificar a capacidade de discriminância do descritor, porém essa competência será explorada com mais detalhes no próximo Capítulo.

Inicialmente, a extração de atributos foi feita em torno da região nasal por ser uma região de estabilidade e, portanto, estar sujeito a variações pequenas da distribuição dos pontos de uma aquisição para outra (DRIRA *et al.*, 2009). Essa região é definida por um dos pontos fiduciais, a extremidade do nariz marcada nas faces.

Para calcular o vetor de características, é necessário definir uma região de interesse, uma vizinhança. Para um determinado ponto, todos os vizinhos dentro de uma esfera de raio R_{feat} são considerados para o cálculo do descritor. Um valor de raio R_{normal} precisa ser também definido para estimar a normal de cada ponto, pois o descritor precisa desse atributo para a construção do histograma (equação 2.4). Os raios foram delimitados em pequenos conjuntos para o teste. $R_{features} = \{15, 25, 35, 45\}$ e $R_{normals} = \{5, 10, 15, 25\}$. Assim, 16 combinações desses dois parâmetros serão utilizadas para realizar alguns dos experimentos. A Figura 19 dá uma noção do espaço de atuação do descritor na região nasal para o conjunto $R_{features}$.

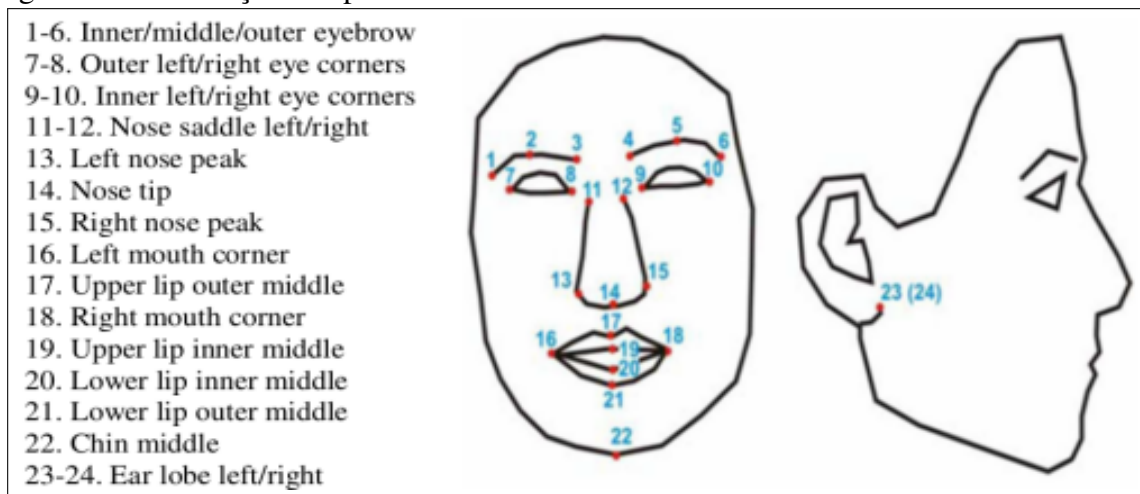
Figura 19 – Vizinhanças utilizadas para o cálculo do descritor na região do nariz.



Fonte: Elaborado pelo autor.

Para aumentar a discriminância da face, a investigação foi expandida para considerar outras regiões, pois realizar a extração de atributos baseando-se apenas em um local é inviável, dado que essa região poderia estar sujeita a oclusão. A base Bosphorus possui 24 pontos marcados manualmente mas, dependendo do tipo da amostra, algumas regiões podem não ter as marcações como no caso das oclusões. Para minimizar o tempo de análise, foram escolhidos, além da extremidade nasal, mais 8 pontos dos 24 mostrados na Figura 20 para compor as regiões que serão experimentadas. Alguns foram renomeados para facilitar a associação ao longo do trabalho.

Figura 20 – Descrição dos pontos fiduciais marcados manualmente.



Fonte: (SAVRAN *et al.*, 2008)

- | | |
|------------------|----------------|
| 1. nose_tip - 14 | 6. eye_li - 8 |
| 2. mouth_cd - 21 | 7. eye_le - 7 |
| 3. nose_cu - 17 | 8. eye_ri - 9 |
| 4. mouth_l - 16 | 9. eye_re - 10 |
| 5. mouth_r - 18 | |

Para cada uma dos pontos escolhidos foram realizados 16 experimentos, cada um com uma combinação de raio diferente de acordo com os conjuntos $R_{features}$ e $R_{normals}$. Em cada região, o vetor possui 33 atributos como explanado no Capítulo anterior. Após isso é realizado um experimento que consiste na fusão dos descritores, ou seja, fundir regiões a fim de dar mais discriminância ao vetor de características. Essa estratégia de combinar regiões é utilizada por Siqueira *et al.* (2018) e Lei *et al.* (2013) e tem produzido resultados comparáveis ao estado da arte. A junção dos vetores consiste numa simples concatenação destes e mantendo a mesma

ordem de concatenação para todas as amostras. Dessa forma, os atributos ficam indexados da mesma maneira em todas as amostras e as correspondências entre os vetores de atributos são mantidas para a etapa de classificação. Nos experimentos utilizando as amostras neutras, todos os vetores foram concatenados, resultando num vetor representativo com 297 atributos (9 regiões x 33). Para os experimentos com oclusão foram selecionadas manualmente somente algumas regiões dependendo do tipo de oclusão, para verificar a possibilidade de classificação com regiões faltantes. Tais regiões são definidas com mais detalhes nos resultados.

3.2.2 Pontos Radiais

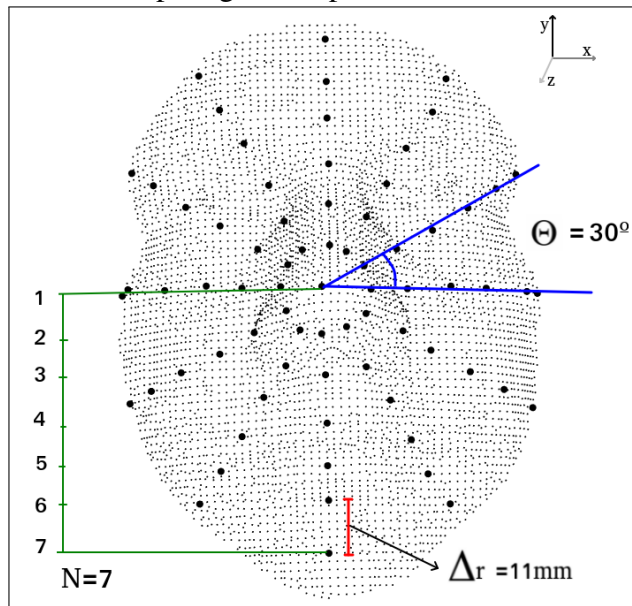
Nessa estratégia, os pontos são posicionados sobre a face de maneira radial e têm como posição base a extremidade do nariz por causa da sua localização centralizada. A dependência do algoritmo a esse ponto não é tão crítica, pois detectá-lo é uma tarefa bem resolvida na comunidade científica. (Kyong I. Chang *et al.*, 2006; BAGCHI *et al.*, 2012). Dependendo do controle do cenário de aquisição e da natureza da captura, seria possível até aproveitar-se da profundidade dos pontos em relação ao plano da câmera para estimá-lo (LEE; YI, 2003). Então, partindo deste ponto os outros pontos são indexados.

Há diversas formas de gerar e distribuir os pontos na superfície facial. Na metodologia empregada nessa etapa existem algumas restrições e definições de parâmetros que precisam ser bem descritos afim de permitir a reprodução da metodologia.

Inicialmente, é necessário que a nuvem esteja corretamente alinhada no plano xy e direcionada para o eixo z . Então, no mesmo plano xy da extremidade nasal, os outros pontos são gerados em níveis circulares. Para cada nível, os pontos são discretizados em coordenadas polares dadas por $(Nose_x + Rsen(\phi), Nose_y + Rcos(\phi))$, sendo $\phi = (L \times 360/\theta)$, tal que $L = \{0, 1, 2, 3, \dots, \theta - 1\}$. $Nose_x$ e $Nose_y$ são as coordenadas x e y da extremidade do nariz, assim, os pontos gerados possuem a mesma coordenada z deste ponto, pois estão projetados no mesmo plano xy do nariz (Figura 22). A cada nível, o raio aumenta de Δr , sendo zero o valor inicial.

A malha foi construída em função de três parâmetros: espaçamento angular dos raios (θ) em graus, numero de níveis (N), e espaçamento radial (Δr) em milímetros. A Figura 21 ilustra espacialmente tais parâmetros, tendo a configuração $\theta = 30^\circ$, $N = 7$ e $\Delta r = 11mm$. Esses parâmetros foram definidos de forma empírica e com auxílio visual.

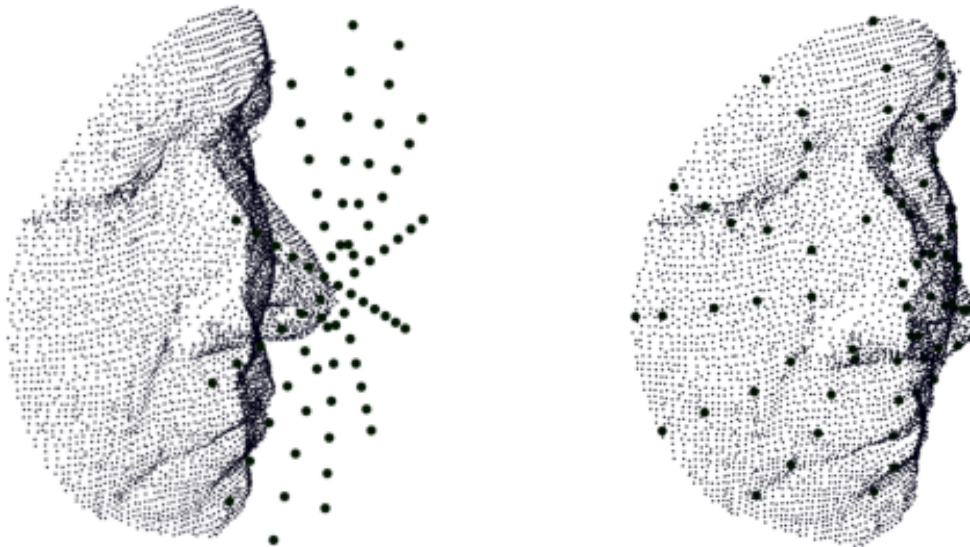
Figura 21 – Parâmetros para gerar os pontos radiais.



Fonte: Elaborado pelo autor

Após definir os pontos radiais no plano, é feita uma busca para encontrar o ponto da face mais próximo a cada ponto da malha, desconsiderando a coordenada z . A Figura a seguir mostra o resultado antes e depois da busca.

Figura 22 – Projeção dos pontos radiais.



(a) Pontos radiais no plano do nariz.

(b) Pontos encontrados na face após a busca.

Fonte: Elaborado pelo autor.

3.2.3 SIFT

Para essa estratégia de localização dos pontos foi escolhido um dos detectores mais conhecidos no contexto de reconhecimento de objetos em geral, o SIFT. Primeiramente, é preciso lembrar que o algoritmo original desenvolvido por Lowe (2004) é composto por duas partes, detecção e descrição, entretanto é utilizado somente a parte de detecção de pontos relevantes que foi adaptado para 3D. Diferente de muitas adaptações para 3D, a utilizada não necessita de reconstrução de malha, isso diminui a complexidade para encontrar os pontos, entretanto é necessário escolher os valores dos parâmetros que mais se adaptem a base de dados.

Para a aplicação do detector é necessário equalizar quatro parâmetros de entrada, o número de *octaves* (N_{oct}), número de escalas por *octave* (N_{esc}), a escala inicial (E_0) e o limiar de contraste (Th_{cont}). Para os experimentos com esse detector foram escolhidos $N_{oct} = 7$, $N_{esc} = 8$, $E_0 = 4$ e $Th_{cont} = 0.0001$. Nessa adaptação é preciso ainda definir que atributo será considerado para ser o equivalente à intensidade do *pixel* em imagens 2D. Nesse caso, o valor da normal é utilizado e o raio de abrangência para o seu cálculo é o mesmo que foi definido para o descritor. Com essa configuração, a quantidade média de pontos detectados nas faces fica em torno de 97. Essa média é a soma de todos os pontos localizados em todas as amostras dividido pelo número de amostras.

3.2.4 ISS

Assim como o SIFT, o ISS é uma técnica de extração de características composta por duas partes, detecção de pontos relevantes e descrição da região circunvizinha aos pontos relevantes, como visto no Capítulo anterior. Entretanto, será utilizado apenas a parte de detecção.

Para a aplicação do detector foi necessário definir cinco parâmetros, estes foram equalizados empiricamente para uma nuvem e replicados para todas as outras faces. Em resumo, os valores estão enumerados a seguir,

1. raio de suporte ($R_{sup} = 10$): raio de definição da vizinhança de pontos considerados para computar a matriz de covariância;
2. $N_{min} = 5$: a quantidade mínima de pontos que a vizinhança deve ter para computar a matriz;
3. $\gamma_{21} = 0.975$: razão entre o segundo e o primeiro autovalor;
4. $\gamma_{32} = 0.097$: razão entre o terceiro e o segundo autovalor;
5. $r_{nms} = 5$: raio usado para definir a vizinhança da etapa de NMS.

3.3 Classificação

A exemplo do que foi feito para a definição das regiões de interesse, há também uma variedade metodológica, a qual diz respeito à sujeição de oclusões e a forma como o extrator de atributos e o classificador lidam tais áreas ocluídas.

O escopo dessa seção se divide em duas estratégias: correspondência indexada e filtragem dinâmica. A primeira consiste em indexar as regiões para classificação sem se preocupar com possíveis distorções (oclusões) na face que provocariam alterações no vetor de características. Ficaria a critério do classificador lidar com essas perturbações. Na segunda estratégia, esse problema foi trazido para a etapa de correspondência, filtrando algumas regiões para compor o vetor de características a fim de prevenir o classificador de ter que lidar com essas oscilações indesejadas que poderiam surgir por causa das oclusões ou até mesmo pequenas flutuações nos músculos mesmo em aquisições neutras da face.

3.3.1 Correspondência Indexada

No procedimento, é montada uma matriz de dados, em que cada linha corresponde a uma amostra de face e cada coluna representa os atributos correspondentes ao vetor de característica extraído e as informações do indivíduo. Essa abordagem foi utilizada para os testes com amostras neutras usando os 9 pontos fiduciais escolhidos. Em alguns testes com amostras oclusas também foi utilizada essa abordagem, porém as regiões são selecionadas manualmente para manter a indexação do vetor de atributos. Após fazer a concatenação dos descritores de cada amostra, tem-se um vetor de $33 \times N_r$ atributos para cada face, sendo N_r a quantidade de descritores utilizados para a concatenação.

Antes da classificação é feito um pré-processamento de normalização simples dos dados estabilizando a média em zero e tornando a variância unitária. Esse Processo é feito utilizando o método de Yeo e Johnson (2000) que está disponível na biblioteca *sklearn* (PEDREGOSA *et al.*, 2011).

3.3.1.1 Classificadores

Para avaliar os descritores utilizando essa estratégia de localização das regiões de maneira indexada, foram escolhidos dois classificadores conhecidos na literatura, o *K-Nearest Neighbors* (KNN) e *Support Vector Machine* (SVM), ambos utilizados também por Ahdid *et al.*

(2017) no contexto de reconhecimento facial.

Em problemas de classificação de alta dimensão com poucas amostras, o classificador SVM é bastante indicado. No experimentos realizados, uma única amostra por classe é utilizada e o vetor de atributos tem um tamanho considerável, 297. Inicialmente, o método foi projetado para ser um classificador linear binário de vetores de alta dimensão que busca separar as classes, estimando um hiperplano entre as classes que maximiza a distância entre os vetores mais próximos ao hiperplano em cada classe (LEI *et al.*, 2013). A forma como é feito o treinamento e a descrição matemática podem ser vistos com mais detalhes no trabalho de Mousavi *et al.* (2008).

O SVM pode ser estendido para um modelo de classificação não linear usando funções de transformação, chamadas de funções de *kernel*, entretanto o modelo linear se ajustou melhor aos dados. Como esse é um problema de classificação multi-classe, a técnica é utilizada no modo "um contra todos", ou seja, o método é aplicado considerando uma classificação binária entre cada classe e todas as outras.

O KNN foi também utilizado como forma de avaliação pela sua simplicidade e sua grande utilização na literatura. Para esse caso, foram utilizadas as distâncias de Manhattan e Euclidiana que podem ser obtidas fazendo $p=1$ e $p=2$, respectivamente, na equação generalizada da distância de Minkowski,

$$D(X, Y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}, \quad \forall X, Y \in \mathbb{R}^n. \quad (3.1)$$

No caso específico da abordagem de classificação utilizada, em que a base de treino possui uma única amostra por classe, o KNN transforma-se em um simples estimador de vizinho mais próximo.

3.3.2 Filtragem Dinâmica

O intuito do método é conseguir classificar um indivíduo mesmo em meio a oclusões na face. Para essa tarefa, é necessário que a face oclusa de um indivíduo tenha uma alta similaridade com a face neutra desse mesmo indivíduo e baixa compatibilidade com as faces neutras de outros indivíduos. Alguns trabalhos aplicam a detecção e descrição de atributos locais e utilizam também estratégias para selecionar descritores que sejam altamente discriminantes em seus processos (DENG *et al.*, 2020; LEI *et al.*, 2016).

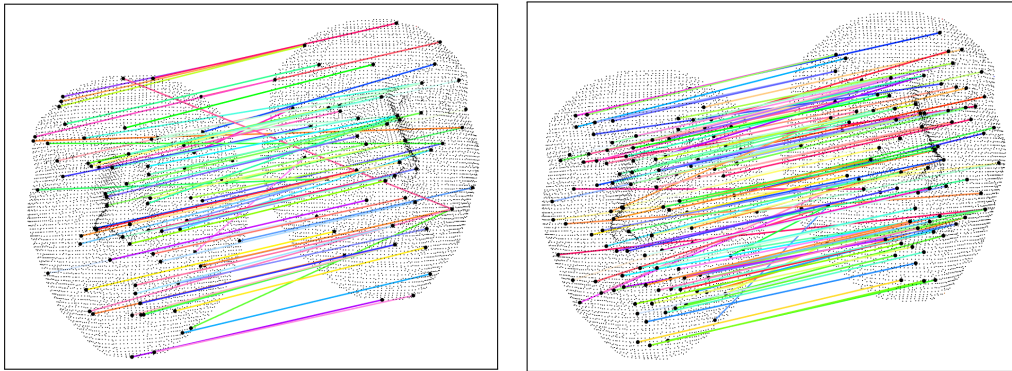
Inspirado no trabalho de Guo *et al.* (2016), adotou-se uma estratégia de correspondência dinâmica da seguinte maneira: primeiro é selecionada uma face de prova i na base de teste. Esta será testada contra cada uma das faces na galeria que, no caso da Bosphorus, são 105. Então, para cada par de nuvens, os seguintes passos devem ser seguidos:

1. são computados todos os descritores de cada ponto localizado na face da galeria e na face de teste;
2. se necessário, são encontradas as correspondências entre cada um dos descritores nas duas amostras;
3. são selecionados os k pares de vetores que tiverem menor distância entre si;
4. com os pares escolhidos, é calculada uma métrica que servirá como medida de dissimilaridade entre a face de prova e a face testada na galeria que consiste na média das distâncias de Manhattan entre cada vetor (μ_d).

Após obter todas as métricas para a face i , a identificação é feita simplesmente escolhendo a face que tem o menor valor de μ_d . No trabalho de Guo *et al.* (2016), essa métrica não teve um bom desempenho, taxa de 42.17% de reconhecimento apenas. Porém, no próximo Capítulo serão mostrados resultados com taxas acima de 90% utilizando essa métrica, pois a técnica apresentada na dissertação se diferencia desta última tanto no método de descrição e detecção como também na forma de seleção dos atributos.

A etapa 2 é realizada somente para o SIFT e o ISS, pois diferente da malha radial, os pontos não estão indexados, logo não é conhecida a correspondência entre os descritores. Nessa etapa, para cada descritor da nuvem fonte é encontrado o seu vizinho mais próximo na nuvem alvo. A nuvem com a menor quantidade de descritores locais é escolhida como a nuvem fonte para a realização da busca. A Figura a seguir mostra um exemplo das correspondências encontradas em duas amostras neutras do mesmo indivíduo usando as técnicas de detecção.

Figura 23 – Correspondências encontradas após a detecção.



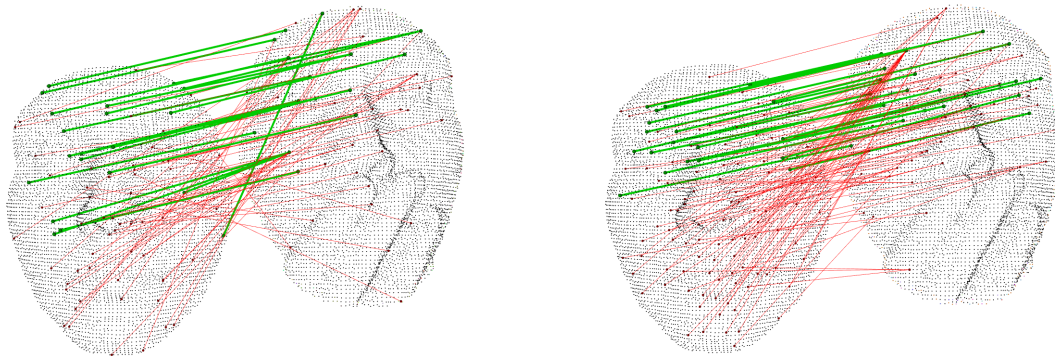
(a) SIFT

(b) ISS

Fonte: Elaborado pelo autor

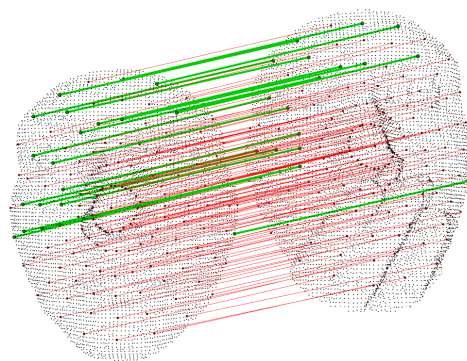
A etapa de seleção das correspondências é uma etapa muito importante do trabalho, pois é nela que o algoritmo ganha flexibilidade para ser tolerante às amostras com oclusão. Isso acontece pelo fato de os descritores encontrados sobre as regiões oclusas divergirem bastante dos encontrados sobre as mesmas regiões nas amostras neutras. Isso gera uma correspondência incoerente entre os vetores de atributos. A Figura 24 ilustra as correspondências com amostras de oclusão na boca utilizando as três estratégias abordadas e destacando de verde as 20 correspondências mais próximas.

Figura 24 – Correspondência entre os pontos destacando as 20 melhores de verde.



(a) SIFT.

(b) ISS.



(c) Malha radial, $\theta = 15^\circ$.

Fonte: Elaborado pelo autor.

Nas três abordagens é possível verificar que as regiões de oclusão não ficam entre as melhores escolhas, fazendo com que na classificação, os descritores encontrados na área oclusa não contribuam pra o cálculo da métrica final. Essa etapa concede ao algoritmo uma certa tolerância à oclusão, entretanto é necessário ajustar o valor de k , pois se for muito grande, falsas correspondências podem ser consideradas no cálculo da métrica, ao passo que se seu valor for muito pequeno, a quantidade de informações pode não ser suficiente para classificar o indivíduo. No Capítulo seguinte, essa análise é feita para um conjunto finito de valores de k , dado por $k_s = \{2, 5, 10, 20, 30, 40, 50\}$.

3.4 Considerações Finais

Neste Capítulo, o *pipeline* foi descrito detalhadamente com o objetivo de permitir a reprodução de cada etapa do trabalho. A metodologia pode ser sintetizada em três grandes blocos. O primeiro, o pré-processamento, consiste nas etapas de remoção de *outliers*, redução da densidade e correção de pose. Neste último, uma adaptação foi feita com a finalidade de corrigir o posicionamento das faces mesmo sob condições de oclusões parciais. O segundo bloco apresenta a descrição detalhada dos parâmetros necessários para a utilização dos dois métodos de detecção de pontos relevantes (discutidos no capítulo anterior) e também apresenta uma forma de localizar os pontos a partir da extremidade do nariz (desenvolvida pelo autor). Por fim, duas maneiras de classificar faces 3D com base nos atributos geométricos são descritas. A primeira consiste em realizar a classificação unindo todos os atributos da face de forma indexada e a outra mostra um forma de selecionar os atributos antes de partir para a classificação de fato. Em síntese, a discussão aqui trazida, portanto, serve de suporte para a compreensão seja dos resultados a seguir apresentados em tabelas e gráficos, seja da importância do sistema proposto para a comunidade científica interessada.

4 RESULTADOS

Neste Capítulo são mostrados os resultados dos experimentos realizados em cada etapa da pesquisa. De forma quantitativa e qualitativa são mostrados e avaliados detalhadamente os resultados de cada experimento das etapas do *pipeline* proposto e feitas algumas discussões sobre os dados apresentados.

4.1 Descrição introdutória aos experimentos

Todos os cenários de experimentos mostrados ao longo dessa seção se inserem no contexto de identificação, por isso as taxas de acurácias aqui explanadas estão relacionadas sempre a esse cenário. A base de dados Bosphorus, introduzida por Savran *et al.* (2008), foi escolhida, pois traz condições bastante desafiadoras no contexto de oclusão e possui ampla aceitação na literatura como forma de validação de métodos de reconhecimento facial 3D. Foram utilizadas 680 amostras das quais 299 são neutras e 381 destas são amostras com oclusão divididas em quatro grupos, denominados: *Glasses* (104), *Eye* (105), *Mouth* (105) e *Hair* (67). Para formar a base de dados da galeria foi utilizada a primeira amostra neutra de cada um dos 105 indivíduos, as outras 194 amostras neutras e as 381 com oclusão foram utilizadas para o teste.

Alexandre *et al.* (2020) e Soltanpour *et al.* (2017) relatam a dificuldade de se fazer comparações entre métodos de reconhecimento facial devido a diferentes cenários de experimentos e diferenças entre quantidade de amostras ou base de dados. Em geral, para o contexto de identificação, a maioria dos autores reportam taxas usando o experimento *rank-1* ou constroem a curva CMC como forma de comparar o desempenho dos algoritmos mostrando vários *ranks*. A configuração mais desafiadora consiste em realizar o experimento contendo somente uma amostra do indivíduo na galeria, desta forma, verifica-se a robustez do método e pressupõe o seu funcionamento no mundo real, visto que em aplicações reais, a base de dados pode conter mais de uma amostra do indivíduo e assim melhorar ainda mais a identificação.

Antes de realizar os experimentos é feita uma normalização nas nuvens, de acordo com a seção 3.1. A fase de pré-processamento demorou cerca de 0,33 segundos em média por amostra, já a etapa de correção de pose é a mais lenta de todo o procedimento, tomando cerca de 7,14 segundos em média para cada amostra.

Os códigos foram escritos na linguagem C++ usando a biblioteca PCL (RUSU; COUSINS, 2011) e a linguagem *Python* usando a biblioteca *sklearn* (PEDREGOSA *et al.*, 2011). Os *scripts* para chamadas de código C++ foram escritos em *Python* e no *Bash*. Para a visualização de alguns dados, a linguagem *R* foi utilizada. Os testes foram realizados em uma máquina com processador Intel i7-8550U, 1.8 GHz, 16GB memória.

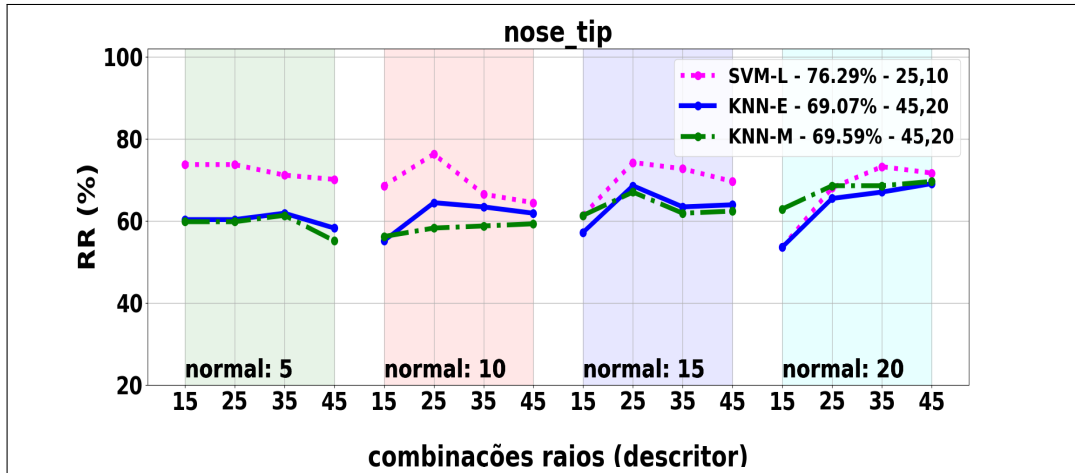
4.2 Resultados da classificação

Os resultados alcançados no processo de investigação são considerados e discutidos nessa parte da dissertação. Nas próximas seções são sumarizados os resultados oriundos da investigação produzida no Capítulo anterior em relação às regiões fiduciais, a malha de pontos criada e os detectores SIFT e ISS. Em todas as abordagens, utiliza-se FPFH como descritor de características.

4.2.1 Regiões Fiduciais

Neste experimento é feita uma análise da influência dos parâmetros de raio da normal calculado sobre o pontos e do raio de abrangência do FPFH, ou seja, o raio da esfera que delimita os pontos que serão considerados para calcular o descritor. O gráfico da Figura 25 mostra a taxa de reconhecimento, *Recognition Rate* (RR), para as combinações de raios definidas na seção 3.2.1. Foram usados os classificadores SVM Linear e o KNN e, neste último, duas formas de calcular a métrica de distância são mostradas, a distância de Manhattan e a Euclidiana. Na legenda da próxima Figura encontra-se a acurácia máxima para cada função colorida no gráfico e a correspondente combinação de raio.

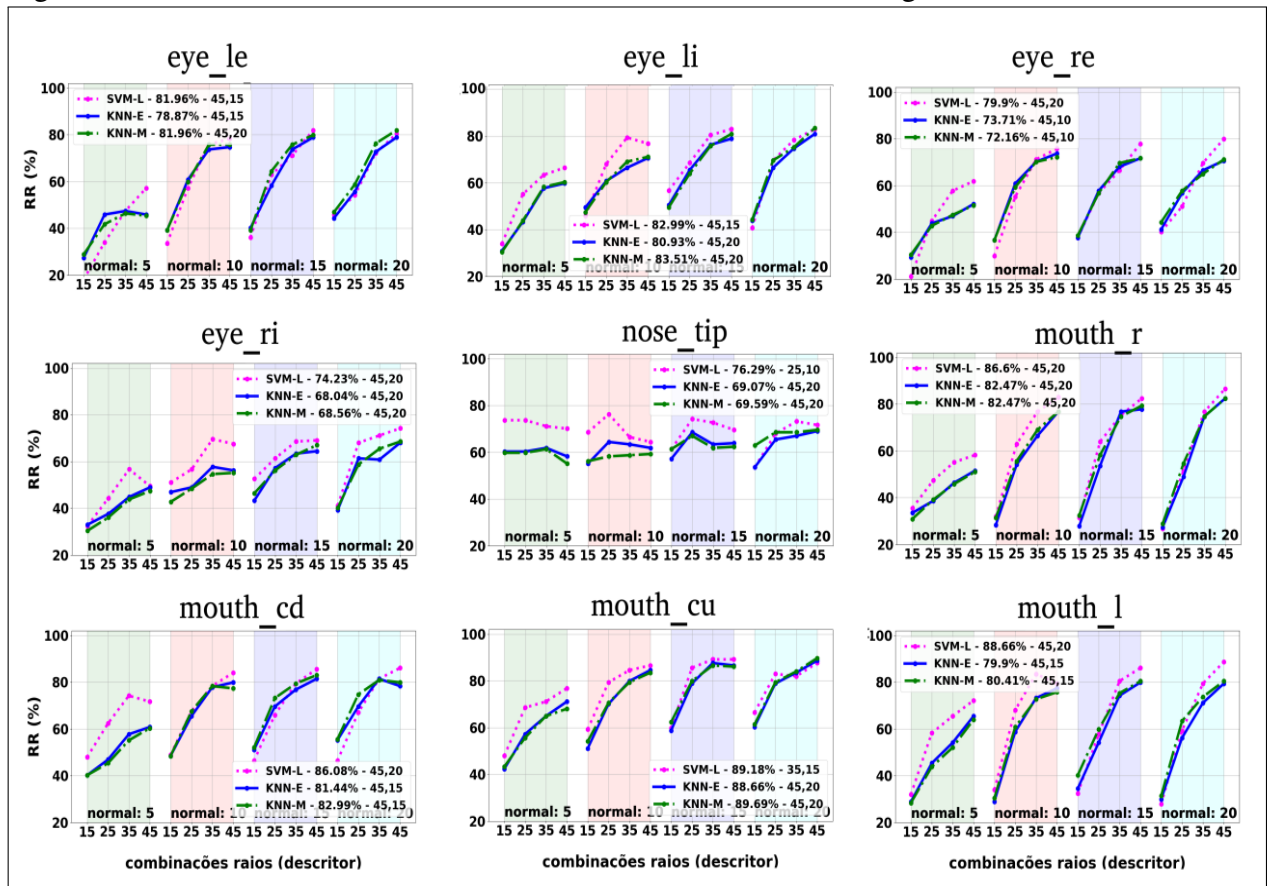
Figura 25 – Taxas de reconhecimento variando raios da normal e do descritor na região em torno do nariz (Neutral vs Neutral)



Fonte: Elaborado pelo autor.

A partir da análise anterior o mesmo ensaio foi estendido para outras regiões da face, definidas no capítulo anterior.

Figura 26 – Taxas de reconhecimento variando raios da normal nas regiões da face



Fonte: Elaborado pelo autor.

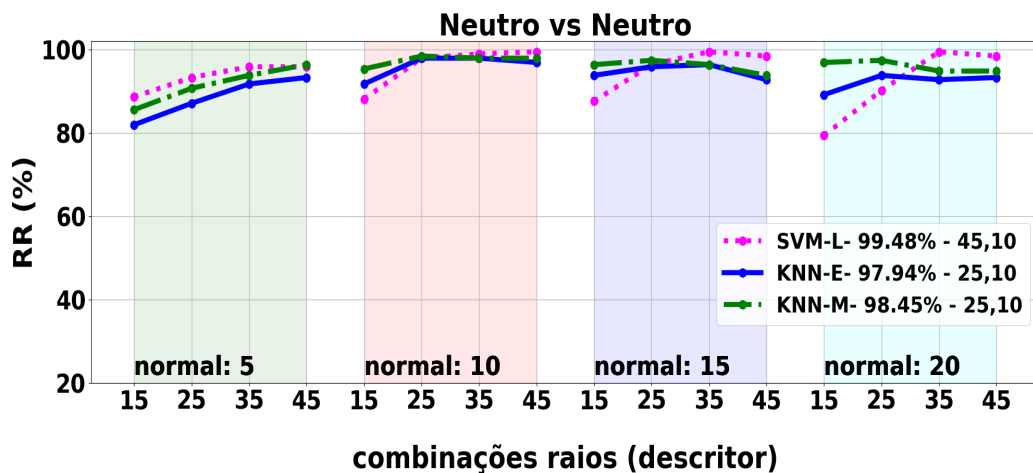
Observa-se pela Figura 26 que em algumas regiões, como em *mouth_cu*, foi possível

obter valores consideráveis de reconhecimento (89,18%) mesmo usando a região de forma isolada. Essa investigação é fundamental em um contexto de faces com regiões oclusas ou amostras de faces com partes faltantes, pois mostra a eficácia do descritor utilizando apenas parte da nuvem de pontos.

Ainda inserido nessa análise, o próximo passo contempla a fusão dessas regiões. Alguns autores (LEI *et al.*, 2013) também exploram a análise de regiões no contexto de reconhecimento facial. Li e Da (2012), por exemplo, buscam investigar as regiões com o intuito de dar robustez à expressão e oclusão por cabelo ao seu método, ou seja, realizar o reconhecimento facial mesmo sob estas condições.

A análise da união dos descritores foi feita de duas maneiras. Na primeira, as melhores combinações de raio encontradas na análise isolada foram consideradas para a fusão. Na segunda abordagem foi utilizada uma mesma combinação de raio para todos os descritores concatenados e verificado qual combinação produziu a maior taxa. As taxas referente à análise global podem ser vistas na Figura 27 e a Tabela 1 faz um comparativo dos resultados de acurácia obtidos com esses dois experimentos.

Figura 27 – Taxas de reconhecimento variando a combinação de raios no experimento de fusão dos descritores.



Fonte: Elaborado pelo autor.

Tabela 1 – Comparativo melhor raio global vs melhores raios locais.

Combinação de raio	RR (%)		
	SVM Linear	KNN-Euc	KNN-Man
Melhores Raios locais	98.45	93.3	94.33
Melhor Raio global	99.48	97.94	98.45

A Tabela 1 mostra que a escolha das melhores combinações de raio analisadas isoladamente não contribui para a maximização das taxas quando se tem o cenário de união dos descritores. Isso indica que a junção desses descritores locais alteram a natureza do vetor de características final utilizado no processo de identificação e, desta forma, modificam o máximo global das funções de classificação para esses parâmetros.

O próximo experimento foi feito em amostras da base que possuem oclusão. Algumas regiões são escolhidas manualmente, isso significa que ao invés de concatenar os nove descritores (nove regiões), somente alguns foram selecionados para formar o descritor final. Dessa forma, foi possível experimentar as amostras que possuem oclusões reais selecionando regiões que não estão na parte oclusa. Neste ensaio foi utilizada a melhor combinação de raio encontrada para cada classificador (legenda da Figura 27). Os experimentos mostrados na Tabela 2 são realizados seguindo a metodologia de indexamento das correspondências, descrita na 3.3.1. Na Tabela são exibidos alguns detalhes, como a quantidade de amostras utilizadas e as regiões que foram selecionadas.

Tabela 2 – Taxas de reconhecimento fazendo a seleção manual das regiões.

Dados de Teste	Nº de amostras	Regiões	RR (%)		
			SVM Linear	KNN-Euc	KNN-Man
NEUTRAL	194	All	99.48	97.94	98.45
MOUTH	105	eye_re; eye_ri eye_le; eye_li nose_tip	46.67	52.38	60.95
EYE	105	mouth_cd; mouth_cu mouth_r; nose_tip eye_re; eye_ri	71.43	79.05	82.86
GLASSES	104	mouth_cd; mouth_cu mouth_r; mouth_l nose_tip	89.42	81.73	81.73
HAIR	67	mouth_cd; mouth_cu mouth_r; mouth_l eye_li; eye_ri nose_tip	40.3	40.3	56.72

Nota: Para o "EYE" os pontos na região do olho direito foram mantidos em virtude das oclusões de olho nessas amostras serem somente na região esquerda.

É possível observar uma queda considerável nas taxas dos testes das faces com oclusão de boca e cabelo. Isso pode ser um indicativo de que o descritor tem uma capacidade de discriminância maior sobre as regiões inferiores nas faces da Bosphorus, pois nas oclusões do tipo *eye* e *glasses* todos os pontos relativos a parte superior foram selecionados. Além disso, a

Figura 26 mostra que as regiões relativas ao *mouth*, mesmo analisadas isoladamente, têm taxas maiores em relação às outras regiões.

4.2.2 Pontos Radiais

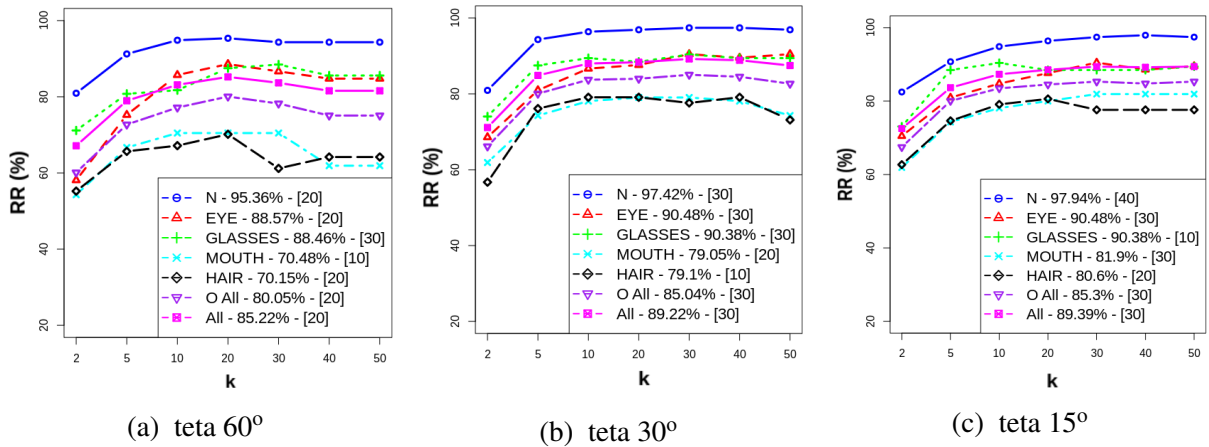
A partir dos próximos passos, os experimentos são realizados usando a seleção automática (explorada na seção 3.3.2) dos pontos com o intuito de evitar locais pouco discriminantes ou regiões em que não há correspondência de pontos, como nas regiões ocluídas na face. Como forma de diminuir a dependência dos pontos fiduciais marcados na base, é utilizada a estrutura de pontos construída de forma radial, a partir da extremidade do nariz. Nos experimentos são utilizadas três configurações diferentes. Na primeira, têm-se sete níveis começando pelo nariz, um espaçamento radial de 11mm e um espaçamento angular de 60° . As outras configurações diferem somente no ângulo, de 30° e 15° .

Como discorrido no Capítulo anterior, nessa etapa de identificação, para cada teste, são selecionados os k melhores pares de regiões correspondentes entre a face de prova e a face da galeria. Para cada par de descritores, é calculada a distância de *Manhattan*, após isso a média dessas distâncias é utilizada como métrica de dissimilaridade entre as duas faces. A distância de *Manhattan* foi preferida visto que produziu melhores resultados do que a distância euclidiana na seção anterior e é recomendada para análises de dados com alta dimensão (AGGARWAL *et al.*, 2001). Além disso, tem se mostrado mais relevante no contexto de classificação de faces como mostrado por Malkauthekar (2013) e Greche *et al.* (2017).

Com base nos resultados do experimento da Figura 27, nos próximos experimentos serão utilizados os valores de raio 25mm e 10mm para cálculo do descritor e da normal respectivamente. Isto porque esta foi a melhor combinação utilizando a técnica do KNN a qual é a mais semelhante ao método dinâmico utilizado. Na Figura 28 é possível verificar as taxas de reconhecimento em função de k para as configurações escolhidas na construção da malha radial, representadas na Figura 29. Estas taxas estão alocadas por cenário de teste, N representa o teste Neutro-Neutro, *EYE*, *GLASSES*, *MOUTH* e *HAIR* são os testes por tipo de oclusão (mostrado na seção 2.1.1); *O All* é o cenário em que todas as amostras com oclusão são submetidas juntas ao teste e *All* representa a realização do experimento com todas as oclusões (381) mais as neutras (194).

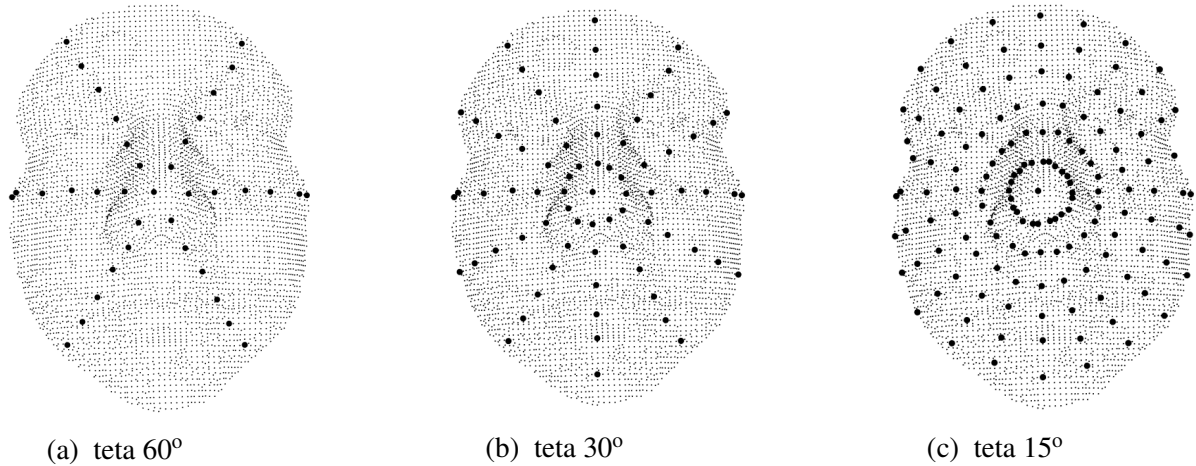
Pela forma como foi gerada a malha radial de pontos, percebe-se uma aglomeração circular de pontos para a malha de 15° (Figura 29c). Pontos muito próximos podem promover

Figura 28 – Resultados dos experimentos em função de k usando os pontos radiais.



Fonte: Elaborado pelo autor.

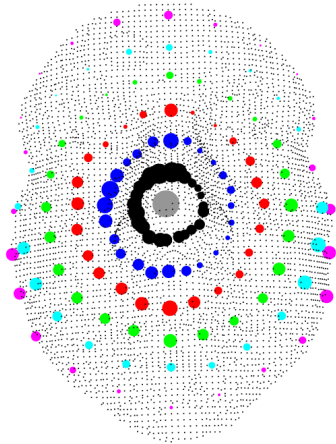
Figura 29 – Representação dos pontos radiais gerados em um indivíduo da base.



Fonte: Elaborado pelo autor.

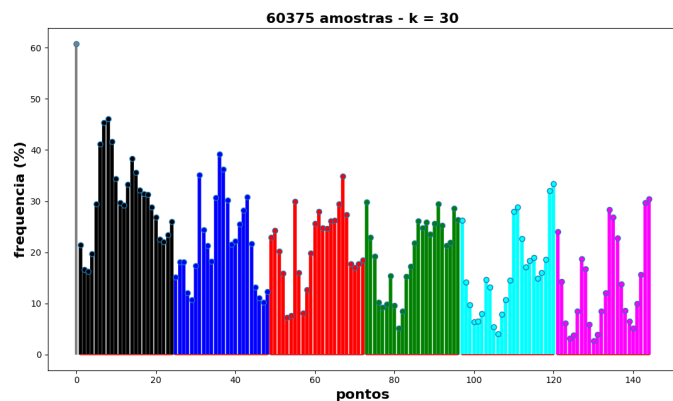
ambiguidade no cálculo dos descritores, fazendo com que o descritor seja menos discriminante e menos representativo da face, pois têm pouca variação da normal, principal atributo utilizado para o cálculo do descritor utilizado.

Figura 30 – Histograma das escolhas dos pontos na etapa de correspondência e seleção dos k descritores.



(a) Representação das escolhas com malha com teta 15° .

Fonte: Elaborado pelo autor.

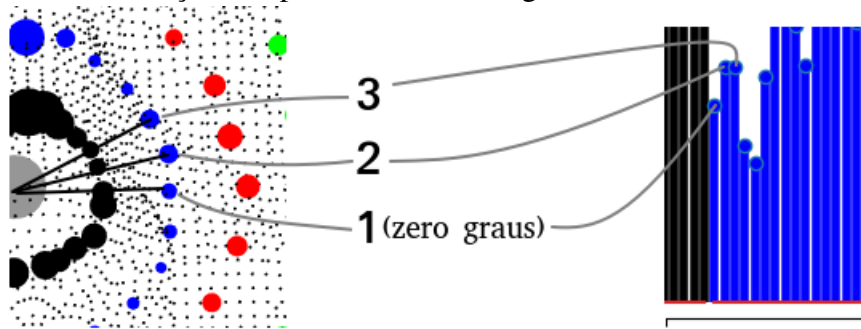


(b) Histograma da seleção dos pontos na etapa de correspondência.

Com o intuito de verificar quais pontos são mais selecionados no decorrer do *pipeline*, foi construído um histograma representando a frequência da seleção dos pontos. A Figura 30b representa os pontos mais escolhidos na etapa de seleção das k correspondências indexadas e a Figura ao lado esquerdo é uma representação visual normalizada da frequência das escolhas de cada ponto.

Cada ponto colorido tem o seu correspondente no histograma e o tamanho do ponto está normalizado de acordo com a intensidade no gráfico. A associação dos pontos é feita da seguinte maneira: em cada nível circular de pontos, o índice do ponto dentro do nível é definido pela sua posição no arco trigonométrico, onde o ponto inicial corresponde ao que está no ângulo zero e a ordem de crescimento é no sentido anti-horário. No histograma esse crescimento é da esquerda para a direita. Esse histograma foi gerado utilizando a malha construída na Figura 29c. A Figura oferece mais informações visuais para a interpretação dessa associação, tomando como exemplo o nível cujo setor está colorido de azul.

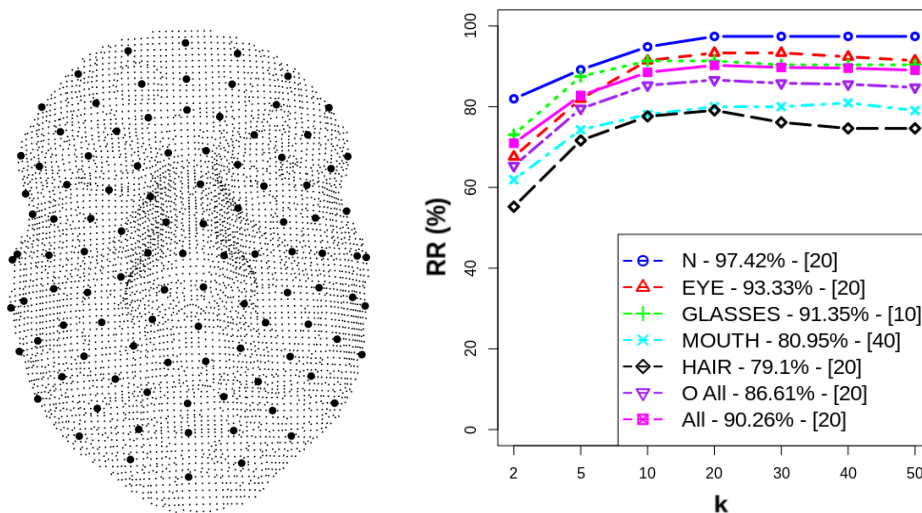
Figura 31 – Associação dos pontos com o histograma.



Fonte: Elaborado pelo autor.

É possível notar que existe uma concentração da frequência de escolhas na região em torno do nariz, principal região da aglomeração circular. A fim de diminuir esse efeito, a malha foi gerada com espaçamento angular diferenciado em alguns níveis, com o intuito de deixar os pontos no entorno do nariz mais esparsos. Utilizando a mesma malha gerada com $\theta = 15^\circ$, foram modificados os espaçamentos angulares para 60° , 40° , 20° nos níveis 2, 3, 4 respectivamente e mantido o mesmo valor de 15° para os outros níveis mais afastados do nariz (5,6,7). Na Figura 32a é possível verificar a nova distribuição da malha e os resultados produzidos, a partir dessa nova configuração. Nessa mesma Figura, os resultados mostram uma flutuação para cima nas taxas de reconhecimento para as amostras com oclusão do tipo *EYE* e *GLASSES*. Também é notável que os resultados alcançados no experimento realizado com neutros e oclusões juntos (*All*) passa de 90% de acerto.

Figura 32 – Resultados dos experimentos em função de k usando malha radial com pontos esparsos na região nasal.



(a) Malha de pontos esparsa.

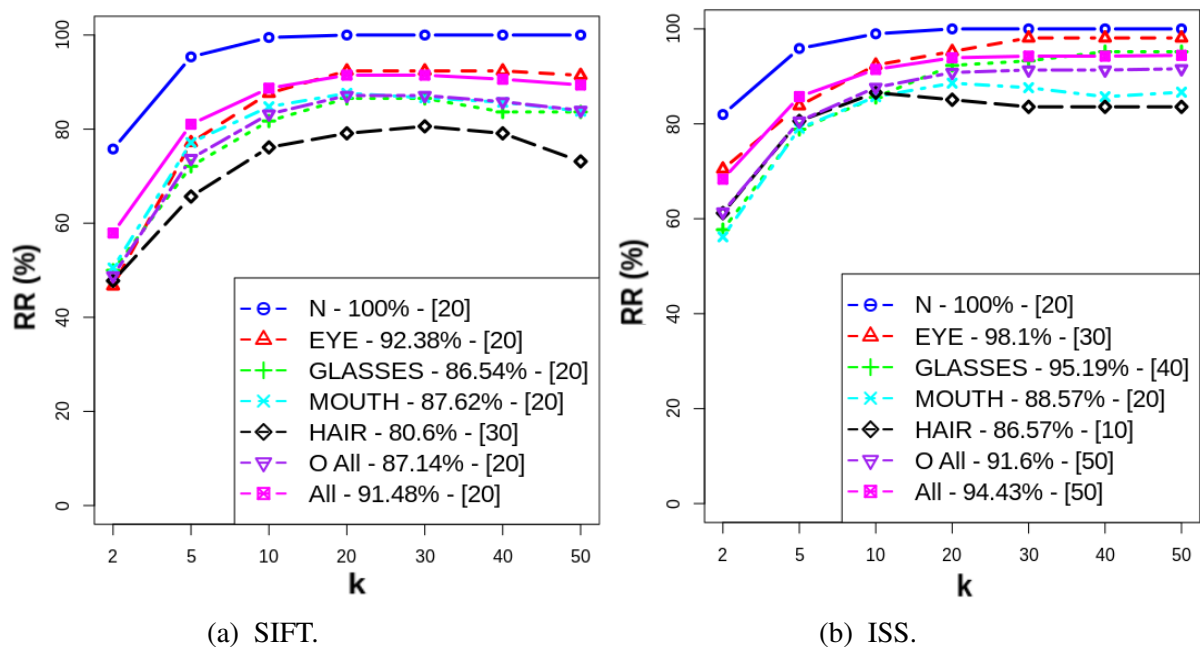
(b) RR em função das k melhores correspondências.

Fonte: Elaborado pelo autor.

4.2.3 Detectores SIFT e ISS

Nessa seção os resultados usando os algoritmos de detecção são explorados. Vale a pena lembrar que, para o uso dessas duas técnicas, foi necessário adicionar uma etapa de busca por pares de descritores correspondentes antes da etapa de seleção dos k melhores (mais próximos). Nos resultados mostrados anteriormente isso não era necessário, pois os pontos já estavam indexados a partir do nariz. A variação das taxas em função de k nas duas técnicas pode ser vista nas Figuras a seguir.

Figura 33 – Resultados dos experimentos em função de k no *rank-1*.



Fonte: Elaborado pelo autor.

Na Tabela a seguir são resumidas as estimativas de tempo para cada etapa usando a malha radial (Figura 32a) e os detectores. É notória a diferença de tempo na etapa de localização dos pontos usando a estratégia da malha radial, quase 46 vezes mais rápido. Isso é devido a estrutura da malha ser bastante simples de gerar e não haver análises refinadas na sua construção. A etapa de classificação é mais lenta (≈ 4 vezes mais) quando as técnicas de detecção são utilizadas, pois nesses casos o processamento de busca das correspondências entre os descritores é adicionado ao *pipeline*, como visto anteriormente. De fato, o tempo de processamento para a localização dos pontos usando a malha possui uma dependência do tempo para localizar a ponta do nariz. Apesar do método de detecção da extremidade do nariz não ter sido implementado, no trabalho de Xu *et al.* (2006) o tempo para a localização foi de 0,4 segundos em uma base com

mais de 70k pontos em cada face (sete vezes maior que a média de pontos das faces utilizadas). Apesar de este tempo superar o tempo de localização das outras técnicas (SIFT e ISS), o tempo global ainda é menor se somado os tempos de localização, cálculo dos descritores e classificação da Tabela 3. Além disso, Breitenstein *et al.* (2008) detecta a extremidade do nariz em imagens de profundidade utilizando a GPU e consegue resultados ainda melhores, 18 ms utilizando imagens com cerca de 150 x 200 valores de profundidade.

Tabela 3 – Comparativo de tempo dos dois métodos utilizados no trabalho.

Método	Nº médio de keypoints	Tempo (ms)		
		Localização dos pontos	Cálculo dos descritores	Classificação
SIFT	≈ 97	160.2	896.5	606.1
ISS	≈ 106	155.7	873.2	632.3
Malha Radial	106	3.5	872.8	162.3

Na Tabela 4 são mostrados os resultados obtidos usando os dois detectores e a malha radial (da Figura 32a) em comparação com algumas técnicas do estado da arte, todas aplicadas a Bosphorus obedecendo o critério *rank-1*.

Tabela 4 – Comparativo de experimentos *rank-1* na base Bosphorus.

	Taxa de reconhecimento (%)						Remoção e/ou Restauração
	Neutro 194	Glasses 104	Eye 105	Mouth 105	Hair 67	Todas Oclusões (381)	
Malha Radial	97.42	91.35	93.33	80.95	79.1	86.61	Não
SIFT	100	86.54	92.38	87.62	80.6	87.14	Não
ISS	100	95.19	98.1	88.57	86.57	91.6	Não
Dagnes <i>et al.</i> (2019)	97.20 ¹	92.55 ¹	96.74 ¹	96.77 ¹	n/a	95.35 ¹	Sim
Deng <i>et al.</i> (2020)	100	96.2	89.5	88.6	88.7 ³	97.9 ³	Não
Gilani <i>et al.</i> (2018)	n/a	100	99.0	96.1	97.3	98.1	Não
Li <i>et al.</i> (2015)	100	100	100	100	95.52	99.21	Não
Al-Osaimi (2016)	98.45	n/a	n/a	n/a	n/a	84.78	Não
Colombo <i>et al.</i> (2011)	n/a	94.23 ²	91.18 ²	74.75 ²	90.47 ²	n/a	Sim
Guo <i>et al.</i> (2016)	n/a	99.04	96.19	96.19	95.52	96.85	Não
Bellil <i>et al.</i> (2016)	n/a	97.12	89.46	92.67	93.91	n/a	Sim
Drira <i>et al.</i> (2013)	n/a	94.2	97.1	78.00	81.00	87.00	Sim
Lei <i>et al.</i> (2016)	n/a	96.20	90.50	94.30	88.10	92.70	Não

¹ O autor utilizou uma mistura da base Bosphorus com sua base privada no experimento neutro e uma quantidade reduzida de amostras na base de teste com oclusão.

² O Autor utilizou 360 amostras das 381 disponíveis, pois 21 destas não foram detectadas no detector de faces apresentado por ele.

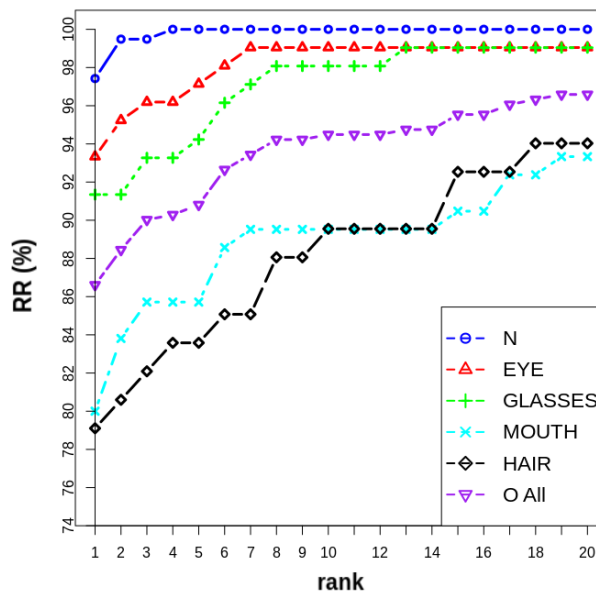
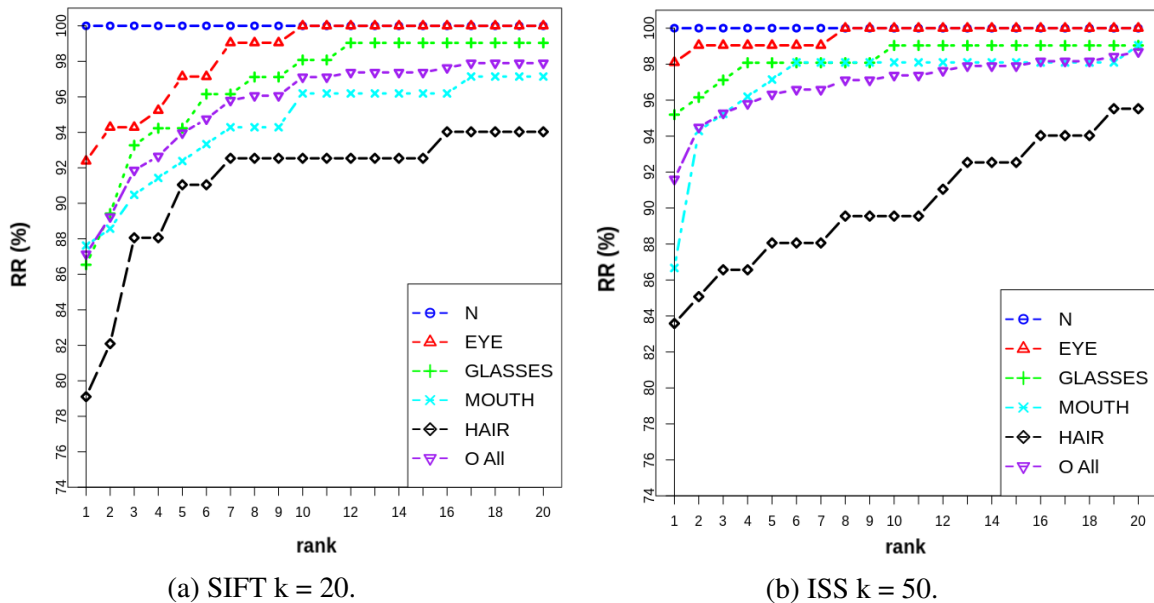
³ Essas taxas apresentam um erro no número de amostras de oclusão do tipo *Hair*. O Autor reportou 97 ao invés de 67. Logo, o cálculo da taxa percentual para a coluna com todas as oclusões está mal colocado.

Assim como a maioria dos trabalhos citados na Tabela anterior, é comum os autores usarem a curva CMC como forma de comparar os algoritmos no contexto de identificação. Russ *et al.* (2006) também usa essa curva para avaliar o seu método e resume que esta mostra a probabilidade de identificação se os primeiros k resultados de identificação forem considerados. Isso quer dizer que no *rank-3*, por exemplo, é contabilizado para o resultado da probabilidade final os casos em que a correta correspondência entre as faces aconteceria na primeira, segunda ou terceira tentativa. Como consequência, o *rank-N*, quando N é a quantidade de classes, é obrigatoriamente 100%.

Alguns métodos, dependendo da sua eficácia em algum *rank*, podem servir como um processo semi-automático, de maneira que a busca retornaria uma lista de k_{rank} prováveis

candidatos para uma decisão final do usuário ou talvez como entrada para uma técnica de pós processamento mais refinada. A Figura 34 demonstra as curvas CMC para cada uma das três estratégias de seleção dos pontos. O valor de k utilizado para cada técnica foi retirado das Figuras 32b, 33a e 33b no experimento *All*, em que todas as amostras são submetidas ao teste, nesse caso são 575 (194 neutras + 381 oclusões).

Figura 34 – Curvas CMC



Fonte: Elaborado pelo autor.

É possível notar que a curva da oclusão do tipo *HAIR* do método SIFT cresce de maneira mais acentuada do que nas outras técnicas, entretanto, a curva da oclusão do tipo

MOUTH no ISS é a que tem essa característica mais notável, subindo de 86.6% para 94.1%, do rank-1 para o rank-2. Apesar de o desempenho geral ter sido inferior utilizando a técnica de pontos radiais, o método não faz busca por pontos relevantes ou processamentos complexos para localizar os pontos. Isso confere à técnica um desempenho de tempo superior às demais (Tabela 3) e, além disso, produz taxas entre 8% e 3% inferiores em relação as outras duas técnicas. É importante destacar que, no caso da oclusão do tipo *GLASSES*, o desempenho usando a grade radial foi aproximadamente 5% superior em relação ao SIFT. No caso da oclusão do tipo *EYE*, superou as taxas reportadas no recente trabalho de Deng *et al.* (2020) e nos trabalhos de Colombo *et al.* (2011), Bellil *et al.* (2016) e Lei *et al.* (2016).

O desempenho utilizando o *ISS* como detetor foi relevante, visto que para amostra com oclusão do tipo *EYE* chegou perto de 100% ficando atrás apenas de três trabalhos no estado da arte (Tabela 4). A técnica apresentada por Li *et al.* (2015), apesar de ser a melhor em acurácia, utiliza uma média de 300 a 350 pontos relevantes e leva mais de 1 minuto para realizar as etapas de localização dos pontos, cálculo dos descritores e classificação. Já os métodos apresentados na dissertação utilizam aproximadamente uma quantidade três vezes menor de pontos relevantes e levam menos de dois segundos para a realização das três etapas.

Também é possível notar que o desempenho geral para oclusões na região dos olhos foi superior às oclusões na região inferior nas três técnicas mostradas. Mesmo assim, em todos os experimentos, o descritor FPFH tem se mostrado eficaz no contexto de reconhecimento facial, contexto no qual esse descritor não é muito conhecido ou as taxas de reconhecimento reportadas na literatura são baixas (AL-OSAIMI, 2016; ATIK *et al.*, 2018).

5 CONCLUSÕES E TRABALHOS FUTUROS

Foi apresentado, ao longo do trabalho, uma metodologia para realizar a identificação de faces 3D com uma certa tolerância a amostras com oclusão. A base Bosphorus foi utilizada principalmente pela disponibilidade de amostras com diferentes tipos de oclusão, a diversificação de indivíduos e sua ampla aceitação na literatura, permitindo uma comparação de resultados mais abrangente. A técnica de registro utilizada antes do processo de localização contribuiu para a correção do posicionamento das faces oclusas, pois o método foi adaptado para permitir o registro mesmo em meio as distorções causadas pelas oclusões.

A técnica de descrição de características locais (FPFH), embora seja mais conhecida no contexto de registro de nuvens de pontos e reconhecimento de objetos, mostrou-se eficaz no contexto de reconhecimento facial, mesmo no caso em que os dados estão submetidos a oclusões. Flexibilizar o processo de classificação por meio da seleção dinâmica dos descritores contribuiu para o reconhecimento das amostras com oclusão, pois o método obteve taxas comparáveis ao estado da arte. A métrica de distância de Manhattan apresentou melhores resultados em relação à euclidiana, confirmando alguns estudos na literatura.

No que concerne a localização das regiões de interesse para a extração de atributos, é importante ressaltar o desempenho utilizando a técnica ISS, que para as oclusões do tipo *EYE*, *GLASSES* e *MOUTH* obteve aproximadamente taxas de 98%, 95% e 87% de reconhecimento no rank-1 e 99%, 96% e 94% no rank-2, respectivamente. A técnica de localização utilizando pontos radiais, apesar de no geral ter desempenho menor, obteve valores acima de 91% para as oclusões do tipo *EYE* e *GLASSES* sem realizar processamentos complexos e num tempo menor tanto na etapa de localização como na etapa de classificação. Os experimentos com amostras neutras alcançaram taxas máximas de reconhecimento utilizando as técnicas SIFT e ISS. Com a malha radial a maior taxa obtida foi de 97.4%.

Há perspectivas de melhoria em quase todas as etapas do trabalho, entretanto são elencadas a seguir as mais pertinentes.

1. Malha radial

- a) Filtragem: Investigar formas de filtrar os pontos na malha radial que sejam mais relevantes para extração de características. Uma análise de variação da normal ou de atributos relacionados com a distribuição ou a forma da região pode gerar taxas melhores.

- b) Dependência: Atualmente, a elaboração da malha radial tem uma dependência forte da localização da extremidade do nariz e do plano de alinhamento da face, pois a malha é gerada sobre esse plano a partir da ponta do nariz. Uma perspectiva melhoraria seria a verificação da possibilidade de construção dessa malha independente do plano de orientação ou até mesmo da localização do nariz, estudando formas de particionamento espacial das nuvens sem a informação de orientação, como o *octree* ou até mesmo o CP-ICP.
2. **Registro:** Verificar formas de aumentar a velocidade do registro ou propor novas técnicas de registro tolerantes a oclusão.
 3. **Ponto relevantes:** Investigação mais apurada dos parâmetros do SIFT e ISS. A investigação do primeiro item também poderia ser aplicada aqui a fim de encontrar os melhores parâmetros.
 4. **Descritor:** Verificar o desempenho do FPFH com as nuvens sem nenhum alinhamento e, no caso de uma piora nos resultados, propor soluções para minimizar esse efeito.
 5. **Classificação** Atualmente uma única métrica é computada na etapa de identificação (média das distâncias dos melhores descritores correspondentes). Elaborar outras métricas ou realizar fusão de métricas poderia melhorar os resultados, como no trabalho de Li *et al.* (2015) e Guo *et al.* (2016).
 6. **Generalização:** Por fim, testar o *pipeline* nas amostras com expressões e testar em outras bases tão complexas e amplas quanto a Bosphorus. É importante também, testar o método no contexto de verificação e fazer análises comparativas com os resultados da literatura.

REFERÊNCIAS

- AGGARWAL, C. C.; HINNEBURG, A.; KEIM, D. A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In: BUSSCHE, J. den; VIANU, V. (Ed.). **Database Theory — ICDT 2001**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 420–434. ISBN 978-3-540-44503-6.
- AHDID, R.; TAIFI, K.; SAID, S.; FAKIR, M.; MANAUT, B. Automatic Face Recognition System Using Iso-Geodesic Curves in Riemannian Manifold. In: **2017 14th International Conference on Computer Graphics, Imaging and Visualization**. [S.l.: s.n.], 2017. p. 73–78.
- AL-OSAIMI, F. R. A Novel Multi-Purpose Matching Representation of Local 3D Surfaces: A Rotationally Invariant, Efficient, and Highly Discriminative Approach With an Adjustable Sensitivity. **IEEE Transactions on Image Processing**, v. 25, n. 2, p. 658–672, 2 2016. ISSN 1057-7149. Disponível em: <<http://ieeexplore.ieee.org/document/7300438/>>.
- ALEXA, M.; BEHR, J.; COHEN-OR, D.; FLEISHMAN, S.; LEVIN, D.; SILVA, C. Computing and rendering point set surfaces. **IEEE Transactions on Visualization and Computer Graphics**, v. 9, n. 1, p. 3–15, 1 2003. ISSN 1077-2626. Disponível em: <<http://ieeexplore.ieee.org/document/1175093/>>.
- ALEXANDRE, G. R.; SOARES, J. M.; THÉ], G. A. P. Systematic review of 3D facial expression recognition methods. **Pattern Recognition**, v. 100, p. 107108, 2020. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320319304091>>.
- ALYUZ, N.; GOKBERK, B.; AKARUN, L. Adaptive Registration for Occlusion Robust 3D Face Recognition. In: FUSIELLO, A.; MURINO, V.; CUCCHIARA, R. (Ed.). **COMPUTER VISION - ECCV 2012, PT III**. [S.l.], 2012. (Lecture Notes in Computer Science, v. 7585), p. 557–566. ISBN 978-3-642-33885-4; 978-3-642-33884-7. ISSN 0302-9743.
- ALYUZ, N.; GOKBERK, B.; SPREEUWERS, L.; VELDHUIS, R.; AKARUN, L. Robust 3D face recognition in the presence of realistic occlusions. In: **2012 5th IAPR International Conference on Biometrics (ICB)**. IEEE, 2012. p. 111–118. ISBN 978-1-4673-0397-2. ISSN 2376-4201. Disponível em: <<http://ieeexplore.ieee.org/document/6199767/>>.
- ATIK, M. E.; DURAN, Z.; BAYRAM, B. Comparison and usage of local feature based algorithms for 3D face recognition. In: **Proceedings - 39th Asian Conference on Remote Sensing: Remote Sensing Enabling Prosperity, ACRS 2018**. [s.n.], 2018. v. 4, p. 2387–2392. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071846532&partnerID=40&md5=9824503d99b8566e5acd6027f6e8dcf2>>.
- BAGCHI, P.; NASIPURI, M.; BHATTACHARJEE, D.; BASU, D. K. A novel approach for nose tip detection using smoothing by weighted median filtering applied to 3D face images in variant poses. In: **International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)**. IEEE, 2012. p. 272–277. ISBN 978-1-4673-1039-0. Disponível em: <<http://ieeexplore.ieee.org/document/6208357/>>.
- BELLEKENS, B.; SPRUYT, V.; WEYN, R. B. M. A survey of rigid 3D pointcloud registration algorithms. In: **Fourth International Conference on Ambient Computing, Applications, Services and Technologies, Proceedings**. IARA, 2014. p. 8–13. ISBN 9781612083568. ISSN 2326-9324. Disponível em: <http://www.thinkmind.org/download.php?articleid=ambient_2014_1_20_40015>.

BELLIL, W.; BRAHIM, H.; AMAR, C. B. Gappy wavelet neural network for 3D occluded faces: detection and recognition. **Multimedia Tools and Applications**, v. 75, n. 1, p. 365–380, 1 2016. ISSN 1380-7501. Disponível em: <<http://link.springer.com/10.1007/s11042-014-2294-6>>.

BENAMOUN, M.; GUO, Y.; SOHEL, F. Feature Selection for 2D and 3D Face Recognition. In: **Wiley Encyclopedia of Electrical and Electronics Engineering**. American Cancer Society, 2015. p. 1–28. ISBN 9780471346081. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/047134608X.W8257>>.

BESL, P. J.; MCKAY, N. D. Method for registration of 3-D shapes. In: SCHENKER, P. S. (Ed.). **Sensor Fusion IV: Control Paradigms and Data Structures**. SPIE, 1992. v. 1611, p. 586–606. Disponível em: <<https://doi.org/10.1117/12.57955>>.

BREITENSTEIN, M. D.; KUETTEL, D.; WEISE, T.; GOOL, L. van; PFISTER, H. Real-time face pose estimation from single range images. In: **2008 IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2008. p. 1–8. ISSN 1063-6919.

CHEN, Y.; MEDIONI, G. Object modeling by registration of multiple range images. In: **Proceedings. 1991 IEEE International Conference on Robotics and Automation**. IEEE Comput. Soc. Press, 1991. p. 2724–2729. ISBN 0-8186-2163-X. Disponível em: <<http://ieeexplore.ieee.org/document/132043/>>.

CHIEM, Q. T.; WILKINSON, R. H.; LECH, M.; CHENG, E. Investigating Keypoint Repeatability for 3D Correspondence Estimation in Cluttered Scenes. In: **2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)**. [S.l.: s.n.], 2017. p. 1–7.

COLOMBO, A.; CUSANO, C.; SCHETTINI, R. Three-Dimensional Occlusion Detection and Restoration of Partially Occluded Faces. **Journal of Mathematical Imaging and Vision**, Springer, v. 40, n. 1, p. 105–119, 5 2011. ISSN 0924-9907. Disponível em: <<http://link.springer.com/10.1007/s10851-010-0252-0>>.

DAGNES, N.; MARCOLIN, F.; NONIS, F.; TORNINCASA, S.; VEZZETTI, E. 3D geometry-based face recognition in presence of eye and mouth occlusions. **International Journal on Interactive Design and Manufacturing (IJIDeM)**, v. 13, n. 4, p. 1617–1635, 12 2019. ISSN 1955-2513. Disponível em: <<http://link.springer.com/10.1007/s12008-019-00582-7>>.

DAROM, T.; KELLER, Y. Scale-Invariant Features for 3-D Mesh Models. **IEEE Transactions on Image Processing**, v. 21, n. 5, p. 2758–2769, 2012. ISSN 1941-0042.

DENG, H.; BIRDAL, T.; ILIC, S. 3D Local Features for Direct Pairwise Registration. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2019. p. 3239–3248. ISSN 2575-7075.

DENG, X.; DA, F.; SHAO, H.; JIANG, Y. A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors. **Computers & Electrical Engineering**, v. 85, p. 106700, 2020. ISSN 0045-7906. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0045790620305553>>.

DRIRA, H.; AMOR, B. B.; SRIVASTAVA, A.; DAOUDI, M. A Riemannian analysis of 3D nose shapes for partial human biometrics. In: **2009 IEEE 12th International Conference on Computer Vision**. IEEE, 2009. p. 2050–2057. ISBN 978-1-4244-4420-5. ISSN 2380-7504. Disponível em: <<http://ieeexplore.ieee.org/document/5459451/>>.

DRIRA, H.; AMOR, B. B.; SRIVASTAVA, A.; DAOUDI, M.; SLAMA, R. 3D Face recognition under expressions, occlusions, and pose variations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 9, p. 2270–2283, 9 2013. ISSN 01628828. Disponível em: <<http://ieeexplore.ieee.org/document/6468044/>>.

EKENEL, H. K.; STIEFELHAGEN, R. Why Is Facial Occlusion a Challenging Problem? In: TISTARELLI, M.; NIXON, M. S. (Ed.). **Advances in Biometrics**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 299–308. ISBN 978-3-642-01793-3.

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

GALBALLY, J.; MARCEL, S.; FIERREZ, J. Biometric Antispoofing Methods: A Survey in Face Recognition. **IEEE Access**, v. 2, p. 1530–1552, 2014. ISSN 2169-3536.

GILANI, S. Z.; MIAN, A. Towards Large-Scale 3D Face Recognition. In: **2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)**. IEEE, 2016. p. 1–8. ISBN 978-1-5090-2896-2. ISSN null. Disponível em: <<http://ieeexplore.ieee.org/document/7797090/>>.

GILANI, S. Z.; MIAN, A. Learning from millions of 3d scans for large-scale 3d face recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2018. p. 1896–1905.

GILANI, S. Z.; MIAN, A.; SHAFAIT, F.; REID, I. Dense 3D Face Correspondence. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 40, n. 7, p. 1584–1598, 2018. ISSN 1939-3539.

GRECHE, L.; JAZOULI, M.; ES-SBAI, N.; MAJDA, A.; ZARGHILI, A. Comparison between Euclidean and Manhattan distance measure for facial expressions classification. In: **2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)**. [S.l.: s.n.], 2017. p. 1–4.

GUO, Y.; LEI, Y.; LIU, L.; WANG, Y.; BENNAMOUN, M.; SOHEL, F. EI3D: Expression-invariant 3D face recognition based on feature and shape matching. **Pattern Recognition Letters**, v. 83, n. 3, p. 403–412, 11 2016. ISSN 01678655. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0167865516300551>>.

KAZMI, I. K.; YOU, L.; ZHANG, J. J. A Survey of 2D and 3D Shape Descriptors. In: **2013 10th International Conference Computer Graphics, Imaging and Visualization**. [S.l.: s.n.], 2013. p. 1–10.

KIM, J.; HAN, D.; HWANG, W.; KIM, J. 3D face recognition via discriminative keypoint selection. In: **2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)**. IEEE, 2017. p. 477–480. ISBN 978-1-5090-3056-9. ISSN null. Disponível em: <<http://ieeexplore.ieee.org/document/7992781/>>.

Kyong I. Chang; BOWYER, K.; FLYNN, P. Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 28, n. 10, p. 1695–1700, 10 2006. ISSN 0162-8828. Disponível em: <<http://ieeexplore.ieee.org/document/1677524/>>.

- LEE, Y.; YI, T. 3D face recognition using multiple features for local depth information. In: **Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No.03EX667)**. Faculty of Electrical Eng. & Comput, 2003. v. 1, p. 429–434. ISBN 953-184-054-7. Disponível em: <<http://ieeexplore.ieee.org/document/1220499/>>.
- LEHTOLA, V.; KAARTINEN, H.; NÜCHTER, A.; KAIJALUOTO, R.; KUKKO, A.; LITKEY, P.; HONKAVAARA, E.; ROSNELL, T.; VAAJA, M.; VIRTANEN, J.-P.; KURKELA, M.; ISSAOUI, A. E.; ZHU, L.; JAAKKOLA, A.; HYYPPÄ, J. Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 9, n. 8, p. 796, 8 2017. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/9/8/796>>.
- LEI, Y.; BENNAMOUN, M.; EL-SALLAM, A. A. An efficient 3D face recognition approach based on the fusion of novel local low-level features. **Pattern Recognition**, v. 46, n. 1, p. 24–37, 1 2013. ISSN 00313203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320312002932>><https://linkinghub.elsevier.com/retrieve/pii/S0031320312002932>>.
- LEI, Y.; GUO, Y.; HAYAT, M.; BENNAMOUN, M.; ZHOU, X. A Two-Phase Weighted Collaborative Representation for 3D partial face recognition with single sample. **Pattern Recognition**, v. 52, p. 218–237, 2016. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320315003660>>.
- LI, H.; HUANG, D.; MORVAN, J.-M.; WANG, Y.; CHEN, L. Towards 3D Face Recognition in the Real: A Registration-Free Approach Using Fine-Grained Matching of 3D Keypoint Descriptors. **International Journal of Computer Vision**, v. 113, n. 2, p. 128–142, 6 2015. ISSN 0920-5691. Disponível em: <<http://link.springer.com/10.1007/s11263-014-0785-6>>.
- LI, X.; DA, F. Efficient 3D face recognition handling facial expression and hair occlusion. **Image and Vision Computing**, v. 30, n. 9, p. 668–679, 9 2012. ISSN 02628856. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0262885612001163>>.
- LI, X.; RUAN, Q.; AN, G. 3D facial expression recognition using delta faces. In: **5th IET International Conference on Wireless, Mobile and Multimedia Networks (ICWMMN 2013)**. [S.l.: s.n.], 2013. p. 234–239.
- LOWE, D. G. Object recognition from local scale-invariant features. In: **Proceedings of the Seventh IEEE International Conference on Computer Vision**. [S.l.: s.n.], 1999. v. 2, p. 1150–1157.
- LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 11 2004. ISSN 0920-5691. Disponível em: <<https://doi.org/10.1023/B:VISI.0000029664.99615.94>><http://link.springer.com/10.1023/B:VISI.0000029664.99615.94>>.
- LU, J.; WANG, W.; SHAO, H.; SU, L. Point Cloud Registration Algorithm Fusing of Super 4PCS and ICP Based on the Key Points. In: **2019 Chinese Control Conference (CCC)**. [S.l.: s.n.], 2019. p. 4439–4444. ISSN 1934-1768.
- MALKAUTHEKAR, M. D. Analysis of euclidean distance and Manhattan Distance measure in face recognition. In: **Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)**. [S.l.: s.n.], 2013. p. 503–507.

MARR, D. Vision: A computational investigation into the human representation and processing of visual information, Henry Holt and Co. Inc., New York, NY, v. 2, n. 4.2, 1982.

MASUDA, T.; SAKAUE, K.; YOKOYA, N. Registration and integration of multiple range images for 3-D model construction. In: **Proceedings of 13th International Conference on Pattern Recognition**. [S.l.: s.n.], 1996. v. 1, p. 879–883. ISSN 1051-4651.

MOUSAVI, M. H.; FAEZ, K.; ASGHARI, A. Three Dimensional Face Recognition Using SVM Classifier. In: **Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008)**. [S.l.: s.n.], 2008. p. 208–213.

PASSALIS, G.; KAKADIARIS, I. A.; THEOHARIS, T.; TODERICI, G.; MURTUZA, N. Evaluation of 3D Face Recognition in the presence of facial expressions: an Annotated Deformable Model approach. In: **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops**. [S.l.: s.n.], 2005. p. 171. ISSN 2160-7516.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, n. 85, p. 2825–2830, 2011. Disponível em: <<http://jmlr.org/papers/v12/pedregosa11a.html>>.

PEREIRA, N. S.; CARVALHO, C. R.; THÉ, G. A. P. Point cloud partitioning approach for ICP improvement. In: **2015 21st International Conference on Automation and Computing (ICAC)**. [S.l.: s.n.], 2015. p. 1–5.

QUAN, S.; MA, J. Keypoint domain triangular features for fast initial alignment of 3D point clouds. **Electronics Letters**, v. 55, n. 14, p. 787–789, 2019. ISSN 0013-5194.

RUSINKIEWICZ, S.; LEVOY, M. Efficient variants of the ICP algorithm. In: **Proceedings Third International Conference on 3-D Digital Imaging and Modeling**. [S.l.: s.n.], 2001. p. 145–152.

RUSS, T.; BOEHNEN, C.; PETERS, T. 3D Face Recognition Using 3D Alignment for PCA. In: **2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)**. IEEE, 2006. v. 2, p. 1391–1398. ISBN 0-7695-2597-0. ISSN 1063-6919. Disponível em: <<http://ieeexplore.ieee.org/document/1640920/>>.

RUSU, R. B.; BLODOW, N.; BEETZ, M. Fast Point Feature Histograms (FPFH) for 3D registration. In: **2009 IEEE International Conference on Robotics and Automation**. [S.l.: s.n.], 2009. p. 3212–3217. ISSN 1050-4729.

RUSU, R. B.; BLODOW, N.; MARTON, Z. C.; BEETZ, M. Aligning point cloud views using persistent feature histograms. In: **2008 IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S.l.: s.n.], 2008. p. 3384–3391. ISSN 2153-0866.

RUSU, R. B.; COUSINS, S. 3D is here: Point Cloud Library (PCL). In: **2011 IEEE International Conference on Robotics and Automation**. [S.l.: s.n.], 2011. p. 1–4. ISSN 1050-4729.

RUSU, R. B.; MARTON, Z. C.; BLODOW, N.; BEETZ, M. Persistent point feature histograms for 3D point clouds. In: . [S.l.: s.n.], 2008.

- SAVRAN, A.; ALYÜZ, N.; DIBEKLIO\UGLU, H.; ÇELIKTUTAN, O.; GÖKBERK, B.; SANKUR, B.; AKARUN, L. Bosphorus Database for 3D Face Analysis. In: SCHOUTEN, B.; JUUL, N. C.; DRYGAJLO, A.; TISTARELLI, M. (Ed.). **Biometrics and Identity Management**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 47–56. ISBN 978-3-540-89991-4.
- SEGAL, A.; HAEHNEL, D.; THRUN, S. Generalized-ICP. In: **Proceedings of Robotics: Science and Systems**. Seattle, USA: [s.n.], 2009.
- SIQUEIRA, R. S.; ALEXANDRE, G. R.; SOARES, J. M.; THÉ, G. A. P. 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, 2018. ISSN 2377-3766.
- SOLTANPOUR, S.; BOUFAMA, B.; WU, Q. J. A survey of local feature methods for 3D face recognition. **Pattern Recognition**, v. 72, p. 391–406, 12 2017. ISSN 00313203. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0031320317303072>>.
- STEDER, B.; RUSU, R. B.; KONOLIGE, K.; BURGARD, W. Point feature extraction on 3D range scans taking into account object boundaries. In: **2011 IEEE International Conference on Robotics and Automation**. [S.l.: s.n.], 2011. p. 2601–2608. ISSN 1050-4729.
- TURK, G.; LEVOY, M. Zippered Polygon Meshes from Range Images. In: **Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques**. New York, NY, USA: Association for Computing Machinery, 1994. (SIGGRAPH '94), p. 311–318. ISBN 0897916670. Disponível em: <<https://doi.org/10.1145/192161.192241>>.
- WAHL, E.; HILLENBRAND, U.; HIRZINGER, G. Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification. In: **Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings**. IEEE, 2003. p. 474–481. ISBN 0-7695-1991-1. Disponível em: <<http://ieeexplore.ieee.org/document/1240284/>>.
- WEIK, S. Registration of 3-D partial surface models using luminance and depth information. In: **Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134)**. [S.l.: s.n.], 1997. p. 93–100.
- XU, C.; TAN, T.; WANG, Y.; QUAN, L. Combining local features for robust nose location in 3D facial data. **Pattern Recognition Letters**, v. 27, n. 13, p. 1487–1494, 2006. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865506000663>>.
- XU, K.; WANG, X.; HU, Z.; ZHANG, Z. 3d face recognition based on twin neural network combining deep map and texture. In: IEEE. **2019 IEEE 19th International Conference on Communication Technology (ICCT)**. [S.l.], 2019. p. 1665–1668.
- YANG, J.; LI, H.; JIA, Y. Go-ICP: Solving 3D Registration Efficiently and Globally Optimally. In: **2013 IEEE International Conference on Computer Vision**. [S.l.: s.n.], 2013. p. 1457–1464. ISSN 2380-7504.
- YEO, I.; JOHNSON, R. A. A new family of power transformations to improve normality or symmetry. **Biometrika**, v. 87, n. 4, p. 954–959, 2000. ISSN 0006-3444. Disponível em: <<https://doi.org/10.1093/biomet/87.4.954>>.

ZHANG, Z.; SUN, L.; ZHONG, R.; CHEN, D.; XU, Z.; WANG, C.; QIN, C.; SUN, H.; LI, R. 3-D Deep Feature Construction for Mobile Laser Scanning Point Cloud Registration. **IEEE Geoscience and Remote Sensing Letters**, v. 16, n. 12, p. 1904–1908, 2019. ISSN 1558-0571.

ZHONG, Y. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In: **2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops**. IEEE, 2009. p. 689–696. ISBN 978-1-4244-4442-7. Disponível em: <<http://ieeexplore.ieee.org/document/5457637/>>.

ZHOU, S.; XIAO, S. 3D face recognition: a survey. **Human-centric Computing and Information Sciences**, SpringerOpen, v. 8, n. 1, p. 35, 12 2018. ISSN 2192-1962. Disponível em: <<https://doaj.org/article/6a7963cf7cde43838564177fb504f03bhttps://hcis-journal.springeropen.com/articles/10.1186/s13673-018-0157-2>>.