



**UNIVERSIDADE FEDERAL DO CEARÁ  
CENTRO DE TECNOLOGIA  
DEPARTAMENTO DE TELEINFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA  
DOUTORADO EM TELEINFORMÁTICA**

**ROBSON DA SILVA SIQUEIRA**

**RECONHECIMENTO DE FACES 3D SEGMENTADAS EM REGIÕES TRIAXIAIS  
UTILIZANDO MOMENTOS ESPACIAIS ADAPTADOS E INVARIANTES À  
ROTAÇÃO**

**FORTALEZA  
2018**

ROBSON DA SILVA SIQUEIRA

RECONHECIMENTO DE FACES 3D SEGMENTADAS EM REGIÕES TRIAXIAIS  
UTILIZANDO MOMENTOS ESPACIAIS ADAPTADOS E INVARIANTES À ROTAÇÃO

Tese apresentada ao 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 doutor em Engenharia de Teleinformática. Área de Concentração: SINAIS E SISTEMAS

Orientador: Prof. Dr. José Marques Soares

Coorientador: Prof. Dr. George André Pereira Thé

FORTALEZA

2018

ROBSON DA SILVA SIQUEIRA

RECONHECIMENTO DE FACES 3D SEGMENTADAS EM REGIÕES TRIAXIAIS  
UTILIZANDO MOMENTOS ESPACIAIS ADAPTADOS E INVARIANTES À ROTAÇÃO

Tese apresentada ao 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 doutor em Engenharia de Teleinformática. Área de Concentração: SINAIS E SISTEMAS

Aprovada em:

BANCA EXAMINADORA

---

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

---

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

---

Prof. Dr. João Paulo Pordeus Gomes  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Ajalmor Rego da Rocha  
Instituto Federal de Educação, Ciência e  
Tecnologia do Ceará (IFCE)

---

Prof. Dr. Adrião Duarte Dória Neto  
Universidade Federal do Rio Grande do Norte  
(UFRN)

À Deus que com seu infinito amor, mostrou-me durante todo o período de estudos a necessidade de ter fé, paciência, humildade e resignação para que o resultado de todo esforço pudesse ter sua conclusão. À minha esposa Karine, aos meus filhos Sofia e Artur e aos meus pais.



## **AGRADECIMENTOS**

Ao Prof. Dr. José Marques Soares por acreditar no meu trabalho mais do que eu mesmo e sempre me lembrar do real sentido de um estudo de doutorado.

Ao Prof. Dr. George André Pereira Thé que nos momentos de afobiação sempre vinha com uma análise crítica que apontava caminhos a serem trilhados.

Ao Prof. Dr. Paulo Cesar Cortez que desde a graduação em Engenharia Elétrica quando foi meu tutor informal, e no mestrado meu orientador, nunca deixou de me ouvir, orientar e cuidar onde quer que estivesse.

À amiga Manoela Rodrigues que nos momentos mais escuros, quando desistir de tudo parecia a única saída, sempre me mostrou o caminho da luz.

À mestrandra Gilderlane Ribeiro pela grande parceria pela construção deste trabalho.

À equipe do grupo Interfaces, criado a partir dos vários temas que surgiram com este trabalho.

Ao Instituto Federal do Ceará, nas pessoas dos professores do Departamento de Computação do *Campus* Maracanaú que me proporcionaram a liberação necessária para a conclusão deste trabalho.

E à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento da pesquisa de doutorado.

'Não importa o quão estreito seja o portão e quão  
repleta de castigos seja a sentença, eu sou o dono  
do meu destino, eu sou o capitão da minha alma.'  
– *Invictus*'

(William Ernest Henley)

## **RESUMO**

Este trabalho apresenta um modelo de fatiamento múltiplo para imagens 3D da face humana, utilizando os planos ortogonal Frontal, Sagital e Transversal. A definição dos segmentos depende apenas de um ponto-chave, a ponta do nariz, o que torna simples e independente da detecção de vários pontos-chave. Para o reconhecimento facial, atributos baseados em momentos espaciais de Hu, momentos 2D adaptados, e momentos 3D invariantes à rotação espacial são extraídos de cada segmento. Testes com o modelo proposto usando a base de dados Bosphorus para experimento ROC-I Neutro *vs* Não-Neutro, aplicando Análise Linear Discriminante como classificador e mais de uma amostra para treinamento, alcançaram 98,7% de taxa de verificação com 0,1% de falso positivo. Usando o Support Vector Machine como classificador, as taxas de reconhecimento de experimento rank-1 de 99% e 95,4% foram alcançadas para os eventos Neutro *vs* Neutro e para um Neutro *vs* Não-Neutro, respectivamente. Esses resultados abordam o estado da arte usando a base de dados do Bosphorus, superarando quando as expressões Raiva e Repugnância são avaliadas. Além disso, também foi avaliado a generalização do método proposto, utilizando a base de dados FRGC v2.0 e resultados competitivos foram obtidos, tornando a técnica promissora, principalmente pela sua simplicidade.

## ABSTRACT

This work presents a multiple slicing model for 3D images of human face, using the Frontal, Sagittal and Transverse orthogonal planes. The definition of the segments depends on just one key point, the nose tip, which makes it simple and independent of the detection of several key points. For facial recognition, attributes based on adapted 2D spatial moments of Hu and 3D spatial Invariant Rotation Moments are extracted from each segment. Tests with the proposed model using the Bosphorus Database for neutral vs non-neutral ROC I experiment, applying Linear Discriminant Analysis as classifier and more than one sample for training, achieved 98.7% of verification rate with 0.1% of false acceptance rate. By using the Support Vector Machine as classifier the rank1 experiment recognition rates of 99% and 95.4% have been achieved for a neutral vs neutral and for a neutral vs non-neutral, respectively. These results approach the state-of-the-art using Bosphorus Database and even surpasses it when Anger and Disgust expressions are evaluated. In addition, it was also evaluate the generalization of the method using the FRGC v2.0 database and competitive results been achieved, making the technique promising, especially for its simplicity.

## LISTA DE FIGURAS

Figura 1 – Pontos marcados baseados nas Action Units (AU’s). Extraído de Savran <i>et al.</i> (2008).	22
Figura 2 – Exemplo das imagens fornecidas pela base de dados Bosphorus.	25
Figura 3 – Exemplo das imagens fornecidas pela base de dados FRGC v2.0.	26
Figura 4 – Exemplo das imagens fornecidas pela base de dados 3DMAD.	27
Figura 5 – Imagem recortada da face em 3 escalas diferentes.	32
Figura 6 – Projeção 2D de nuvem de pontos 3D.	32
Figura 7 – Duas perspectivas de nuvens de pontos rotacionadas e transladadas entre si.	36
Figura 8 – Nuvem de pontos neutra submetida a uma subamostragem através do algoritmo do <i>Voxel Grid</i> .	41
Figura 9 – Nuvem de pontos 3D da face rotacionada em torno no eixo z com diferentes ângulos. Nuvem de pontos da base de dados Bosphorus bs003_N_N_0.	47
Figura 10 – Nuvem de pontos 3D da face rotacionada nos três eixos com diferentes ângulos. Nuvem de pontos da base de dados Bosphorus bs003_N_N_0.	48
Figura 11 – Recorte da Região útil da face com 80 mm a partir da ponta do nariz. Base de dados 3DMAD.	52
Figura 12 – Recorte de uma projeção da Figura 11 mostrando a separação das nuvens de pontos do indivíduo e cenário.	52
Figura 13 – Recorte da Região útil da face. 80 milímetros a partir da ponta do nariz. Base de dados FRGC v2.0.	53
Figura 14 – Recorte da Região útil da face. 80 milímetros a partir da ponta do nariz. Base de dados Bosphorus.	53
Figura 15 – Faces extraídas de 3 bases de dados: 3DMAD, FRGC v2.0 e Bosphorus.	54
Figura 16 – Orientação da Nuvem de pontos da face em relação ao eixo de coordenadas que tem sua origem na ponta do nariz. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.	57
Figura 17 – Nuvem de pontos da Face com cortes Frontais de 20 milímetros de espessura. Cortes Frontais numerados de 0 a 2. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.	57

Figura 18 – Nuvem de pontos da Face com cortes Sagitais de 20 milímetros de espessura. Cortes Frontais numerados de -3 a 3. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face. . . . .	58
Figura 19 – Nuvem de pontos da Face com cortes Transversal de 20 mm de espessura. Cortes Frontais numerados de -3 a 3. A Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face. . . . .	58
Figura 20 – Indivíduo bs003 extraído da base de dados Bosphorus. Amostra neutra, mulher e atriz. . . . .	59
Figura 21 – Cortes Frontais do indivíduo bs003 extraído da base de dados Bosphorus. Amostras: Neutra, Expressões emocionais, LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 mm da ponta do nariz. . . . .	60
Figura 22 – Cortes Sagitais do indivíduo bs003, extraído da base de dados Bosphorus. Amostras: Neutra e Expressões emocionais, LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 mm da ponta do nariz. . . . .	61
Figura 23 – Cortes Transversais do indivíduo bs003, base de dados Bosphorus, amostras 0 Neutra e Expressões emocionais; LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 milímetros da ponta do nariz. . . . .	62
Figura 24 – Cortes frontais de algumas expressões emocionais simples: Tristeza, Surpresa e Felicidade. . . . .	67
Figura 25 – Combinação dos cortes Frontal, Sagital e Transversal. Bosphorus rank-1. SVM-SVC-rbf. Análise do comportamento das combinações de cortes. Subgrupo das Emoções. . . . .	70
Figura 26 – Combinação dos cortes Frontal, Sagital e Transversal. Bosphorus rank-1. SVM-SVC-rbf. Análise do comportamento das combinações de cortes. Grandes grupos. . . . .	71
Figura 27 – VR para Neutral vs Não neutro, experimento rank-1 com combinação dos cortes ST, utilizando SVM-SVC-rbf. . . . .	73

## LISTA DE TABELAS

Tabela 1 – Sensores 3D com suas resoluções, preço e bases de dados associadas. . . . .	21
Tabela 2 – Distribuição por subconjuntos das amostras da base de dados Bosphorus. . .	24
Tabela 3 – Momentos de Hu para rotação de imagens 2D em tons de cinza. . . . .	33
Tabela 4 – Momentos de Hu para rotação de imagens 2D binárias. . . . .	33
Tabela 5 – Momentos HuXY para variação de densidade de pontos com base no voxel grid. . . . .	36
Tabela 6 – Momentos 3DIT para variação de densidade de pontos com base no voxel grid. .	36
Tabela 7 – Momentos HuXY para variação de densidade de pontos com base no voxel grid. . . . .	42
Tabela 8 – Momentos 3DIT para variação de densidade de pontos baseado no voxel grid. .	42
Tabela 9 – Momentos HuXY para rotação de nuvens de pontos em torno do plano XY. .	48
Tabela 10 – Momentos HuYZ para rotação de nuvens de pontos em torno do plano YZ. .	48
Tabela 11 – Momentos HuXZ para rotação de nuvens de pontos em torno do plano XZ. .	49
Tabela 12 – Momentos 3DITM para rotação de nuvens de pontos. . . . .	49
Tabela 13 – RR para Neutral <i>vs</i> Não neutro, experimento rank-1 com combinação dos cortes ST, utilizando SVM-SVC-rbf. . . . .	72
Tabela 14 – Taxas de Verificação a 0.1% FAR. Experimento ROC I na base Bosphorus com VR e RR para Neutro <i>vs</i> Não Neutro; e experimento ROC I e ROC III na base FRGC v2.0 com VR para Neutro <i>vs</i> Não Neutro. . . . .	74
Tabela 15 – RR para categorias distintas de conjuntos de teste na base de dados Bosphorus, experimento rank-1. Ming (2014), Emambakhsh e Evans (2017) e Al-Osaimi (2016) . . . . .	76

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>15</b>
<b>1.1</b>	<b>Objetivos Gerais e Específicos . . . . .</b>	<b>17</b>
<b>1.2</b>	<b>Resumo das contribuições e publicações . . . . .</b>	<b>17</b>
<b>1.3</b>	<b>Organização do documento de tese . . . . .</b>	<b>18</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>19</b>
<b>2.1</b>	<b>Faces 3D . . . . .</b>	<b>19</b>
<b>2.1.1</b>	<b><i>Nuvem de Pontos</i> . . . . .</b>	<b>19</b>
<b>2.1.2</b>	<b><i>Sensores e Resolução</i> . . . . .</b>	<b>20</b>
<b>2.1.3</b>	<b><i>Bosphorus</i> . . . . .</b>	<b>21</b>
<b>2.1.4</b>	<b><i>FRGC v2.0</i> . . . . .</b>	<b>24</b>
<b>2.1.5</b>	<b><i>3DMAD</i> . . . . .</b>	<b>26</b>
<b>2.1.6</b>	<b><i>Experimentos padronizados para reconhecimento de Faces 3D</i> . . . . .</b>	<b>27</b>
<b>2.2</b>	<b>Momentos Espaciais 2D e 3D . . . . .</b>	<b>29</b>
<b>2.2.1</b>	<b><i>Momentos de Hu</i> . . . . .</b>	<b>29</b>
<b>2.2.2</b>	<b><i>Momentos Invariantes 3D (3DITM)</i> . . . . .</b>	<b>32</b>
<b>2.2.3</b>	<b><i>Equalização da Densidade nas nuvens de pontos 3D</i> . . . . .</b>	<b>34</b>
<b>2.3</b>	<b>Alinhamento de objetos 3D . . . . .</b>	<b>35</b>
<b>2.3.1</b>	<b><i>RANSAC-Plane</i> . . . . .</b>	<b>37</b>
<b>2.3.2</b>	<b><i>Matriz Jacobiana</i> . . . . .</b>	<b>38</b>
<b>2.3.3</b>	<b><i>ICP</i> . . . . .</b>	<b>38</b>
<b>3</b>	<b>MÉTODO DE RECONHECIMENTO DE FACES 3D SEGMENTADAS EM REGIÕES TRIAXIAIS . . . . .</b>	<b>40</b>
<b>3.1</b>	<b><i>Equalização da Densidade nas nuvens de pontos 3D</i> . . . . .</b>	<b>40</b>
<b>3.2</b>	<b>Adaptações Momentos de Hu e 3DITM . . . . .</b>	<b>43</b>
<b>3.2.1</b>	<b><i>Momentos de Hu para nuvens de pontos</i> . . . . .</b>	<b>44</b>
<b>3.2.2</b>	<b><i>Momentos de Hu para múltiplos planos</i> . . . . .</b>	<b>45</b>
<b>3.2.3</b>	<b><i>Experimentos com Momentos de Hu e Rotações em Múltiplos Planos</i> . . . . .</b>	<b>46</b>
<b>3.3</b>	<b>Definição da Área Útil da Face . . . . .</b>	<b>48</b>
<b>3.3.1</b>	<b><i>Localização da Ponta do Nariz</i> . . . . .</b>	<b>50</b>
<b>3.3.2</b>	<b><i>Recorte da Face</i> . . . . .</b>	<b>51</b>

<b>3.4</b>	<b>Cortes Triaxiais da Face . . . . .</b>	54
<i>3.4.1</i>	<i>Efeito das Expressões nos Cortes . . . . .</i>	56
<i>3.4.2</i>	<i>Combinação dos Cortes . . . . .</i>	59
<b>3.5</b>	<b>Classificadores . . . . .</b>	61
<i>3.5.1</i>	<i>LDA . . . . .</i>	63
<i>3.5.2</i>	<i>SVM . . . . .</i>	64
<i>3.5.3</i>	<i>Escolha dos classificadores . . . . .</i>	65
<b>4</b>	<b>RESULTADOS E DISCUSSÃO . . . . .</b>	67
<b>4.1</b>	<b>Discussão Preliminar . . . . .</b>	67
<i>4.1.1</i>	<i>Resultados . . . . .</i>	70
<i>4.1.2</i>	<i>Melhor combinação de cortes e momentos . . . . .</i>	72
<i>4.1.3</i>	<i>Experimentos ROC I, ROC III e rank-1 . . . . .</i>	73
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS . . . . .</b>	77
	<b>REFERÊNCIAS . . . . .</b>	79
	<b>APÊNDICES . . . . .</b>	85
	<b>ANEXOS . . . . .</b>	85
	<b>ANEXO A – Artigo Aceito na IEEE RA-L . . . . .</b>	85

## 1 INTRODUÇÃO

Biometria refere-se à identificação de indivíduos a partir de características fisiológicas ou comportamentais que podem ser medidas. Existem diversas modalidades biométricas, dentre as mais populares, as realizadas a partir de: impressão digital, íris, retina, voz e face (RATYAL *et al.*, 2016; JAFRI; ARABNIA, 2009).

Segundo Jafri e Arabnia (2009), o reconhecimento facial oferece diversas vantagens sobre as demais modalidades biométricas. Ao contrário de métodos como identificação de impressão digital, o reconhecimento facial pode ser feito de forma passiva, sem que qualquer ação explícita do indivíduo seja solicitada, o que torna essa modalidade especialmente atraente para propósitos de segurança e monitoramento, por exemplo. Além disso, se comparada a modalidades como identificação de retina e íris, requer equipamentos de custo mais acessível.

Assim, reconhecimento facial é um tópico de pesquisa que desperta interesse para aplicação em diversas áreas: sistemas de monitoramento, sistemas de controle de acesso, sistemas para suporte na aplicação da lei e entretenimento (RATYAL *et al.*, 2016; JAFRI; ARABNIA, 2009; SOLTANPOUR *et al.*, 2017).

Muito têm sido produzido nas últimas décadas com o intuito de encontrar métodos mais eficientes e eficazes para reconhecimento facial a partir das convencionais imagens 2D (JAFRI; ARABNIA, 2009; SOLTANPOUR *et al.*, 2017). Apesar de resultados consideráveis terem sido alcançados, esses são sempre condicionados ao controle de aspectos como pose e iluminação durante a aquisição, o que traz limitações à sua aplicabilidade em cenários reais. Segundo Li *et al.* (2015), os maiores desafios associados ao reconhecimento facial 2D incluem fortes similaridades entre indivíduos distintos e variações entre amostras de um mesmo indivíduo. Esse último aspecto é significativamente afetado por mudanças na iluminação, variações de pose e expressões, variabilidade do *background* e oclusões (ZHAO *et al.*, 2003; HATEM *et al.*, 2015).

Recentemente, o reconhecimento facial 3D, impulsionado pelo desenvolvimento das tecnologias de aquisição de imagens 3D, tem sido investigado como alternativa para superar as limitações do domínio de imagens bidimensionais como sensibilidade à pose e iluminação. Além disso, imagens 3D oferecem informação de características geométricas da face capazes de melhorar a performance de sistemas de reconhecimento em comparação àqueles baseados em imagens 2D (HARIRI *et al.*, 2016) (SOLTANPOUR *et al.*, 2017).

Por outro lado, as técnicas 3D não são tão maduras e bem estabelecidas quanto as abordagens 2D. Algumas tarefas podem ser mais triviais para imagens 2D do que para imagens

3D (detecção automática de pontos fiduciais, por exemplo). Por essa razão, abordagens que combinam os dois tipos de dados, 2D e 3D, são também empregadas (KHADHRAOUI *et al.*, 2014).

No que se refere aos métodos de extração de características, três principais categorias são apresentadas na literatura: com base em atributos holísticos ou globais, com base em atributos locais e métodos híbridos.

Os métodos holísticos procuram por um conjunto de atributos globais de modelos 3D inteiros. Geralmente, nessa categoria, algoritmos de seleção de atributos como *Principal Component Analysis* (PCA) também são amplamente utilizados. Liu *et al.* (2013) utiliza o método holístico baseado em atributos esféricos harmônicos e depois emprega algoritmos de seleção de atributos para identificar um conjunto ótimo para a tarefa de reconhecimento.

Já aqueles com base em atributos locais focam na extração destes da face ou de regiões relevantes desta (região dos olhos e nariz ou *keypoints*, por exemplo). Por não requererem modelos completos da face, métodos baseados em atributos locais lidam melhor com situações de oclusão, autoclusão e expressões, sendo mais apropriados para propósitos de identificação e verificação (SOLTANPOUR *et al.*, 2017; ZHOU *et al.*, 2014).

Emambakhsh e Evans (2017) empregaram um método baseado em atributos locais para investigar o efeito da região nasal para o reconhecimento facial 3D robusto a expressões. Seu método se estabelece sobre a precisa determinação de *keypoints* da região do nariz e realiza extração de atributos baseados em vetores normais de superfície sobre regiões e curvas da região nasal para, finalmente, determinar as regiões mais estáveis sob variação de expressões e que são bem úteis para fins de reconhecimento.

Al-Osaimi (2016) utilizou múltiplos *keypoints* para derivar regiões superficiais e apresentou técnica de extração de descritores invariantes à rotação para aplicação em reconhecimento facial em imagens 3D estáticas e dinâmicas (vídeos 3D). Seu método não abordou situações de oclusão parcial da face.

Li *et al.* (2015) também propuseram um método considerando atributos locais para realizar reconhecimento facial 3D. A partir de algoritmos de detecção de *keypoints* 3D baseados em curvatura, regiões de alta curvatura foram identificadas nos modelos de faces 3D e atributos invariantes foram extraídos de cada *keypoint*. Esses atributos combinados descrevem a forma das regiões que contém os *keypoints*, alcançando resultados significativos e exploram situações de variabilidade de expressões, pose e oclusão.

Os métodos híbridos, por sua vez, são menos populares e usam uma combinação de atributos globais e locais para reconhecimento facial (RATYAL *et al.*, 2016).

## 1.1 Objetivos Gerais e Específicos

Neste trabalho tem-se como objetivo geral apresentar uma nova estratégia para o reconhecimento de faces 3D com base no seccionamento tri-axial da nuvem de pontos que representa uma face 3D. Tem-se como hipótese que é possível extrair características em diferentes regiões da superfície da face, ainda que grande parte dessa superfície possa ser deformada em consequência das expressões. Além disso, adota-se como premissa o uso mínimo de pontos-chave, cuja localização pode ser custosa e sujeita a erros, estabelecendo uma forte dependência *a priori* e comprometendo a técnica. Como objetivos específicos, elencam-se:

- Avaliar o uso da ponta do nariz como ponto central para delimitação da área significativa para o reconhecimento da face;
- Avaliar e propor formas de segmentar a superfície da face em regiões, ou fatias, compreendidas entre três conjuntos de planos paralelos (Frontal, Sagital e Transversal);
- Avaliar e adaptar momentos espaciais invariantes à rotação como discriminantes para as regiões definidas para a face 3D;
- Propor um algoritmo de reconhecimento facial usando os resultados das avaliações anteriores e submetê-los a diferentes classificadores;
- Avaliar a generalização da proposta para diferentes cenários, comparando os resultados encontrados com os apresentados em publicações especializadas no mesmo domínio.

## 1.2 Resumo das contribuições e publicações

As principais contribuições deste trabalho são:

1. a adaptação dos momentos de Hu como descritores, originalmente aplicados no domínio 2D;
2. a extração desses descritores de regiões de superfície ao longo dos três eixos ortogonais e suas combinações (a literatura usualmente considera um único eixo);
3. a mínima dependência de pontos-chave para o reconhecimento de faces 3D. Como consequência, os resultados de classificação são comparáveis aos trabalhos do estado da arte, mas obtidos em menor tempo.

Os principais resultados deste trabalho foram aprovados para publicação com o título *Tri-axial slicing for 3D face recognition from adapted rotational invariants spatial moments and minimal keypoints dependence* nos seguintes veículos:

- Periódico *IEEE Robotics and Automation Letters*, doi: 10.1109/LRA.2018.2854295
- Conferência *Intelligent Robots and Systems (IROS 2018)*, Barcelona - ESPANHA, de 1 a 5 de Outubro de 2018

### **1.3 Organização do documento de tese**

Este documento está organizado em 5 capítulos. No Capítulo 1 foi feita a contextualização do trabalho bem como apresentados os objetivos da pesquisa; No Capítulo 2 são discutidos os aspectos que fundamentam e dão suporte ao reconhecimento de faces 3D em nuvens de pontos; O método de reconhecimento de faces proposto é detalhado no Capítulo 3. No Capítulo 4 são discutidos os resultados alcançados e, finalmente, no último Capítulo são elencadas as conclusões e as perspectivas de aprimoramento deste trabalho.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são discutidos os aspectos teóricos que fundamentam o reconhecimento de faces 3D e algumas das técnicas usadas neste trabalho. São discutidos os conceitos de Faces 3D e a sua representação em forma de nuvens de ponto. Em seguida são apresentados os momentos espaciais invariantes à rotação utilizados para discriminar regiões predefinidas da face 3D. Finalmente, são apresentadas técnicas que permitem alinhar objetos 3D, em especial para o ajuste da orientação da face, etapa importante do processo de reconhecimento proposto.

### 2.1 Faces 3D

As faces podem ser representadas de diferentes formas, dependendo dos objetivos aos quais se aplicam. Algumas representações são baseadas em imagens 2D e outras em imagens 3D. Para algumas aplicações, as duas formas de representação podem ser conjugadas.

Para dar suporte a trabalhos como este, uma lista de banco de dados de faces 3D é apresentada por Soltanpour *et al.* (2017), onde são detalhados aspectos como número de indivíduos considerados, número de amostras, se possuem ou não imagens 2D associadas e o sensor utilizado na aquisição.

Neste trabalho, as faces são representadas em nuvens de pontos, cujo conceito é apresentado na próxima subseção.

#### 2.1.1 Nuvem de Pontos

Uma nuvem de pontos é um conjunto de pontos no espaço de 3 dimensões, elas são produzidas por scanners 3D(YIN *et al.*, 2006a; ERDOGMUS; MARCEL, 2013; SAVRAN *et al.*, 2008) diretamente ou reconstruídas por um conjunto de câmeras 2D, sendo pelo menos duas câmeras, em posições distintas. O termo utilizado de reconstrução é apropriado pois as imagens 2D são projeções dos objetos em 3 dimensões que estão sendo fotografados.

As técnicas de reconstrução de imagens 3D a partir de imagens 2D estão tornando-se comuns pois alguns novos modelos de celulares já vem com mais de uma câmera na parte de trás do aparelho. Essas 2 câmeras tanto ajudam a tirar fotos com múltiplos pontos focais, quanto possibilitam a reconstrução do ambiente 3D que está sendo capturado.

A nuvem de pontos pode ser densa como a fornecida por equipamentos como tomógrafos que analisam as partes internas de órgãos humanos como pode ser visto em Salama *et*

*al.* (2018). O uso de nuvens e pontos 3D de tomografias são largamente utilizadas em estudos para identificação e reconhecimento de doenças como câncer(DIEDERICH *et al.*, 2001; SANTOS *et al.*, 2014).

O uso de imagens 3D também está presente na indústria na área de metrologia e inspeção industrial, em que objetos são escaneados e comparados, computacionalmente, à modelos pré-existentes de forma a identificar tanto as medidas quanto alguma alteração ocorrida que represente um risco ao bom funcionamento de algum equipamento (SCHUMACHER *et al.*, 2018; SKARŻYŃSKI; SUCHORZEWSKI, 2018).

### **2.1.2 Sensores e Resolução**

As bases de dados de faces 3D são criadas com objetivos específicos e por isso possuem características que as credenciam a ser utilizadas em determinados tipos de problemas.

Embora deseje-se saber o número de indivíduos da base, bem como sua distribuição em termos de gênero, etnia, se são atores ou não, se possuem barba ou não, ou, ainda, deseje-se conhecer o número de amostras da base, há uma informação importante relacionada com o tipo de sensor utilizado para realizar as aquisições.

Dependendo do sensor utilizado, a base pode ter, além da imagem 3D, uma imagem 2D associada com a informação Imagem colorida - (Red, Green, Blue) (RGB) do pixel 2D associada ao voxel ( $x, y, z$ ) da imagem 3D. Outra característica importante do sensor é a sua resolução, ou seja, a quantidade de pontos por área do objeto mapeado. Esta resolução varia com a distância do objeto em relação ao sensor e tem seus valores máximos descritos em seus documentos de características. Além disto, cada sensor tem um limite de distância de funcionamento, alguns são para aquisições de curtas distâncias, outros de média ou longa distâncias. Assim, além da resolução variar de sensor para sensor, e dentro de cada sensor com a distância, há uma distância específica a qual o objeto deve ser posicionado para que tenha uma boa aquisição.

De acordo com a forma de aquisição, é possível que todas as amostras adquiridas pelo sensor possuam a mesma densidade de voxels, mas isso vai depender das condições definidas. O indivíduo pode estar em um local preparado para realizar as aquisições, como em uma cadeira posicionada a uma distância sempre igual ao sensor, ou ela pode ser realizada ao ar livre, onde a distância até o indivíduo não possa ser controlada. Do ponto de vista da aquisição, não há qualquer problema nos dois cenários descritos anteriormente, desde que estejam adequadamente

Sensor	Base de Dados	Resolução Máxima	Preço
3DMD digitizer	BU-3DFE de Yin <i>et al.</i> (2006a)	< 0,5 mm	> US\$ 20.000
Minolta Vivid 3D scanner	FRGC v2.0 de Phillips <i>et al.</i> (2005a)	0,1 mm	> US\$ 20.000
Inspeck Mega Capturor II 3D	Bosphorus de Savran <i>et al.</i> (2008)	0,3 mm	< US\$ 5000
Kinect	3DMAD de Erdogmus e Marcel (2013)	1 cm	< US\$ 100
SR300	-	1 mm	< US\$ 200
GO!SCAN 20©	-	0,3 mm	> US\$ 20.000

Tabela 1 – Sensores 3D com suas resoluções, preço e bases de dados associadas.

anotados e este seja um dos objetivos da base de dados em questão.

Amostras 3D de faces com densidades de pontos diferentes podem afetar métodos de reconhecimento. Os atributos extraídos para realizar o reconhecimento podem ter uma resolução mínima de densidade de pontos para alcançar altas taxas de reconhecimento.

Outro fator importante é o preço do sensor que pode tornar o processo de aquisição muito oneroso para algumas aplicações. Sensores como 3DMD digitizer, Minota Vivid 3D scanner, GO!SCAN 20© custam algumas dezenas de milhares de dólares mas possuem uma resolução da ordem de 0,1 milímetros. Atualmente, sensores como o Kinect e mais recentemente o Real Sense© da INTEL (equipamento SR300, por exemplo) tem trazido a possibilidade de realizar aquisições de imagens 3D e 2D a baixo custo, menos de US\$ 200, o que já pode permitir o uso comercial de equipamentos que realizem identificação ou verificação com faces 3D.

A tabela 1 mostra alguns sensores com seus preços, resolução e se foram utilizados na aquisição de alguma base de dados.

A seguir serão apresentados os detalhamentos de três bases de dados utilizadas neste trabalho.

### 2.1.3 *Bosphorus*

Muitas bases de dados tem sido utilizadas com o propósito de realizar o reconhecimento facial de imagens 3D. A Bosphorus certamente é a mais desafiadora pois traz uma diversidade de expressões que não é vista em outra com a mesma finalidade. Dentre os indivíduos selecionados, constam atores profissionais, inseridos propositalmente para garantir que as expressões manifestadas pudessem ter o maior grau de dificuldade possível, uma vez que são treinados para manifestá-las exageradamente.

A base de dados Bosphorus foi proposta por Savran (SAVRAN *et al.*, 2008) e é uma

das mais completas e complexas utilizada com o propósito de realizar reconhecimento facial envolvendo técnicas 3D. Possui amostras 2D e 3D com as respectivas coordenadas dos pontos chave localizadas nas Unidades de Ação (Action Units - AU), definidas por Ekman e Friesen (EKMAN; FRIESEN, 1976). Além disso, fornece um mapeamento entre os dados 3D e seus correspondentes pixels de imagem 2D. Assim, torna-se possível trabalhar em espaços 2D e 3D, ou em uma combinação destes.

Ademais, consiste em 4.666 amostras faciais de 105 indivíduos em várias expressões, poses e condições de oclusão capturadas usando o scanner Inspec 3D com base em um princípio óptico, que combina o Triângulo Óptico Ativo e o método Moiré Deslocado pela Fase. A maioria dos indivíduos tem entre 25 e 35 anos, dos quais 60 são homens e 45 são mulheres, e são caucasianos. Há 29 atores/atrizes, profissionais incorporados ao banco de dados. Esta presença confere uma complexidade adicional, pois pode-se notar que as emoções e outros movimentos amplos fazem uma variação de rosto muito expressiva. Barba/bigode estão presentes em 18 homens e 15 outros sujeitos têm pelos faciais curtos. A maioria dos indivíduos tem até 54 amostras, mas 34 têm apenas 31 amostras. Desta forma, não há uma distribuição equitativa nem de amostras neutras nem de amostras de expressões. Alguns indivíduos possuem apenas uma amostra neutra (29 indivíduos), enquanto outros chegam a ter até 4 (58 indivíduos), o mesmo acontecendo para as expressões de emoções, 6 ao todo, que estão presentes em 65 dos 105 indivíduos e em 34 há apenas uma das expressões de emoção.

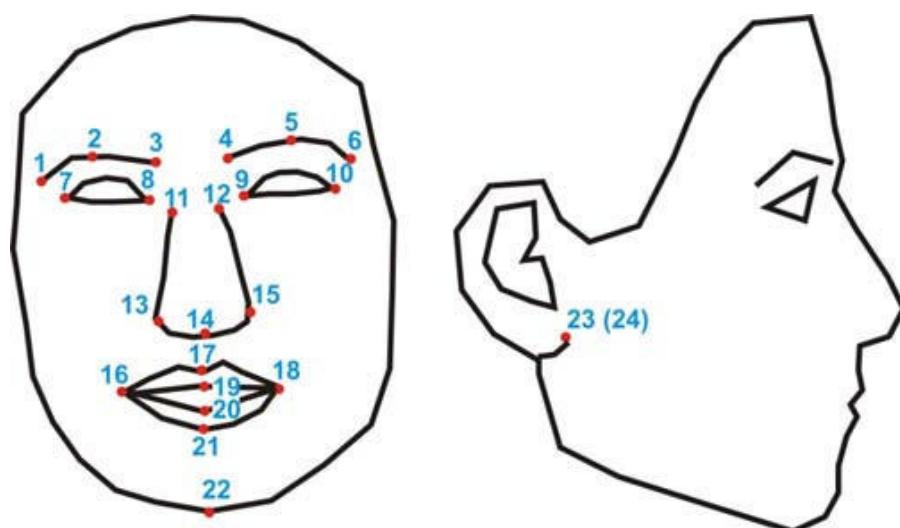


Figura 1 – Pontos marcados baseados nas AU's. Extraído de Savran *et al.* (2008).

Os rótulos das amostras possuem relação com as Unidades de Ação (Action Units), ou simplesmente AU, definidas por nada (a) e que definem pontos representativos da face rela-

cionado com boca, olhos, nariz, mandíbula e demais estruturas. As AU utilizadas pela base Bosphorus podem ser vistas na Figura 1.

As amostras podem ser divididas em subconjuntos, descritos a seguir:

### 1. 34 Expressions

- a) 6 Emotional Expressions(E): Anger, Disgust, Fear, Happy, Sadness, Surprise.
- b) 1 Neutral (N).
- c) 20 Lower Face Action Unit (LFAU); usando principalmente as AU's envolvidos com lips, jaw, nose and mouth
- d) 5 Upper Face Action Unit (UFAU); usando principalmente as AU's envolvidos com brow and eyes;
- e) 3 Combined Action Units (CAU); combinação de algumas AU's envolvendo lips and jaw.

### 2. 13 Poses

- a) 7 Yaw Rotation (YR): Right (R) with 10° (YR\_R10), 20°(YR\_R20), 30°(YR\_R30), 45°(YR\_R45) e 90°(YR\_R90) e Left(L) with 45°(YR\_L45) and 90°(YR\_L90).
- b) 4 Pitch Rotation (PR): Strong upwards (PR\_SU), Slight upwards(PR\_U), Slight downwards(PR\_D) and Strong downwards(PR\_SD).
- c) 2 Cross Rotation (CR): Yaw Right 45° and and Pitch Upwards 20° (CR\_RU); ;Yaw Right 45° and Pitch Downwards 20° (CR\_RD);

### 3. 4 Occlusions

- a) Eyes by hands (O\_EYE)
- b) Eyes by hands (O\_GLASSES)
- c) Mouth by hands (O\_MOUTH)
- d) Face by hair (O\_HAIR)

### 4. Ignored samples are present with the label IGN.

Nos experimentos realizados neste trabalho, foram utilizadas 2.902 amostras de Expressões (62% do total de amostras), como definido por Ocegueda *et al.* (2011), onde há uma subdivisão entre os subconjuntos Neutros (N) e Não-Neutros(NN) . O subconjunto Neutros é composto por 299 amostras, enquanto o subconjunto Não-Neutros é composto por 2.603 amostras. como definido na Tabela 2.

Ocegueda *et al.* (2013) considera que essas amostras não possuem qualquer tipo de oclusão, seja as diretas realizadas pelas mãos, cabelo e óculos, sejam as relacionadas às

Conjunto/Subconjunto	Rótulo	Amostras (#)
Expressões	E	2902
Neutras	N	299
Raiva	ANGER	71
Disgust	DISGUST	69
Medo	FEAR	70
Felicidade	HAPPY	106
Tristeza	SAD	66
Surpresa	SURPRISE	71
Lower Face Action Unit	LFAU	1549
Upper Face Action Unit	UFAU	432
Combined Action Unit	CAU	169
Rotações		1365
Yaw Rotation	YR	735
Pitch Rotation	PR	419
Cross Rotation	CR	211
Oclusões	O	381
Eye by hand	EYE	105
Mouth by hands	MOUTH	105
Eye by Glasses	GLASSES	104
Face by Hair	HAIR	67
Sem rótulo	IGN	18

Tabela 2 – Distribuição por subconjuntos das amostras da base de dados Bosphorus.

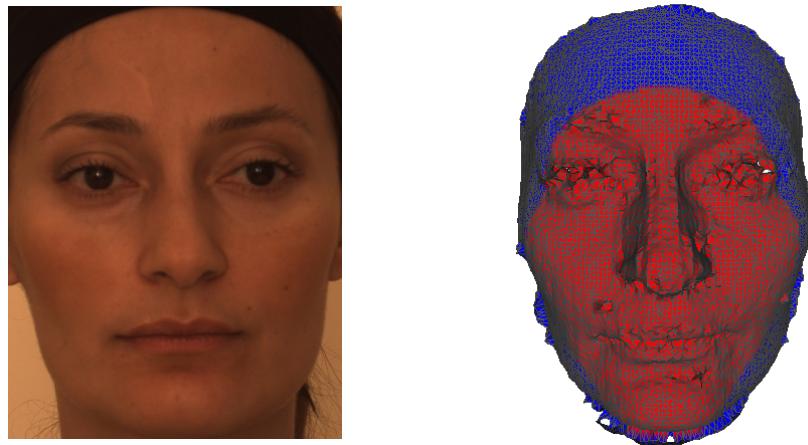
inclinações da face em relação ao sensor, definidas pelas poses, ou seja, quando uma das metades da face oculta a outra devido à inclinação no momento da amostragem. Esta escolha também pode ser considerada para efeito de uma comparação mais justa quando estão sendo analisados resultados de outras base de dados como a FRGC v2.0 de Phillips *et al.* (2005a) e Yin *et al.* (2006b).

Como declarado por Soltanpour *et al.* (2017), é difícil encontrar uma comparação justa entre métodos na literatura sobre reconhecimento de faces, principalmente devido aos diferentes experimentos serem apresentados em diversas situações e em distintas condições e bases de dados. Por isso, as principais comparações realizadas neste trabalho foram com resultados de trabalhos que utilizaram condições idênticas ou muito similares.

Na Figura 2 podem ser vistas as duas amostras de um mesmo indivíduo tomada simultaneamente pela câmera RGB e pelo scanner 3D.

#### 2.1.4 FRGC v2.0

Como descrito por Phillips *et al.* (2005b), dados da base FRGC v2.0 foram coletados na Universidade de Notre Dame e fazem parte de um projeto progressivo de coleta de dados



(a) Imagem RGB.

(b) Imagem de Profundidade.

Figura 2 – Exemplo das imagens fornecidas pela base de dados Bosphorus.

biométricos. Considera-se uma sessão o conjunto de todas as imagens de uma pessoa realizadas no momento da coleta de dados biométricos. Uma sessão consiste de 4 imagens controladas, 2 não controladas e uma tridimensional.

Ainda segundo Phillips *et al.* (2005b), as imagens controladas são realizadas em um estúdio, todas são realizadas de maneira frontal sob duas condições de luminosidade (com 2 e 3 luzes do estúdio) e com 2 expressões faciais (neutra e sorrindo). As imagens não controladas foram realizadas com uma configuração variada de iluminação, sendo tomadas em corredores, átrios ou ao ar livre, estas imagens também são tomadas com 2 expressões faciais (neutra e sorrindo).

As imagens 3D são tomadas em um ambiente de iluminação controlado que é apropriado ao sensor Vivid 900/910, onde as pessoas eram postas à distância de 1,5 metros do sensor. As imagens RGB foram realizadas por uma câmera 4 Megapixel Canon PowerShot G2 nos modos 1704x2272 pixels or 1200x1600 pixels.

Vale salientar que a base FRGC foi incrementada e possui dados do Outono de 2002, Primavera de 2003, Outono de 2003 e Primavera de 2004. No total, há 6 experimentos realizados na aquisição desta base de dados.

Nosso trabalho, bem como os mais recentes como Berretti *et al.* (2014), Liu *et al.* (2013), Li *et al.* (2015), fazem uso do experimento que utiliza as amostras do período que compreende o Outono de 2003 e a Primavera de 2004. Esta fração da base de dados FRGC v2.0 compreende 466 indivíduos e um total de 4.007 amostras (PHILLIPS *et al.*, 2005c) coletadas durante o ano letivo de 2003-2004. É composto por 2.410 amostras 3D neutras e 1.597 amostras 3D não neutras de diferentes expressões faciais (nojo, feliz, tristeza, surpresa, outras).

Além de possuir um sensor de alta resolução para imagens 3D, esta base de dados permite realizar um experimento que consiste em treinar com amostras tomadas em um determinado período e testar com amostras de um período de vários meses, do Outono de 2003 para a Primavera de 2004.

Na Figura 3 podem ser vistas as duas amostras de um mesmo indivíduo tomada simultaneamente pela câmera RGB e pelo scanner 3D para a base de dados FRGC v2.0.



(a) Imagem RGB.

(b) Imagem de Profundidade.

Figura 3 – Exemplo das imagens fornecidas pela base de dados FRGC v2.0.

### 2.1.5 3DMAD

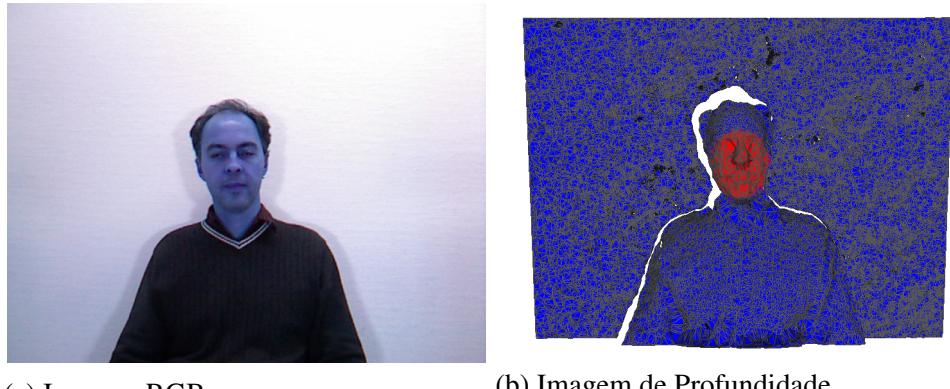
Esta base de dados foi desenvolvida por Erdogmus e Marcel (2013) utilizando Microsoft Kinect for Xbox 360. Como outros sensores, o Kinect, possui duas câmeras que fornecem imagens RGB (24-bits) e de profundidade (11 - bits), ambas com tamanho de 640x480, com uma frequencia de 30 imagens por segundo.

O objetivo da construção desta base de dados era testar técnicas que detectem o uso de máscaras (por uma máscara com o rosto de outra pessoa) para poder realizar uma falsa autenticação. Esse tipo de problema é bem comum em sistemas de autenticação que utilizam imagens RGB, onde uma foto impressa pode ser colocada em frente a câmera e o sistema pode realizar uma falsa autenticação. O uso de imagens 3D nestes sistemas permite uma outra camada de autenticação que inviabilize o uso de máscaras ou imagens impressas.

A base de dados é composta por 17 indivíduos, foram realizadas duas sessões de aquisição de dados com uma diferença de 2 semanas entre elas. Em cada uma das sessões, cada indivíduo realizou 5 filmagens de 10 segundos cada. Uma terceira sessão, nas mesmas condições é realizadas, mas esta com a presença de uma máscara na face dos indivíduos. Cada uma das 5

filagens por sessão realizou a aquisição de 600 amostras por indivíduo, 300 de imagens RGB e 300 de imagens de profundidade.

As condições de aquisição foram bem controladas com uma tela na parte posterior de onde os indivíduos eram posicionados, e a iluminação também foi controlada de forma a minimizar sombras na face. Uma amostra da imagem RGB e seu homólogo em profundidade pode ser visto em 4.



(a) Imagem RGB.

(b) Imagem de Profundidade.

Figura 4 – Exemplo das imagens fornecidas pela base de dados 3DMAD.

Uma vantagem desta base de dados, em relação às demais já citadas, é que ela é obtida de um sensor que tem um preço acessível, logo, um método que funcione nele pode ser implementado em sistemas reais que utilizem o sensor em questão. Trabalhos recentes tem utilizado essa base de dados para testar algoritmos que detectem tentativa de sabotar sistemas de autenticação que utilizam imagens 2D/3D como em Siddiqui *et al.* (2016), Li *et al.* (2016).

### **2.1.6 Experimentos padronizados para reconhecimento de Faces 3D**

Nesta seção serão apresentados os tipos de experimentos realizados nas diversas bases de dados de face 3D.

As bases de dados de faces 3D, principalmente as utilizadas para reconhecimento facial, possuem algumas configurações comuns, embora não padronizadas, de realizar aquisições de amostras Neutras e Não-Neutras. As amostras Neutras são aquelas que indicam a expressão natural do indivíduo enquanto as Não-Neutras expressam algum tipo de emoção ou movimento de determinada região da face. As bases FRGC v2.0, Bosphorus e BU-3DFE possuem essas amostras marcadas.

Dentre as expressões, as mais comuns são as de emoção: Raiva (ANGER), Repugnância (DISGUST), Medo (FEAR), Felicidade (HAPPY), Tristeza (SADNESS) e Surpresa

(SURPRISE). Estas podem ser vistas nas 3 bases de dados citadas acima, embora nem todas estejam presentes na FRGC v2.0, e mesmo considerando as outras duas bases, nem todas os indivíduos possuem amostras de todas as expressões.

A Bosphorus é a que possui uma abrangência maior de expressões, como mencionado no item 2.1.2 desta tese.

Outra característica importante é a diferença de tempo nas aquisições. Algumas bases realizam aquisições de amostras de um mesmo indivíduo em períodos distintos. A FRGC v2.0, por exemplo, tem 2 aquisições com 6 meses de diferença, uma sendo realizada no Outono de 2003 (Fall2003) e a segunda na Primavera de 2004 (Spring2004).

Além disso, é possível haver mais de uma amostra por indivíduo de determinada expressão. Na Bosphorus há até 4 amostras Neutras por indivíduos, já as expressões emocionais possuem até uma amostra por indivíduo e nem todos possuem todas as 6 emoções.

A partir destes cenários há alguns tipos de experimentos que podem ser realizados, quais sejam:

**ROC-I** É o experimento mais comum e basicamente faz o confronto entre as amostras Neutras e Não-Neutras. O conjunto de treinamento (gallery) é composto de todas as amostras Neutras da base de dados, enquanto o conjunto de testes é composto pelas amostras Não-Neutras. A base Bosphorus possui 105 indivíduos, 299 amostras Neutras e 2.603 amostras Não-Neutras (envolvendo expressões de emoção, LFAU, UFAU e CAU). A base de dados FRGC v2.0 possui 466 indivíduos com 2.399 amostras Neutras e 1.608 amostras Não-Neutras.

**ROC-III** É um experimento que apenas as bases de dados que possuem aquisições em períodos diferentes para os mesmos indivíduos podem executar. Das bases citadas, a FRGC v2.0 é a que possui amostras com estas características: 1.834 amostras no outono de 2003 (Fall2003) e 1.628 amostras na primavera de 2004 (Spring2004). É um experimento interessante do ponto de vista que é capaz de demonstrar se o método é robusto o suficiente para reconhecer indivíduos com o passar do tempo, onde a aparência pode ter sido modificada pelo corte de cabelo, aumento ou diminuição da barba, e mudança comportamental em duas estações tão diferentes como o outono e a primavera, que motivam emoções mais contidas e mais expansivas, respectivamente.

**rank-1** É o experimento mais completo e complexo uma vez que utiliza como base de dados de treinamento apenas uma amostra neutra por indivíduo. É considerado o teste mais

complexo porque o método deve ser de tal forma robusto que com apenas uma amostra de um indivíduo, seja possível reconhecer outras amostras do mesmo indivíduo com ou sem expressões. Desta forma, os atributos devem ser tais que não sofram muita variação mesmo com grandes deformações da face. Utiliza-se as primeiras amostras neutras como sendo as de treinamento. Neste caso, há dois tipos de sub-experimentos: o Neutro *vs* Neutro e o Neutro *vs* Não-Neutro. O primeiro sub-experimento considera no conjunto de testes as amostras neutras que não estejam presentes no conjunto de treinamento (como na base Bosphorus que possui um total de 299 amostras neutras, sendo 105 do conjunto de treinamento e 194 do conjunto de teste). O outro sub-experimento é o teste com as amostras de expressões, neste caso, são 2.603 amostras Não-Neutras.

## 2.2 Momentos Espaciais 2D e 3D

Como este trabalho tinha o objetivo de identificar indivíduos através da face e esta sofre variações em relação à rotação, translação, expressões e movimentos musculares, foi necessário selecionar atributos com propriedades invariantes a essas alterações.

Há, na literatura, dois momentos que preservam suas características quando submetidos às alterações de translação e rotação, são eles: momentos de Hu (HuM) e momentos invariantes 3D (3DIM).

### 2.2.1 Momentos de Hu

Os momentos de Hu, por exemplo, são invariantes à translação, à rotação e à escala para imagens 2D e possuem um número fixo de atributos que é 7. Os momentos de Hu (HU, 1962) são originados dos momentos de funções contínuas de 2 variáveis  $f(x,y)$  e foram utilizados originalmente para solucionar o problema de reconhecimento de letras em imagens binárias.

Os momentos de Hu derivam dos momentos espaciais que podem ser vistos na Equação 2.1.

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy \quad (2.1)$$

Quando os momentos de Hu são utilizados para imagens binárias ou em tons de cinza, o termo  $f(x,y)$  é substituído pelo valor do pixel  $p(x,y)$ , como pode ser visto na versão

discreta na Equação 2.2, já em sua forma normalizada para translação.

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j f(x, y) \quad (2.2)$$

Para serem invariantes à escala, os momentos 2D são normalizados de acordo com a equação abaixo:

$$\psi_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}}, (i+j) \geq 2 \quad (2.3)$$

Finalmente, os momentos de Hu invariantes à rotação são dados pelas equações 2.4 a 2.10:

$$\Psi_1 = \psi_{20} + \psi_{02} \quad (2.4)$$

$$\Psi_2 = (\psi_{20} - \psi_{02})^2 + 4\psi_{11}^2 \quad (2.5)$$

$$\Psi_3 = (\psi_{30} - 3\psi_{12})^2 + (3\psi_{21} - \psi_{03})^2 \quad (2.6)$$

$$\Psi_4 = (\psi_{30} + \psi_{12})^2 + (\psi_{21} + \psi_{03})^2 \quad (2.7)$$

$$\begin{aligned} \Psi_5 = & (\psi_{30} - 3\psi_{12})(\psi_{30} + \psi_{12})[(\psi_{30} + \psi_{12})^2 - 3(\psi_{21} - \psi_{03})^2] + \\ & (3\psi_{21} - \psi_{03})(\psi_{21} + \psi_{03})[3(\psi_{30} + \psi_{12})^2 - (\psi_{21} + \psi_{03})^2] \end{aligned} \quad (2.8)$$

$$\Psi_6 = (\psi_{20} - \psi_{02})[(\psi_{30} + \psi_{12})^2 - (\psi_{21} + \psi_{03})^2] + 4\psi_{11}(\psi_{30} + \psi_{12})(\psi_{21} + \psi_{03}) \quad (2.9)$$

$$\begin{aligned} \Psi_7 = & (3\psi_{21} - \psi_{03})(\psi_{30} + \psi_{12})[(\psi_{30} + \psi_{12})^2 - 3(\psi_{21} - \psi_{03})^2] + \\ & (\psi_{30} - \psi_{12})(\psi_{21} + \psi_{03})[3(\psi_{30} + \psi_{12})^2 - (\psi_{21} + \psi_{03})^2] \end{aligned} \quad (2.10)$$

Os momentos 2D que precisam ser calculados para compor os momentos de Hu são  $(\mu_{00}, \mu_{02}, \mu_{03}, \mu_{11}, \mu_{12}, \mu_{20}, \mu_{21}, \mu_{30})$ .

Para demonstrar esta propriedade foi realizado um experimento com uma imagem 2D de uma face 3D projetada com cortes Frontais a cada 10 milímetros. A Figura 5a é o recorte original colorido, passado por um processo de transformação em tons de cinza e outro para imagem binária, que podem ser vistos nas Figuras 5b e 5c, respectivamente. A imagem 2D em tons de cinza e binaária foram rotacionadas a cada  $30^\circ$  graus e os momentos de Hu extraídos. O efeito da rotação pode ser visto na Figura 6. Neste ponto, é necessário salientar que as imagens foram submetidas apenas à rotação, sem translação ou alteração de escala.

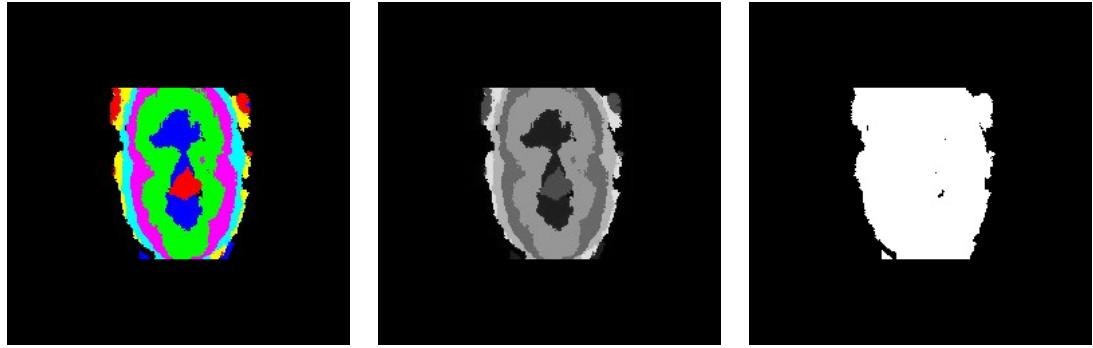
A Tabela 3 mostra os valores dos 7 momentos de Hu para as rotações de  $0^\circ$  a  $330^\circ$  para a imagem em tons de cinza e a Tabela 4 os mesmos 7 momentos para as rotações da imagem binária.

Algumas informações importantes podem ser extraídas das Tabelas 3 e 4:

- Os momentos de Hu para as imagens em tons de cinza rotacionadas possuem pequenas variações que podem ser vistas em alguns dos elementos da Tabela;
- O mesmo efeito citado anteriormente acontece com as rotações das imagens binárias;
- Os momentos de Hu são diferentes para a imagem binária e para a imagem em tons de cinza;
- As escalas dos momentos de Hu são bem distintas umas das outras variando, para os dois casos em questão, de  $10^{-3}$  a  $10^{-28}$ .

A diferença entre os momentos da imagem em tons de cinza para a imagem binária acontece exatamente devido ao valor que substitui a função  $f(x,y)$  para cada caso. Para as imagens binárias  $f(x,y) = 1$  para todos os pontos e para as imagens em tons de cinza  $f(x,y) = p(x,y)$ , onde  $p(x,y)$  é o tom de cinza do pixel na posição  $(x,y)$ . Este experimento foi importante para que tirar a dúvida de que a função  $f(x,y)$  da equação 2.1 precisa ser necessariamente um valor constante para todos os pixels da imagem.

Trabalhos recentes continuam utilizando momentos de Hu em imagens 2D como o de Thangaraj *et al.* (2018). E a utilização de momentos de Hu para nuvens de pontos é algo mais recente e pode ser visto em Yamashita *et al.* (2017) para descrever movimento humano.

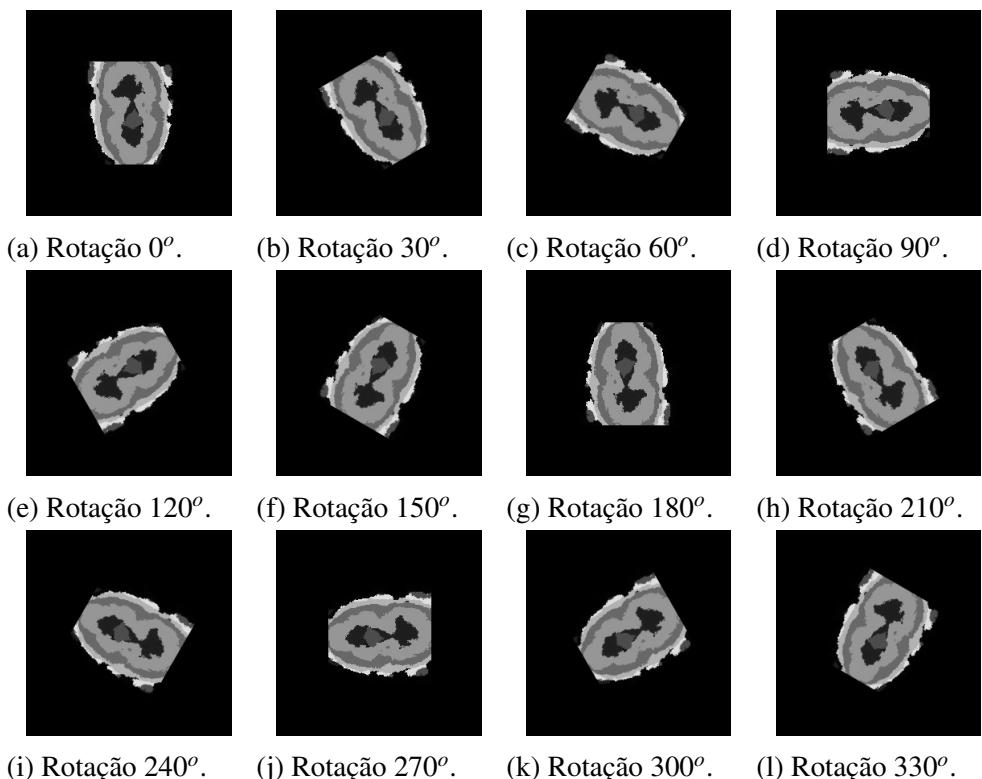


(a) Original colorida.

(b) Tons de cinza.

(c) Binária.

Figura 5 – Imagem recortada da face em 3 escalas diferentes.



(a) Rotação 0°.

(b) Rotação 30°.

(c) Rotação 60°.

(d) Rotação 90°.

(e) Rotação 120°.

(f) Rotação 150°.

(g) Rotação 180°.

(h) Rotação 210°.

(i) Rotação 240°.

(j) Rotação 270°.

(k) Rotação 300°.

(l) Rotação 330°.

Figura 6 – Projeção 2D de nuvem de pontos 3D.

### 2.2.2 Momentos Invariantes 3D (3DIM)

Outros momentos utilizados no trabalho foram os Momentos Invariantes à Rotação 3D (3DRIM). Como trabalhamos com imagens em 3D, a aplicação desses momentos pareceu mais adequada e também uma forma de comparar resultados em relação à aplicação dos Momentos de Hu.

As fórmulas dos Momentos de Hu são uma simplificação dos 3DIM para duas dimensões. As equações dos momentos invariantes 3D, desde sua forma contínua, passando pelas equações discretas são mostradas a seguir.

Tabela 3 – Momentos de Hu para rotação de imagens 2D em tons de cinza.

Rotação	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
$0^\circ$	1.44e-03	1.02e-07	9.69e-11	6.31e-14	-1.50e-25	-1.66e-17	4.38e-26
$30^\circ$	1.44e-03	1.02e-07	9.71e-11	6.37e-14	-1.52e-25	-1.68e-17	4.52e-26
$60^\circ$	1.44e-03	1.02e-07	9.71e-11	6.43e-14	-1.55e-25	-1.68e-17	4.31e-26
$90^\circ$	1.44e-03	1.02e-07	9.69e-11	6.31e-14	-1.50e-25	-1.66e-17	4.38e-26
$120^\circ$	1.44e-03	1.02e-07	9.71e-11	6.37e-14	-1.52e-25	-1.67e-17	4.51e-26
$150^\circ$	1.44e-03	1.02e-07	9.71e-11	6.41e-14	-1.54e-25	-1.68e-17	4.44e-26
$180^\circ$	1.44e-03	1.02e-07	9.69e-11	6.31e-14	-1.50e-25	-1.66e-17	4.38e-26
$210^\circ$	1.44e-03	1.02e-07	9.71e-11	6.40e-14	-1.52e-25	-1.69e-17	4.68e-26
$240^\circ$	1.44e-03	1.02e-07	9.71e-11	6.41e-14	-1.54e-25	-1.68e-17	4.38e-26
$270^\circ$	1.44e-03	1.02e-07	9.69e-11	6.31e-14	-1.50e-25	-1.66e-17	4.38e-26
$300^\circ$	1.44e-03	1.02e-07	9.71e-11	6.38e-14	-1.52e-25	-1.68e-17	4.56e-26
$330^\circ$	1.44e-03	1.02e-07	9.71e-11	6.46e-14	-1.55e-25	-1.69e-17	4.44e-26

Tabela 4 – Momentos de Hu para rotação de imagens 2D binárias.

Rotação	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
$0^\circ$	6.76e-04	3.48e-08	1.18e-11	8.74e-15	-2.71e-27	-1.29e-18	7.47e-28
$30^\circ$	6.76e-04	3.49e-08	1.18e-11	8.71e-15	-2.70e-27	-1.28e-18	7.09e-28
$60^\circ$	6.76e-04	3.49e-08	1.18e-11	8.70e-15	-2.69e-27	-1.28e-18	7.10e-28
$90^\circ$	6.76e-04	3.48e-08	1.18e-11	8.74e-15	-2.71e-27	-1.29e-18	7.47e-28
$120^\circ$	6.76e-04	3.49e-08	1.18e-11	8.71e-15	-2.70e-27	-1.28e-18	7.04e-28
$150^\circ$	6.76e-04	3.49e-08	1.18e-11	8.70e-15	-2.70e-27	-1.28e-18	7.07e-28
$180^\circ$	6.76e-04	3.48e-08	1.18e-11	8.74e-15	-2.71e-27	-1.29e-18	7.47e-28
$210^\circ$	6.76e-04	3.49e-08	1.18e-11	8.72e-15	-2.71e-27	-1.28e-18	7.08e-28
$240^\circ$	6.76e-04	3.49e-08	1.18e-11	8.70e-15	-2.69e-27	-1.28e-18	7.06e-28
$270^\circ$	6.76e-04	3.48e-08	1.18e-11	8.74e-15	-2.71e-27	-1.29e-18	7.47e-28
$300^\circ$	6.76e-04	3.49e-08	1.18e-11	8.71e-15	-2.70e-27	-1.28e-18	7.14e-28
$330^\circ$	6.76e-04	3.49e-08	1.18e-11	8.73e-15	-2.71e-27	-1.28e-18	7.06e-28

Equação Contínua dos momentos invariantes 3D (SUK *et al.*, 2015):

$$\phi_{pqr} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q z^r f(x, y, z) dx dy dz \quad (2.11)$$

Equação Discreta dos momentos invariantes 3D:

$$\phi_{pqr} = \sum_x \sum_y \sum_z x^p y^q z^r f(x, y, z) \quad (2.12)$$

Equação Discreta e normalizada para translação:

$$\phi_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r f(x, y, z) \quad (2.13)$$

Assim como na equação 2.3, utiliza-se o centro de massa da nuvem de pontos  $(\bar{x}, \bar{y}, \bar{z})$ .

Momentos 3D invariantes:

$$\Phi_1^{3D} = \frac{(\phi_{220} + \phi_{020} + \phi_{002})}{\phi_{000}^{\frac{5}{3}}} \quad (2.14)$$

$$\Phi_2^{3D} = \frac{(\phi_{220}^2 + \phi_{020}^2 + \phi_{002}^2 + 2\phi_{110}^2 + 2\phi_{101}^2 + 2\phi_{011}^2)}{\phi_{000}^{\frac{10}{3}}} \quad (2.15)$$

$$\Phi_3^{3D} = (\phi_{200}^3 + 3\phi_{200}\phi_{110}^2 + 3\phi_{200}\phi_{101}^2 + 3\phi_{020}\phi_{110}^2 + \quad (2.16)$$

$$3\phi_{002}\phi_{101}^2 + \phi_{020}^3 + 3\phi_{020}\phi_{011}^2 + 3\phi_{002}\phi_{011}^2 + \quad (2.17)$$

$$\phi_{020}^3 + 6\phi_{110}\phi_{101}\phi_{011})/\phi_{000}^5 \quad (2.18)$$

### 2.2.3 Equalização da Densidade nas nuvens de pontos 3D

Densidade da nuvem de pontos pode ser definida como a quantidade de pontos em determinado volume tridimensional unitário. Cabe ressaltar que a nuvem de pontos de faces utilizada neste trabalho não foi densa, como as obtidas por tomografias de pulmão por exemplo, e sim nuvens de superfícies 3D. O que se deseja destacar é que pode haver uma diferença de densidade entre as faces quando a aquisição for feita por sensores distintos, e quando as amostras forem coletadas a distâncias suficientemente diferentes, de modo a variar essa densidade de pontos de maneira considerável.

Nesta seção serão apresentados resultados de experimentos que mostram o efeito da variação de densidades nos momentos de Hu e 3DIT.

O experimento consiste em tomar uma nuvem de pontos, submetê-la a uma variação de densidade utilizando o algoritmo Voxel Grid, implementado pela biblioteca PCL (Point Cloud Library), e posteriormente submeter a nuvem de pontos resultante ao cálculo dos momentos.

A Tabela 7 foi montada tomando-se os 7 momentos de HuXY para a nuvem de pontos original e a mesma nuvem submetida à variação de densidade de 1,0 mm até 4,0 mm com passo de 0,5 mm. A Tabela 8 faz o mesmo para os momentos 3DIT.

Os resultados mostram que nenhum dos momentos de Hu ou 3DITM preserva seus valores e que a variação pode ir de uma potência de 10 no momento  $\Psi_1$ , chegando a uma diferença de  $10^{10}$  como pode ser visto no momento  $\Psi_7$  entre a nuvem original e a com voxel grid de 4,0 mm.

Essa informação é importante pois 2 nuvens de pontos de um mesmo indivíduo, tomadas nas mesmas condições de emoção (Neutra, por exemplo), se tiverem uma variação

considerável de densidade, terão momentos de Hu e 3DITM completamente diferentes. No processo de identificação, essa diferença pode causar tanto uma interferência intra grupo, ou seja, considerar duas amostras do mesmo indivíduo como sendo de indivíduos diferentes; quanto uma interferência intergrupo, quando amostras de indivíduos diferentes podem ser consideradas como sendo de um mesmo indivíduo.

Uma outra situação pode ser analisada quando se mescla as bases de dados. Imagine que seja possível definir um experimento onde as nuvens de pontos da base Bosphorus e FRGC v2.0 sejam utilizadas em conjunto. Mesmo que fosse possível garantir que, dentro de cada uma das bases, as nuvens de pontos de todas as amostras teriam a mesma densidade de pontos, isso não necessariamente aconteceria entre as amostras de bases diferentes.

É importante salientar que alguns métodos não são aplicáveis a todos os tipos de nuvens de pontos. Berretti *et al.* (2014) cita que o seu método, apesar dos resultados estarem no topo do estado da arte para a base Bosphorus, não é aplicável a nuvens de pontos de face de sensores de baixa densidade como o Kinect. Os últimos trabalhos publicados utilizam principalmente as seguintes 3 bases de dados: Bosphorus, FRGC v2.0 e BU-3DFE, que possuem amostras com resoluções da ordem de, 0,3 mm , XX mm e YY mm, respectivamente.

Lembrando que os sensores utilizados pelas bases de dados Bosphorus e FRGC v2.0 possuem resolução da ordem de 0,3 mm. Alguns sensores com essa precisão custam mais de 20.000 dólares, veja GoSCAN20©da CREAFORM©, mostrando que realizar uma aquisição com esse nível de resolução ainda é uma tarefa cara. A câmera Intel©RealSense©SR300 tem resolução de 1 mm para objetos a 1 m de distância, custando menos de 200 dólares.

A partir desta análise é possível concluir que a equalização da densidade é um processo importante para que se possa utilizar os momentos de Hu e 3DITM para identificação de indivíduos, além de que a resolução das nuvens de pontos pode afetar a eficácia dos algoritmos de reconhecimento de face.

### **2.3 Alinhamento de objetos 3D**

Nesta seção serão descritos os procedimentos para alinhamento da face em relação ao plano XY de forma a manter as faces dentro de um mesmo condicionamento de orientação.

A Figura 9 mostra duas nuvens de pontos desalinhadas, uma delas é a nuvem objetivo, a da direita, e a outra a nuvem de pontos fonte. O objetivo é encontrar o melhor método que consiga registrar as duas nuvens de pontos com a melhor acurácia e com o mínimo de esforço

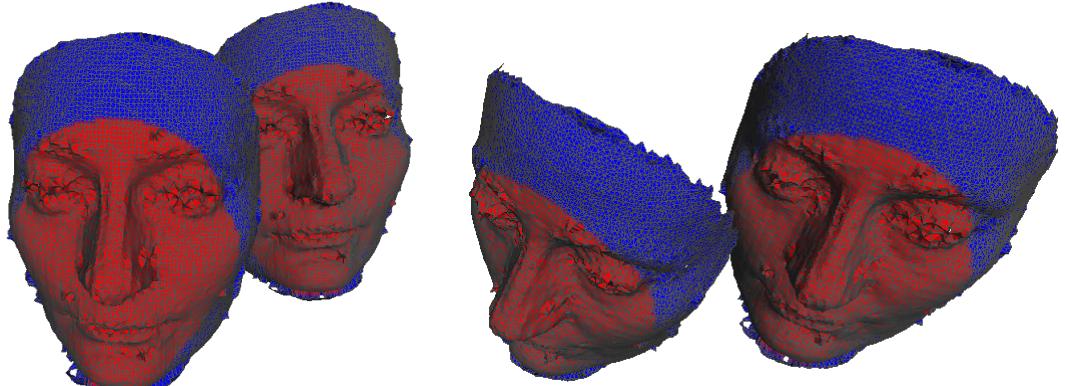
Tabela 5 – Momentos HuXY para variação de densidade de pontos com base no voxel grid.

Voxel Grid (mm)	voxels (#)	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
0,0	46.321	5,89e-04	3,23e-08	1,14e-12	6,55e-15	3,03e-28	8,59e-19	-4,76e-28
1,0	31.468	9,04e-04	6,19e-08	4,06e-12	1,29e-13	-8,34e-26	3,12e-17	-4,10e-26
1,5	18.090	1,62e-03	1,55e-07	2,97e-11	1,44e-12	-9,30e-24	5,67e-16	-1,47e-24
2,0	11.519	2,58e-03	3,34e-07	1,17e-10	1,36e-11	-4,57e-22	7,62e-15	-2,85e-22
2,5	7.845	3,85e-03	7,50e-07	4,21e-10	4,30e-11	-5,72e-21	3,72e-14	-9,52e-22
3,0	5.682	5,36e-03	1,36e-06	1,38e-09	1,03e-10	-3,26e-20	1,18e-13	-2,15e-20
3,5	4.283	7,23e-03	2,35e-06	2,63e-09	4,00e-10	-3,88e-19	6,13e-13	-1,33e-19
4,0	3.360	9,16e-03	3,63e-06	5,64e-09	1,06e-09	-1,96e-18	1,95e-12	-1,68e-18

Tabela 6 – Momentos 3DIT para variação de densidade de pontos com base no voxel grid.

Voxel Grid (mm)	voxels (#)	$\Phi_1^{3D}$	$\Phi_2^{3D}$	$\Phi_3^{3D}$
0,0	46.321	7,77e-01	4,13e-01	2,90e-01
1,0	31.468	9,39e-01	5,99e-01	4,67e-01
1,5	18.090	1,35e+00	1,23e+00	1,36e+00
2,0	11.519	1,83e+00	2,23e+00	3,33e+00
2,5	7.845	2,36e+00	3,69e+00	7,08e+00
3,0	5.682	2,92e+00	5,67e+00	1,35e+01
3,5	4.283	3,53e+00	8,20e+00	2,34e+01
4,0	3.360	4,15e+00	1,14e+01	3,82e+01

computacional.



(a) Perspectiva 0.

(b) Perspectiva 1.

Figura 7 – Duas perspectivas de nuvens de pontos rotacionadas e transladadas entre si.

Alguns trabalhos de reconhecimento facial demonstraram a importância da orientação da face na extração de características (Berretti *et al.* (2014), Elaiwat *et al.* (2015) e Erdoganmus e Marcel (2013)).

Nesse contexto, foram avaliadas 3 técnicas para alinhamento da face: RANSAC-Plane (RANdom SAmple Consensus), Two-Sided Jacobi SVD (Singular Value Decomposition) e ICP (Interactive Closest Point). Detalhes de cada um destes métodos serão apresentados a seguir.

### 2.3.1 RANSAC-Plane

Algoritmo RANSAC (FISCHLER; BOLLES, 1981) é largamente utilizado para encontrar nuvens de pontos a partir de um modelo pré-estabelecido. Este método é muito utilizado também para rejeitar falsas correspondências (*false matching*) (GAO *et al.*, 2018), mas devido à sua amostragem randômica, o algoritmo pode não resultar nem na melhor aproximação e nem em um tempo razoável, a depender da nuvem de pontos a ser analisada.

Como o algoritmo é pensado para encontrar coincidências entre modelos, uma das implementações da biblioteca *Point Cloud Library* (PCL), dentre outras, usa alguns modelos preexistentes como planos, cilindros, cones, dentre outros. Uma forma de utilizar o RANSAC para realizar alinhamento de uma nuvem de pontos é tentar encontrar o melhor plano possível que possa corresponder à ela, e utilizar a normal do plano como direção de orientação.

Schnabel *et al.* (2007) foi quem primeiro adaptou o algoritmo RANSAC para detecção de planos, encontrando uma maneira mais rápida e precisa desde que os parâmetros estivessem adequadamente configurados. Schnabel *et al.* (2007) definiu que para um plano, os parâmetros  $(p_1, p_2, p_3)$  constituem o conjunto mínimo para definir seu vetor normal. E que o melhor plano é encontrado quando a diferença entre as normais entre dois planos consecutivos da iteração possuem desvio menor que um determinado ângulo  $\alpha$  predefinido.

Com a informação da normal do plano, que é retirada diretamente da equação do próprio plano, pode-se alinhar qualquer nuvem para qualquer direção. Por utilizar o plano como modelo, esta técnica é chamada de RANSAC-Plane. O algoritmo encontra os parâmetros  $p_1, p_2, p_3$  e  $p_4$  do plano e a partir deles é encontrada a normal ao plano apresentada na Equação 2.19.

$$p_1\bar{x} + p_2\bar{y} + p_3\bar{z} = p_4 \quad (2.19)$$

Por conta de ser um algoritmo iterativo e ser muito sensível ao número de pontos a serem analisados, várias propostas de otimização deste método são propostas e algumas podem ser vistas em Gao *et al.* (2018), Fujiwara *et al.* (2013) e Alehdaghi *et al.* (2015).

### 2.3.2 Matriz Jacobiana

Uma das maneiras de realizar o alinhamento de duas nuvens de pontos é encontrar o vetor normal das duas nuvens, uma fonte e outra objetivo, é encontrar a transformação que leva normal da nuvem fonte para a mesma direção da nuvem objetivo. Alguns trabalhos utilizam esta transformação

Uma das formas de encontrar estas normais é utilizar a matriz Jacobiana. Uma complicação é que esta matriz utiliza diferenças finitas e que pode levar a um grande esforço computacional, mas utilizando o algoritmo SVD(Single Value Decomposition), em sua versão que implementa o cálculo pelos dois lados da equação (Two-Sided Jacobi SVD) é possível realizar o cálculo desta matriz, e desta forma, encontrar os vetores normais(TWIGG; KAVCIC-ALESIC, 2010).

A decomposição é representada pela equação 2.20, onde  $U$  é uma matrix unitária  $n \times n$ ,  $S$  é uma matriz retangular diagonal  $m \times n$  e  $V$  é uma matrix unitária  $n \times n$ .

$$A = USV^* \quad (2.20)$$

Os ângulos da normal à nuvem de pontos são extraídos da matriz  $V^*$ .

Este método pode ser utilizado em cascata com outros métodos de alinhamento, como é feito na implementação do ICP na biblioteca PCL, que utiliza o algoritmo Two-Sided Jacobi SVD para realizar o que ele chama de ajuste mais grosso de alinhamento. Isto é feito tomando-se as normais das duas nuvens, fonte e objetivo, e alinhando-se a nuvem de pontos fonte na direção da normal da nuvem de pontos objetivo. A partir desta primeira reorientação o algoritmo ICP é utilizado.

Muitos trabalhos concentram-se em otimizar o cálculo da decomposição de forma a minimizar o seu tempo como pode ser visto nos trabalhos de Wang e Zambreno (2014), (GUENTHER *et al.*, 2016), Bečka *et al.* (2010) e Bečka *et al.* (2010).

### 2.3.3 ICP

Algoritmo ICP, introduzido (MUDAR ESSA PALAVRA!) por Chen e Medioni (1991) e Besl e McKay (1992), funciona assumindo-se uma nuvem de pontos fonte  $P$  e uma nuvem de pontos objetivo  $Q$ . Para cada um dos pontos da nuvem  $P$  são calculadas as distâncias, duas a duas, para os pontos da nuvem  $Q$  e busca-se o ponto que possua a menor distância Eucli-

ana para fazer a correspondência. A partir desta correspondência é realizada uma transformação mínima possível, com uma função objetivo demonstrada pela equação 2.21. Esta transformação inclui a matriz de rotação  $R$  e o vetor de deslocamento  $T$ , e é aplicada a nuvem de pontos  $P$  até que um determinado critério de convergência seja atingido.

Determinar a correta, ou melhor, a correspondência entre os pontos das nuvens  $P$  e  $Q$  é a chave do processo do algoritmo ICP. A acurácia da correspondência está diretamente ligada à precisão e velocidade de convergência do registro.

$$f(R, T) = \frac{1}{N} \sum_{i=1}^N \|q_i - (Rp_i + T)\|^2 \quad (2.21)$$

Considerando-se  $P = \{p_i | i = 1, 2, 3, \dots\}$  e  $Q = \{q_i | i = 1, 2, 3, \dots\}$ , Guan *et al.* (2018) resume o algoritmos aos seguintes passos

**Passo 1** Defina  $k = 0$  e escolha um determinador limiar  $\delta$

**Passo 2** De acordo com a distância euclidiana, encontre o ponto  $q_i$  da nuvem objetivo mais próximo do ponto  $p_i$  da nuvem fonte.

**Passo 3** De acordo com a correspondência encontrada para os pontos das nuvens  $P$  e  $Q$ , as matrizes de rotação e translação, respectivamente,  $R$  e  $T$  são obtidas pelo método dos 4 elementos ou algoritmo Single Value Decomposition (SVD).

**Passo 4** Nova nuvem de pontos fonte é encontrada sendo,  $P^j = RP^{j-1} + T$ , onde  $j$  é a iteração e a nuvem de pontos fonte original  $P = P^0$ .

**Passo 5** Se o erro da iteração  $j$  com a iteração anterior satisfizer  $f_j - f_{j-1} < \delta$  o algoritmo encerra, caso contrário, executa os passos de 2 a 5.

Originalmente, o algoritmo *Interactive Closest Point* (ICP) foi pensado para superfícies rígidas e não deformáveis, pois desta forma, pode-se considerar o que diz Besl e McKay (1992) que garante que o algoritmo ICP sempre converge, monotonicamente, para o mínimo local mais próximo utilizando-se a distância euclidiana.

O uso em superfícies deformáveis pode não garantir essa convergência, e neste trabalho utiliza-se nuvens de pontos deformáveis, e em algumas situações com alto grau de deformação.

### 3 MÉTODO DE RECONHECIMENTO DE FACES 3D SEGMENTADAS EM REGIÕES TRIAXIAIS

Neste capítulo é detalhado o método proposto para reconhecimento de faces 3D. São discutidas as diversas etapas do processo, que são apresentadas na sequência: (i) equalização da densidade da nuvem ; (ii) adaptação dos momentos de Hu e 3DIT; (iii) extração da área útil da face usando como referência a ponta do nariz; (iv) realização dos cortes triaxiais na face extraída, seguida pela correção de pose; (v) uso de diferentes classificadores, discutindo-se a aplicabilidade de cada um para cada cenário.

#### 3.1 Equalização da Densidade nas nuvens de pontos 3D

A densidade da nuvem de pontos pode ser definida como a quantidade de pontos em determinado volume tridimensional unitário. Cabe ressaltar que a nuvem de pontos de faces utilizadas neste trabalho não são nuvens densas, como as obtidas por tomografias de pulmão, por exemplo, e sim nuvens de superfícies 3D. É preciso considerar que pode haver diferença de densidade entre as faces oriundas de aquisições por sensores distintos ou, ainda, devido ao fato de as amostras terem sido coletadas a distâncias suficientemente diferentes, de modo que a variação da densidade de pontos comprometa o cálculo dos momentos. Assim, como primeira etapa do método de identificação proposto, a densidade é equalizada.

Estudos foram realizados e os resultados de experimentos mostram o efeito da variação de densidades nos momentos de Hu e 3DIT.

O experimento consiste em tomar uma nuvem de pontos, submetê-la a uma variação de densidade utilizando o algoritmo *Voxel Grid* implementado pela biblioteca PCL (*Point Cloud Library*), apresentado por Rusu e Cousins (RUSU; COUSINS, 2011). Em seguida, a nuvem de pontos resultante é submetida ao cálculo dos momentos de Hu e 3DIT. O efeito da diminuição da densidade de pontos pode ser visto na Figura 8, em que uma mesma nuvem de pontos é subamostrada com vários valores de *voxel grid*.

Na Figura 8 é possível perceber que, a partir de um *voxel grid* maior que 4 milímetros, detalhes de olhos e boca começam a desaparecer, e com valores acima de 10 milímetros o rosto perde seus detalhes e passa a ser apenas uma máscara sem qualquer informação que permita identificar o indivíduo que a gerou.

A Tabela 7 foi montada tomando-se os 7 momentos de HuXY para a nuvem de pontos original e a mesma nuvem submetida à variação de densidade de 1,0 mm até 4,0 mm

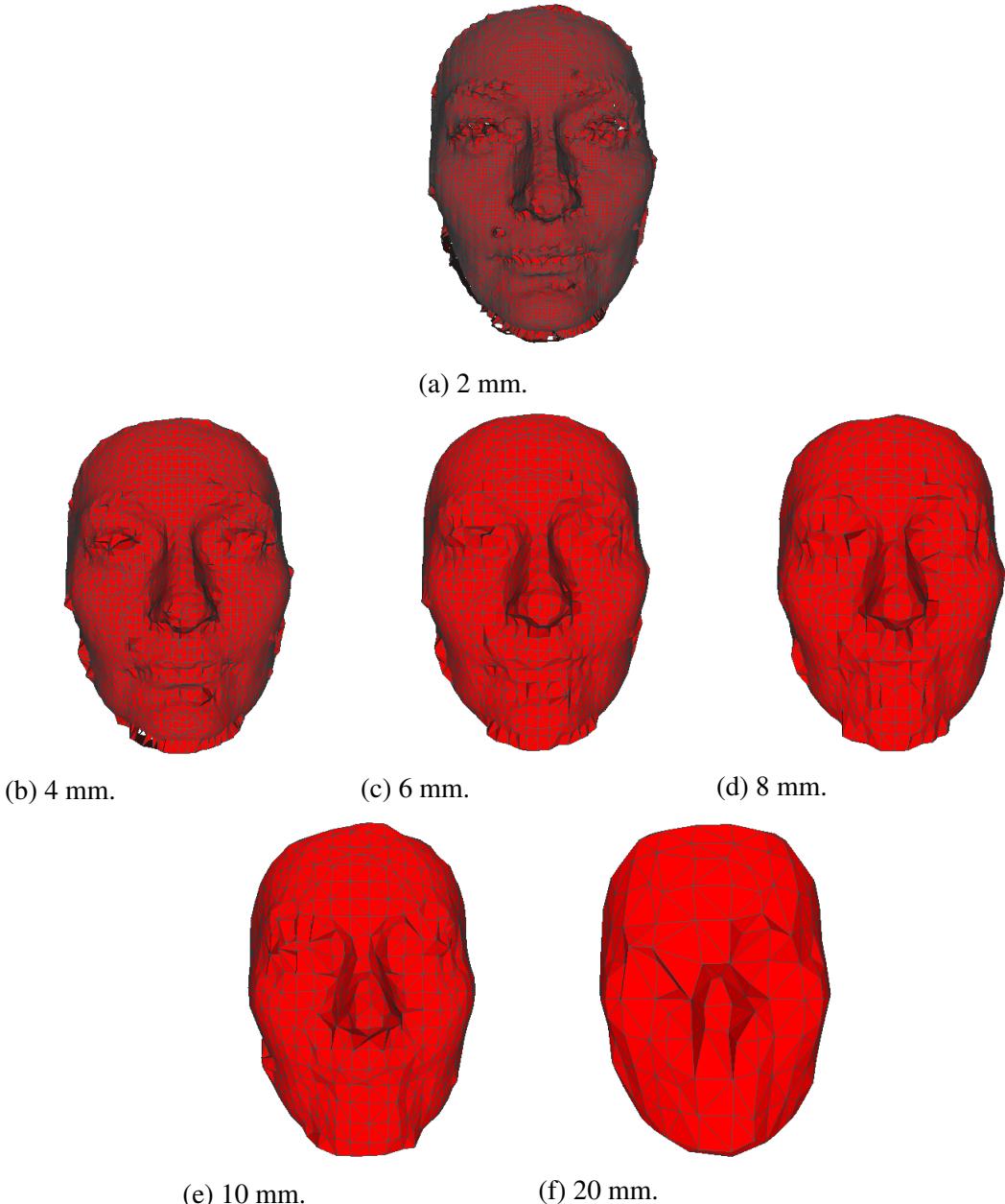


Figura 8 – Nuvem de pontos neutra submetida a uma subamostragem através do algoritmo do *Voxel Grid*.

com passo de 0,5 mm. A Tabela 8 faz o mesmo para os momentos 3DIT.

Os resultados mostram que nenhum dos momentos de Hu ou 3DITM preservam seus valores e, além disso, pode-se variar de uma potência de 10 no momento  $\Psi_1$ , a  $10^{10}$ , como pode ser visto no momento  $\Psi_7$  entre a nuvem original e a com *voxel grid* de 4,0 mm.

Isto mostra que duas nuvens de pontos de um mesmo indivíduo, tomadas nas mesmas condições de emoção (Neutra por exemplo), se tiverem uma variação considerável de densidade, terão momentos de Hu e 3DITM completamente diferentes. Essa diferença pode causar, no processo de identificação, tanto uma interferência intragrupo, ou seja, considerar duas amostras

Tabela 7 – Momentos HuXY para variação de densidade de pontos com base no voxel grid.

Voxel Grid (mm)	voxels (#)	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
0.0	46321	5.89e-04	3.23e-08	1.14e-12	6.55e-15	3.03e-28	8.59e-19	-4.76e-28
1.0	31468	9.04e-04	6.19e-08	4.06e-12	1.29e-13	-8.34e-26	3.12e-17	-4.10e-26
1.5	18090	1.62e-03	1.55e-07	2.97e-11	1.44e-12	-9.30e-24	5.67e-16	-1.47e-24
2.0	11519	2.58e-03	3.34e-07	1.17e-10	1.36e-11	-4.57e-22	7.62e-15	-2.85e-22
2.5	7845	3.85e-03	7.50e-07	4.21e-10	4.30e-11	-5.72e-21	3.72e-14	-9.52e-22
3.0	5682	5.36e-03	1.36e-06	1.38e-09	1.03e-10	-3.26e-20	1.18e-13	-2.15e-20
3.5	4283	7.23e-03	2.35e-06	2.63e-09	4.00e-10	-3.88e-19	6.13e-13	-1.33e-19
4.0	3360	9.16e-03	3.63e-06	5.64e-09	1.06e-09	-1.96e-18	1.95e-12	-1.68e-18

Tabela 8 – Momentos 3DIT para variação de densidade de pontos baseado no voxel grid.

Voxel Grid (mm)	voxels (#)	$\Phi_1^{3D}$	$\Phi_2^{3D}$	$\Phi_3^{3D}$
0.0	46321	7.77e-01	4.13e-01	2.90e-01
1.0	31468	9.39e-01	5.99e-01	4.67e-01
1.5	18090	1.35e+00	1.23e+00	1.36e+00
2.0	11519	1.83e+00	2.23e+00	3.33e+00
2.5	7845	2.36e+00	3.69e+00	7.08e+00
3.0	5682	2.92e+00	5.67e+00	1.35e+01
3.5	4283	3.53e+00	8.20e+00	2.34e+01
4.0	3360	4.15e+00	1.14e+01	3.82e+01

do mesmo indivíduo como sendo de indivíduos diferentes, quanto uma interferência intergrupo, quando amostras de indivíduos diferentes podem ser consideradas como sendo de um mesmo indivíduo.

É importante salientar que alguns métodos de identificação de faces 3D não são aplicáveis a todos os tipos de nuvens de pontos, sendo comprometidos pelo uso de nuvens com baixa densidade. Berretti *et al.* (2014), cujos resultados estavam no topo do estado da arte para a base Bosphorus no momento em que esta tese foi escrita, alegam que seu método não é aplicável a nuvens de pontos de face escaneadas por sensores de baixa resolução, como o Kinect. Os últimos trabalhos publicados utilizam principalmente as bases de dados Bosphorus, FRGC v2.0 e BU-3DFE, que possuem amostras com resoluções da ordem de 0,3 milímetros.

Alguns sensores com resoluções na ordem de 0,3 mm de precisão possuem alto custo. GoSCAN20©, da CREAFORM©, é um exemplo de quão dispendiosa é a aquisição com esse nível de resolução. Em contrapartida, a câmera Intel©RealSense©SR300 possui resolução de 1,0 milímetro para objetos a 1,0 metro de distância, custando menos de 200 dólares.

Conclui-se que há várias razões para realizar a normalização de densidade como

a primeira etapa de pré-processamento. Uma delas é diminuir o número de pontos em uma nuvem, tornando o processamento nos próximos passos mais rápido. A outra razão, e não menos importante, refere-se ao tipo de descritores que devem ser extraídos, uma vez que eles podem ser sensíveis às variações de densidade intra e inter-indivíduo. A ideia é diminuir a densidade das nuvens de pontos até que todas tenham a mesma densidade. Idealmente, um bom aparelho de aquisição deve fornecer imagens com quase a mesma densidade de ponto, no entanto, isso não é observado na prática. Temos visto amostras com nuvens mais densas, o que provavelmente sugere que o indivíduo se aproxime da câmera em algumas sessões. Essas variações de densidade podem ocorrer tanto para amostras do mesmo assunto quanto para amostras distintas. Dependendo do tipo de descritores empregados, o processo de reconhecimento pode ser insensível a essa variação ou ser significativamente afetado por ela. É importante notar que a diminuição da densidade não pode ser extrema, de modo que os rostos se descaracterizem, mas também não pode ser mínima de forma a não beneficiar o algoritmo de alinhamento. A técnica proposta neste trabalho utiliza o algoritmo de grade voxel 3D para realizar a normalização da densidade com uma folha de voxel de 2,0 milímetros para todas as três dimensões.

Finalmente, antes de passar para a próxima etapa, é realizada a remoção de outliers. Esta é uma etapa comum entre os métodos que realizam a análise de superfície, pois podem causar picos que descharacterizam a superfície facial. No nosso caso, eles podem adicionar ruído aos descritores de face. Embora sejam predominantes em imagens oclusas ou auto-oclusas, em geral, os outliers de qualquer origem podem impedir a detecção de pontos importantes, como a ponta do nariz, assunto discutido na próxima seção.

### **3.2 Adaptações Momentos de Hu e 3DITM**

Nesta seção serão discutidas as adaptações dos momentos de Hu e 3DITM para serem utilizados em nuvens de pontos. Importante destacar que enquanto os momentos 3DITM são intrínsecos a estes, os momentos de Hu são utilizados para imagens 2D.

Primeiramente será demonstrado que se pode utilizar os conceitos dos momentos de Hu para nuvens de pontos e posteriormente as variações propostas neste trabalho.

### 3.2.1 Momentos de Hu para nuvens de pontos

Duas situações foram definidas para aplicação dos momentos de Hu para nuvens de pontos, a primeira foi a função  $f(x, y)$  da equação 2.3 e a segunda foi como fazer a rotação de forma que a propriedade de invariância fosse preservada.

Uma vez que a nuvem de pontos é composta pela tripla  $(x, y, z)$ , considerou-se, para este caso, que as variáveis  $x$  e  $y$  eram dependentes e que  $z$  era independente.

A primeira situação relacionada à função  $f(x, y)$  é de fácil resolução, bastando para tanto substituir a função pelo valor da variável independente  $z$ , neste caso, a equação 2.3 passaria para a equação 2.19:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j z \quad (3.1)$$

Da mesma forma que há um efeito relacionado à translação dentro do plano  $XY$ , pode haver uma translação na direção  $\vec{z}$ , o que pode ser solucionado com a normalização em relação à coordenada  $\bar{z}$  do centro de massa da nuvem de pontos. Então, a equação 3.2 passa a ser:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j (z - \bar{z}) \quad (3.2)$$

De fato, não há uma limitação relacionada à função  $f(x, y)$  dentro das condições relacionadas aos momentos espaciais 2D. Desta forma, pode-se utilizar uma função distância, como a função da coordenada  $(x, y, z)$ , até o centro de massa  $(\bar{x}, \bar{y}, \bar{z})$ . Considerando esta última alternativa, a equação 2.3 torna-se:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2} \quad (3.3)$$

Embora as variações apresentadas nas equações 3.2 e 3.3 tenham uma origem lógica e que pode ser justificada pela própria construção das equações de momentos, a função  $f(x, y)$  virtualmente pode ser qualquer uma, com a condição que ela não varie quando forem realizadas

operações de rotação e translação. Seguem algumas opções para a função:

$$f(x, y) = \begin{cases} 1. \\ (z - \bar{z}). \\ |z - \bar{z}|. \\ \sqrt[2]{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2}. \end{cases} \quad (3.4)$$

Considerando que a rotação seja realizada apenas no plano  $XY$ , a equação de transformação é dada pelo produto da equação 3.5, de forma que as apresentações da equação 3.4 permanecem as mesmas para o conjunto de pontos onde será executada a operação.

Abaixo segue a equação de rotação de uma nuvem de pontos 3D no plano  $XY$ :

$$\begin{pmatrix} x' \\ y' \\ z \\ 1 \end{pmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 & 0 \\ \sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.5)$$

Como já foi descrito nesta seção, os momentos mantêm suas características de invariância à rotação e translação no plano  $XY$ , o que faz necessário considerar que as equações dos momentos de Hu, para nuvens 3D, sejam referenciadas como  $Hu_{XY}$ .

Desta forma, a equação 2.2 pode ser reescrita como:

$$\mu_{ij}^{XY} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j \sqrt[2]{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2} \quad (3.6)$$

### 3.2.2 Momentos de Hu para múltiplos planos

Uma vez considerada a adaptação dos momentos de Hu apresentados na equação 3.6, considerando o plano  $XY$ , é necessário realizar a mesma consideração para os outros dois planos ortogonais  $XZ$  e  $YZ$ . Por similaridade:

$$\mu_{ij}^{XZ} = \sum_x \sum_z (x - \bar{x})^i (z - \bar{z})^j f(x, z) \quad (3.7)$$

$$\mu_{ij}^{YZ} = \sum_y \sum_z (y - \bar{y})^i (z - \bar{z})^j f(y, z) \quad (3.8)$$

Ainda por similaridade, as funções  $f(x, z)$  e  $f(y, z)$  podem ser escritas como:

$$f(x, z) = \begin{cases} 1. \\ (x - \bar{x}). \\ |x - \bar{x}|. \\ \sqrt[2]{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2}. \end{cases} \quad (3.9)$$

$$f(y, z) = \begin{cases} 1. \\ (y - \bar{y}). \\ |y - \bar{y}|. \\ \sqrt[2]{(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2}. \end{cases} \quad (3.10)$$

É importante considerar que cada uma dessas variações mantém suas propriedades dentro das operações de rotação e translação no plano ao qual estão associados  $XY$ ,  $XZ$  ou  $YZ$ .

As equações 3.11 e 3.12 representam a rotação nos planos  $XZ$  e  $YZ$ , respectivamente:

$$\begin{pmatrix} x' \\ y \\ z' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \alpha & \cos \alpha & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.11)$$

$$\begin{pmatrix} x \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta & 0 \\ 0 & \sin \beta & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (3.12)$$

### 3.2.3 Experimentos com Momentos de Hu e Rotações em Múltiplos Planos

Uma vez que foram realizados experimentos para demonstrar a invariância à rotação dos momentos de Hu em imagens 2D, esta seção apresenta experimentos realizados para

demonstrar que é possível utilizar as adaptações apresentadas nas seções anteriores de acordo com o plano ao qual estão associados.

A Figura 9 mostra uma nuvem de pontos em sua postura original e com várias rotações em torno do eixo  $\vec{z}$ , colorida artificialmente para melhorar a visualização. A Figura 10 mostra a mesma nuvem de pontos rotacionada em torno dos três eixos. É importante compreender o efeito destas rotações, apenas em um eixo e em mais de um eixo, para que seja possível analisar o conjunto de experimentos a seguir.

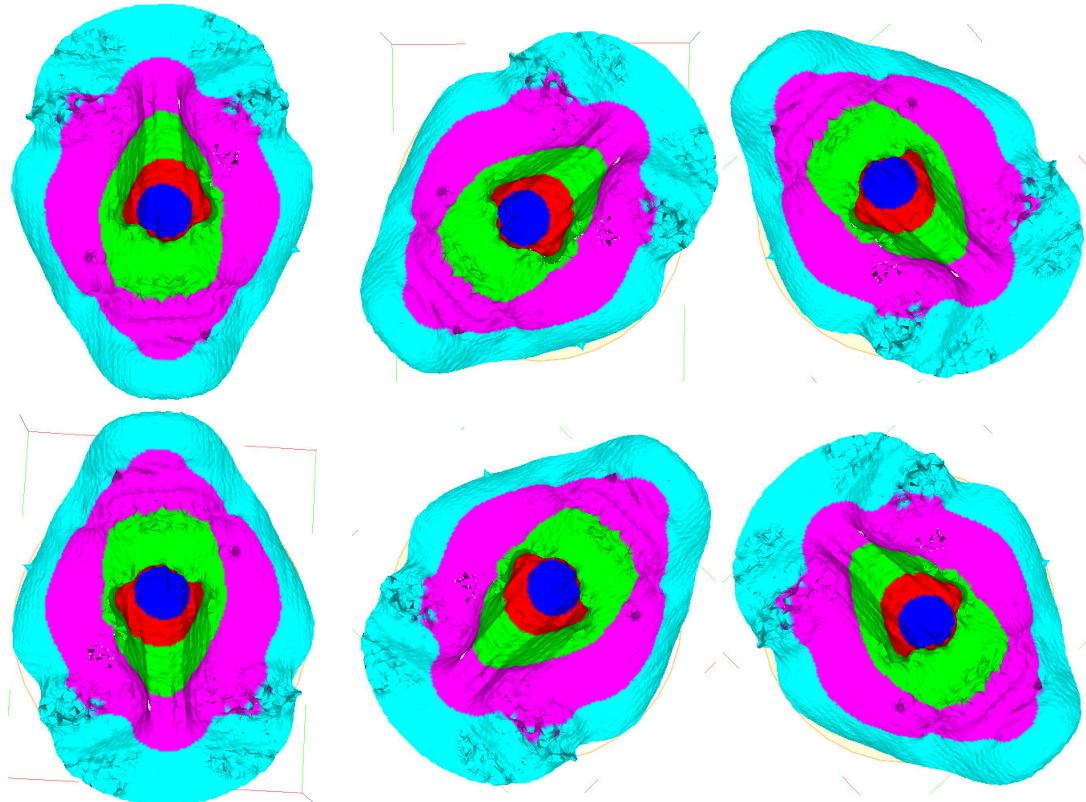


Figura 9 – Nuvem de pontos 3D da face rotacionada em torno no eixo z com diferentes ângulos. Nuvem de pontos da base de dados Bosphorus bs003\_N\_N\_0.

O experimento basicamente calcula os momentos de Hu ( $HuXY$ ,  $HuXZ$  e  $HuYZ$ ) para várias rotações nos três eixos e verifica em que condições eles garantem a característica de invariância à rotação.

As tabelas 9, 11 e 10 mostram que os momentos  $HuXY$ ,  $HuXZ$  e  $HuYZ$  são invariantes, respectivamente, às rotações que ocorram apenas nos planos  $XY$ ,  $XZ$  e  $YZ$ .

Já a tabela 12 mostra que os momentos 3DITM são invariantes independente do eixo em que a nuvem é rotacionado.

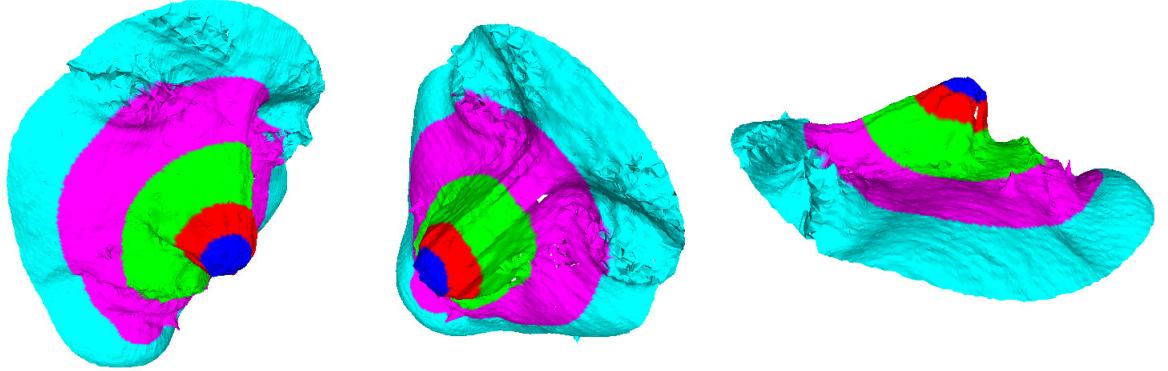


Figura 10 – Nuvem de pontos 3D da face rotacionada nos três eixos com diferentes ângulos.  
Nuvem de pontos da base de dados Bosphorus bs003\_N\_N\_0.

Tabela 9 – Momentos HuXY para rotação de nuvens de pontos em torno do plano XY.

$\hat{X}$	$\hat{Y}$	$\hat{Z}$	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
0°	0°	0°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	30°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	60°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	90°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	120°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	150°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	180°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	210°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	240°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	270°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	300°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10
0°	0°	330°	-1,58e+00	4,05e-02	1,98e-05	6,88e-05	2,47e-09	-1,28e-05	5,87e-10

Tabela 10 – Momentos HuYZ para rotação de nuvens de pontos em torno do plano YZ.

$\hat{X}$	$\hat{Y}$	$\hat{Z}$	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
0°	0°	0°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	30°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	60°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	90°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	120°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	150°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	180°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	210°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	240°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	270°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	300°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22
0°	0°	330°	2,24e-03	3,20e-07	9,88e-11	3,78e-11	2,27e-21	2,04e-14	4,19e-22

### 3.3 Definição da Área Útil da Face

Nesta etapa busca-se, inicialmente, encontrar um ponto a partir do qual é possível definir a área útil da face. Esta área define a região de interesse sobre a qual são extraídos

Tabela 11 – Momentos HuXZ para rotação de nuvens de pontos em torno do plano XZ.

$\hat{X}$	$\hat{Y}$	$\hat{Z}$	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$	$\Psi_7$
$0^\circ$	$0^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$30^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$60^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$90^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$120^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$150^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$180^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$210^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$240^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$270^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$300^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24
$0^\circ$	$330^\circ$	$0^\circ$	2,57e-03	1,02e-06	6,21e-11	5,47e-13	1,23e-24	2,46e-16	2,94e-24

Tabela 12 – Momentos 3DITM para rotação de nuvens de pontos.

$\hat{X}$	$\hat{Y}$	$\hat{Z}$	$\Phi_1^{3D}$	$\Phi_2^{3D}$	$\Phi_3^{3D}$
$0^\circ$	$0^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$0^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$0^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$120^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$120^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$120^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$240^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$240^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$0^\circ$	$240^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$0^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$0^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$0^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$120^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$120^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$120^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$240^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$240^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$120^\circ$	$240^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$0^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$0^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$0^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$120^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$120^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$120^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$240^\circ$	$0^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$240^\circ$	$120^\circ$	1,91e+00	2,83e+00	4,70e+00
$240^\circ$	$240^\circ$	$240^\circ$	1,91e+00	2,83e+00	4,70e+00

os segmentos da nuvem que, por sua vez, fornecem as propriedades capazes de identificar o indivíduo. A área útil preserva as principais estruturas da face, como nariz, boca, testa, olhos, bochechas e, ao mesmo tempo, exclui outras estruturas do indivíduo que comprometem a identificação, como os cabelos, as orelhas, os ombros, o pescoço, o tórax, além de outros objetos diversos eventualmente capturados no cenário em torno do indivíduo.

Não há um padrão adotado pelas bases de dados com faces 3D. Na base Bosphorus, por exemplo, a nuvem de pontos relativa à região da face encontra-se previamente extraída, de forma a simplificar a construção de algoritmos de identificação de indivíduos ou de expressões. Na base de dados 3DMAD, por outro lado, a nuvem de pontos relativa à face está integrada à parte superior do corpo do indivíduo, acima do abdômen, bem como com o restante do cenário de captura. Já a base FRGC v2.0 entrega a nuvem de pontos com parte do corpo, incluindo pescoço e ombros, mas sem detalhes do cenário. Portanto, para cada base de dados deve-se realizar um pré-processamento para extração da nuvem de pontos relativa à região sobre a qual será aplicado o algoritmo de identificação.

### **3.3.1 Localização da Ponta do Nariz**

O nariz é o único ponto-chave que precisa ser previamente localizado para execução do método proposto. Alguns trabalhos, como o de Elaiwat *et al.* (2015), também utilizam a ponta do nariz como ponto chave. A questão é como encontrar a ponta do nariz e, para respondê-la, há algumas considerações a fazer.

A primeira delas é saber quais imagens estão disponíveis para realizar esta localização. No caso da base de dados da face fornecer apenas a nuvem de pontos 3D, recentes estudos realizados por Emambakhsh e Evans (2017) demonstram que a localização do nariz não consiste em um grande esforço computacional e pode ser localizado inclusive quando a face está submetida a rotações.

Caso a base de dados de faces forneça tanto as nuvens de pontos quanto as imagens RGB, e se for possível realizar o pareamento entre o voxel da imagem 3D e o pixel da imagem RGB, é possível realizar a localização da região do nariz através do algoritmo Haar adaptado para localização de face e nariz. Algoritmos para imagens 2D são muito robustos e foram aprimorados por décadas, portanto, utilizá-los integrados a técnicas 3D pode ser um atalho enquanto algumas técnicas 3D são aprimoradas, uma vez que os dispositivos 3D estão se tornando acessíveis e de uso comum. Os algoritmos Haar, adaptados para localizar faces e nariz, suportam recursos de

wavelets (VIOLA; JONES, 2001) e LBP (*Local Binary Patterns*) (LIAO *et al.*, 2007). Utilizar o método para localizar faces, e dentro da ROI (*Region of Interest*) da face buscar o nariz é uma forma de maximizar a taxa de acerto desta região.

Uma vez mapeada a região da nuvem de pontos onde o nariz está localizado, basta localizar o ponto mais extremo desta, máximo ou mínimo de acordo com o sistema de orientação ser na direção  $\vec{z}$  ou  $-\vec{z}$ , como mostra a Figura 16. Nesse aspecto, os *outliers* próximos ao sensor podem competir com a ponta do nariz, mas um motivo pelo qual devem ser previamente removidos a esta etapa.

### 3.3.2 Recorte da Face

A extração da área útil da face é realizada de maneira a garantir correspondência entre todas as amostras do mesmo indivíduo. Garante-se também, assim, a correspondência entre os cortes Frontais, Sagitais e Transversais que são posteriormente realizados sobre essa área útil.

Na base de dados Bosphorus, em que cada nuvem contém apenas a parte recortada do rosto onde as *Action Units* estão presentes (exceto para os casos de rotações, que não são consideradas neste trabalho) a área útil é previamente extraída. Entretanto, algumas diferenças nas partes extremas dessas regiões, mesmo para amostras do mesmo indivíduo em estado neutro, foram observadas. Por isso, no caso da Bosphorus, é necessário limitar ainda mais a região da face que será utilizada na extração de atributos e, para isso, utiliza-se a ponta do nariz e determina-se a região composta pelos pontos que estão a determinada distância dela. Neste trabalho utilizou-se a distância de 80 mm para todas as extrações de região do rosto a partir da ponta do nariz.

As Figuras 11, 13 e 14 correspondem às bases de dados 3DMAD, FRGC v2.0 e Bosphorus, respectivamente. Em cada uma delas é destacada em vermelho a área útil da face extraída, podendo-se perceber as situações distintas de como as nuvens de pontos são fornecidas em cada bases de dados.

A Figura 11 um cenário completo, da forma como foi adquirido pelo sensor. É possível identificar a pessoa e o *background*, neste caso, uma tela branca na parte posterior no fundo de cena. Neste caso é simples separar a região da nuvem de pontos relativa a pessoa da dos pontos relativos ao ambiente, pois eles estão fisicamente separados e há uma divisão bem nítida entre eles que pode ser vista na Figura 12, em uma outra projeção. Nesta figura ainda é possível identificar a sombra da pessoa projetada no anteparo. Enquanto as sombras de imagens RGB

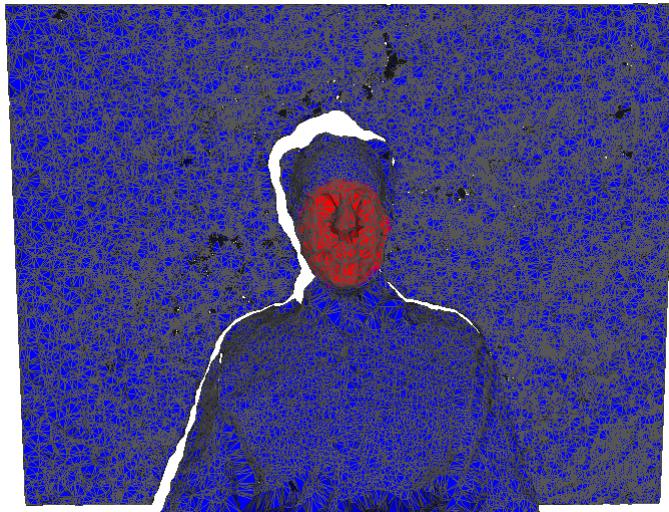


Figura 11 – Recorte da Região útil da face com 80 mm a partir da ponta do nariz. Base de dados 3DMAD.

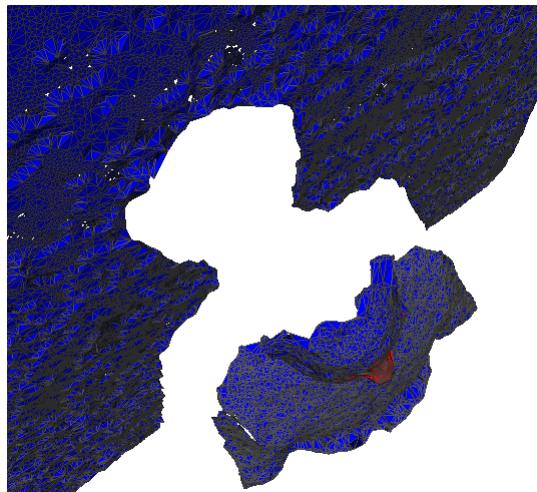


Figura 12 – Recorte de uma projeção da Figura 11 mostrando a separação das nuvens de pontos do indivíduo e cenário.

estão relacionadas à intensidade da luz, as sombras em imagens de profundidade são causadas literalmente pela obstrução do raio do sensor.

A amostra da Figura 13 pertence à base de dados FRGC v2.0 e embora tenha sido adquirida em um ambiente externo complexo, este não aparece na nuvem de pontos apresentada pela base. Ou seja, houve um pré-processamento realizado pela equipe que trabalhou na montagem da base, que separou o indivíduo do cenário externo. Ainda assim, permaneceram, além da face, o pescoço, tronco e parte anterior dos braços.

A amostra da Figura 14 pertence à base de dados Bosphorus e é possível perceber que além de extrair o cenário, a equipe realizou a extração apenas da região do rosto do indivíduo, retirando-se os demais componentes que poderiam dificultar o processamento dos dados da face.

Na Figura 15a é possível perceber a diferença de resolução dos sensores na aquisição

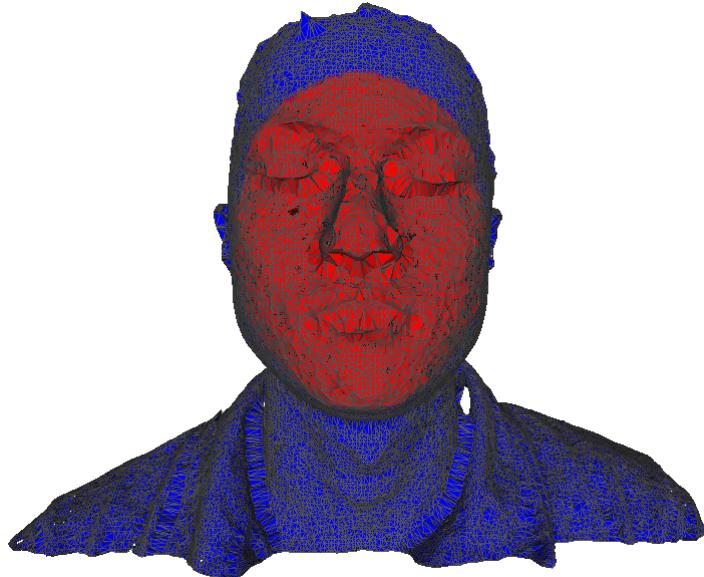


Figura 13 – Recorte da Região útil da face. 80 milímetros a partir da ponta do nariz. Base de dados FRGC v2.0.

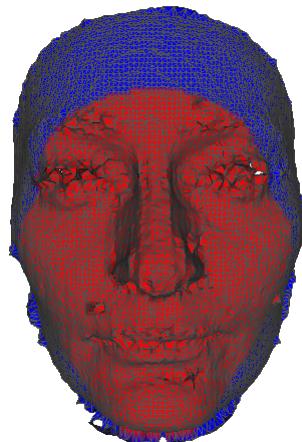
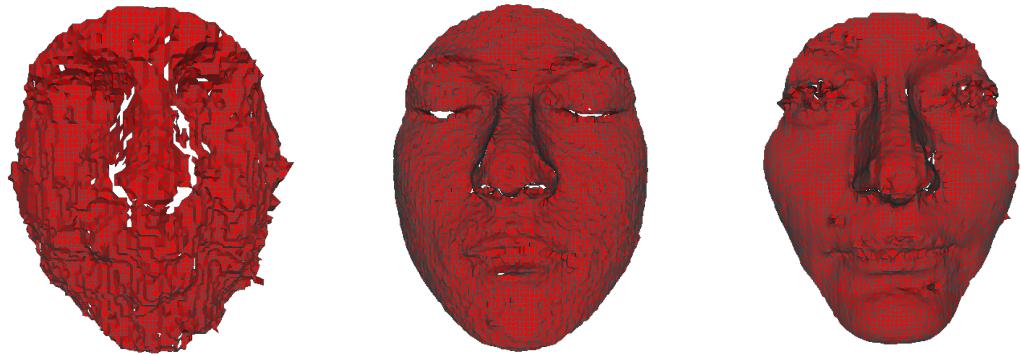


Figura 14 – Recorte da Região útil da face. 80 milímetros a partir da ponta do nariz. Base de dados Bosphorus.

das amostras. O Kinect obtém nuvens de pontos com nível de ruído bem superior aos dos sensores das outras bases aqui citadas. Outra constatação está nos vazios da face, que também são sombreamentos, que embora estejam presentes nas outras bases, são muito maiores na 3DMAD.

Observando-se as Figuras 11, 13, 14 e 15 é possível compreender a complexidade de encontrar e restringir as regiões de interesse das faces, bem como as diferenças de passos a serem executados, dependendo da base de dados utilizada ou, em situações reais, do cenário configurado para extração.

Em resumo, neste trabalho a extração da área útil da face estende-se até a distância Euclidiana  $d = 80$  mm da ponta do nariz, o suficiente para compor as regiões dos olhos, nariz e



(a) 3DMAD.

(b) FRGC v2.0.

(c) Bosphorus.

Figura 15 – Faces extraídas de 3 bases de dados: 3DMAD, FRGC v2.0 e Bosphorus.

boca. O recorte facial é executado por vários autores, incluindo Elaiwat *et al.* (2015), Berretti *et al.* (2014) e Li *et al.* (2015) para restringir a região do rosto que melhor caracteriza o indivíduo. A próxima seção discute a correção da nuvem de pontos, etapa realizada após a extração da face, juntamente com a realização dos cortes triaxiais da face.

### 3.4 Cortes Triaxiais da Face

Nesta seção são apresentados os procedimentos para que a nuvem de pontos da face seja dividida em outras nuvens menores, de acordo com o ponto-chave que corresponde à ponta do nariz, e com os planos  $XY$ ,  $XZ$  e  $YZ$ , ortogonais entre si.

No método proposto, delimitam-se regiões bem definidas e reproduzíveis, segmentadas a partir de um ponto-chave exclusivo e facilmente identificável. A partir dessas regiões são extraídos os descritores locais, suficientemente robustos para identificar indivíduos, mesmo com grande variação de expressões faciais. Isto ocorre porque, sendo os cortes relativos a regiões da face correspondentes entre amostras de um mesmo indivíduo, deformações consideráveis da superfície causadas por expressões concentram-se em algumas partes dessas regiões, mantendo grande parte da superfície sem alterações consideráveis.

A ideia por trás da proposta de método de extração de características apresentada neste trabalho está relacionada à natureza das deformações na superfície da face sob variações de expressão. Diferentes regiões sofrem diferentes perfis de deformação ao longo de direções preferíveis, enquanto a expressão facial muda. Observamos esse comportamento após estimar a deformação como feito no campo do registro não rígido de imagens médicas 3D (CORROCHANO, 2005). A partir dos autovalores do tensor de covariância dos dados tridimensionais, procedeu-se ao cálculo da anisotropia, omnidirecionalidade e outras características geométricas e,

em seguida, foi investigado como esses números variam entre os indivíduos e entre as diferentes regiões de um determinado indivíduo. O resultado desta análise revelou que alguns fragmentos são menos sensíveis à deformação do que outros, confirmando assim a expectativa de que partes que não sofrem grandes deformações ao longo de uma ou mais das três direções são robustas o suficiente para representar indivíduos.

A essência da técnica está no fatiamento da nuvem de pontos de uma imagem da face ao longo dos três eixos espaciais, dando origem ao que chamamos de fatias Frontais, Sagitais e Transversais, como mostrado nas Figuras 17, 18 e 19. O método desenvolvido requer algumas etapas de pré-processamento: primeiro, a nuvem de pontos precisa passar por normalização de densidade, seguida por remoção de outliers, detecção do ponto-chave e corte da face, passos apresentados nas seções precedentes. Finalmente, é feita uma correção de pose para alinhar as faces.

A correção de posicionamento com o *Iterative Closest Point* (ICP) é aplicada para corrigir iterativamente a pose das faces 3D. As amostras de teste estão alinhadas em relação às amostras do treinamento, como em Elaiwat *et al.* (2015), os quais empregam o algoritmo ICP somente para as amostras com grande rotação e usam PCA para as demais, uma vez que o ICP, sendo iterativo, pode tornar todo o procedimento mais lento. Neste trabalho, essa desvantagem foi atenuada, reduzindo a densidade da nuvem e cortando também a face ao redor do nariz. Também foi limitado o número de iterações para 100. Além disso, foram testadas três possibilidades de correção de pose, duas deles tinham como objetivo encontrar o vetor normal da face e alinhá-la com o eixo z: RANSAC-Plane (BORRMANN *et al.*, 2011), com o intuito de encontrar o melhor plano que se encaixe no rosto e usando seu normal para correção de pose; e o método SVD two-sided block-Jacobi (BEČKA *et al.*, 2010) para calcular a face normal. Nenhum deles superou os resultados do ICP. O SVD de dois lados em bloco-Jacobi é rápido e altamente paralelizável (BEČKA *et al.*, 2010) enquanto o RANSAC-Plane tem aproximadamente o mesmo tempo de processamento que o ICP. O melhor arranjo encontrado foi executando primeiramente um alinhamento grosseiro com SVD de dois lados em bloco-Jacobi e depois aplicando ICP para correção fina, o que minimizou o número de iterações necessárias. O número máximo de iterações foi definido como 100, já que valores superiores não tiveram efeito considerável no alinhamento.

A correção da pose é um passo importante para mitigar a inclinação da cabeça durante as sessões de aquisição de imagens, evitando assim a má formação e a indexação das fatias. Tal

procedimento permite que o ponto-chave correspondente à ponta do nariz seja encontrado nas amostras de teste. Nos parágrafos a seguir, o ponto referente à ponta do nariz é referido como  $\mathbf{nu} = (x_n, y_n, z_n)$ . Para cada amostra neutra ou não neutra de uma expressão individual, o ponto  $v$  deve ser coincidente ou estar a uma distância mínima aceitável em relação às outras amostras.

O fatiamento é realizado de tal forma que cada corte contém apenas os pontos  $\mathbf{p} = (x, y, z)$  da nuvem de pontos  $\mathbf{C}$ , situada entre dois planos de cortes adjacentes. Estas seções são igualmente espaçadas ao longo dos eixos. Os planos de referências para as fatias frontais são aqueles paralelos ao plano  $XY$  da Figura 17 e a espessura da fatia é  $c_f$ . Planos de referência para as fatias Sagital e Transversal são, por sua vez, aqueles paralelos aos planos  $YZ$  e  $XZ$  das Figuras 18 e 19, e as respectivas espessuras são  $c_s$  e  $c_t$ . Assim, a tripla  $\mathbf{C} = (c_f, c_s, c_t)$ , definida em milímetros, junto com as coordenadas da ponta do nariz,  $x_k, y_k$  e  $z_k$ , definem os índices  $n$  dos cortes frontais, sagitais e transversais, como mostrado nas Figura 17, 18 e 19.

As seções Frontal, Sagital e Transversal são representadas respectivamente pelos conjuntos de pontos  $F_n, S_n$  e  $T_n, n \in Z$ , conforme Equações 3.13, 3.14 e 3.15:

$$p \in F_n \rightarrow \left\lfloor \frac{|z_k - z|}{c_f} \right\rfloor = n \quad (3.13)$$

$$p \in S_n \rightarrow \left\lfloor \frac{|x_k - x|}{c_s} \right\rfloor \cdot sign(x_k - x) = n \quad (3.14)$$

$$p \in T_n \rightarrow \left\lfloor \frac{|y_k - y|}{c_t} \right\rfloor \cdot sign(y_k - y) = n. \quad (3.15)$$

### 3.4.1 Efeito das Expressões nos Cortes

Nesta seção é apresentado o efeito das expressões faciais nos cortes Frontais, Sagitais e Transversais em relação aos mesmos cortes nas amostras neutras.

São utilizadas as amostras do indivíduo *bs003* cujas imagens RGB e nuvens de pontos tem autorização para serem apresentadas.

As expressões faciais (emocões, LFAU, UFAU e CAU) podem modificar consideravelmente a estrutura da face, ainda mais quando limitado a um corte de 80 mm a partir da ponta do nariz.

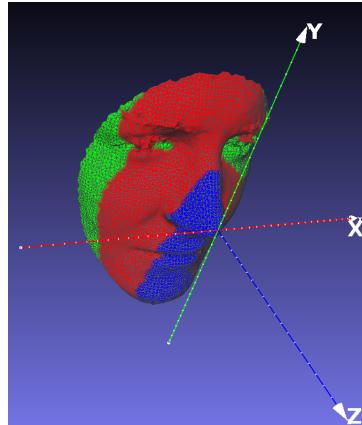


Figura 16 – Orientação da Nuvem de pontos da face em relação ao eixo de coordenadas que tem sua origem na ponta do nariz. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.

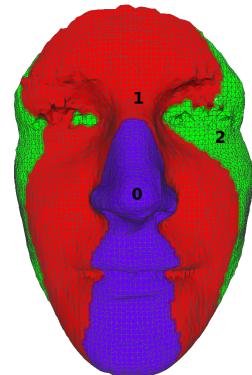


Figura 17 – Nuvem de pontos da Face com cortes Frontais de 20 milímetros de espessura. Cortes Frontais numerados de 0 a 2. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.

A Figura 20 mostra a imagem RGB original (Figura 20a) e os cortes Frontais, Sagitais e Transversais desta face, já com a ponta do nariz identificada e com a restrição de 80 mm de distância a partir do nariz.

Na Figura 21 são apresentados os cortes frontais de algumas amostras do indivíduo bs003 da base de dados Bosphorus, relacionados com as poses Neutra e expressões emocionais. Mesmo pertencendo a um mesmo indivíduo, é fácil perceber a grande variação dos cortes frontais, seja o primeiro (azul), o segundo (vermelho) ou o terceiro (verde). Pode-se dizer, inclusive, que poderiam pertencer a indivíduos distintos.

A depender da expressão, a nuvem de pontos têm seu contorno modificado, podendo ficar mais alongada, como na Figura 21b, ou mais arredondada, como na Figura 21c. Essa variação pode apenas incluir, apenas excluir ou incluir e excluir voxels em determinados cortes. São essas variações que podem alterar os valores dos atributos extraídos destes cortes em relação aos atributos dos mesmos cortes na nuvem de pontos Neutra. Vale relembrar que o treinamento

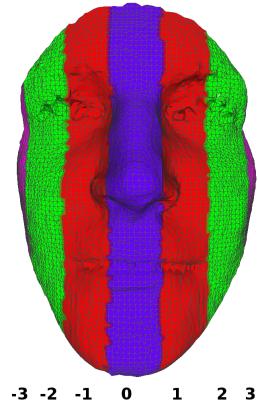


Figura 18 – Nuvem de pontos da Face com cortes Sagitais de 20 milímetros de espessura. Cortes Frontais numerados de -3 a 3. Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.

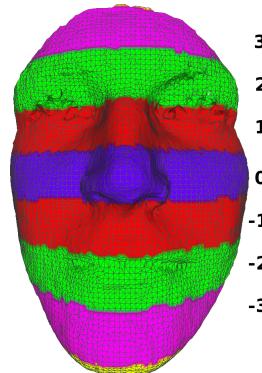


Figura 19 – Nuvem de pontos da Face com cortes Transversal de 20 mm de espessura. Cortes Frontais numerados de -3 a 3. A Triangulação e Renderização da superfície foram realizadas para melhorar a visualização da face.

das faces acontece com as nuvens neutras e o teste com as outras faces neutras, se existirem, e com as faces com expressões.

Nas Figuras 22 e 23 pode-se identificar o efeito das expressões nos cortes Sagitais e Transversais, respectivamente. As maiores diferenças acontecem nos cortes Sagitais ou Transversais, os quais compreendem as bordas que mais se modificaram, como acontece com os cortes verdes Sagitais, Figuras 22a e 22b, correspondentes à região dos olhos, bastante afetada com a expressão de Raiva.

Nos cortes Transversais da Figura 23 pode-se perceber alguns de cor anil, após os lilás (3 ou -3), de acordo com a padronização da Figura 19. Embora estejam na imagem, não fazem parte dos cortes que participam da extração de atributos.

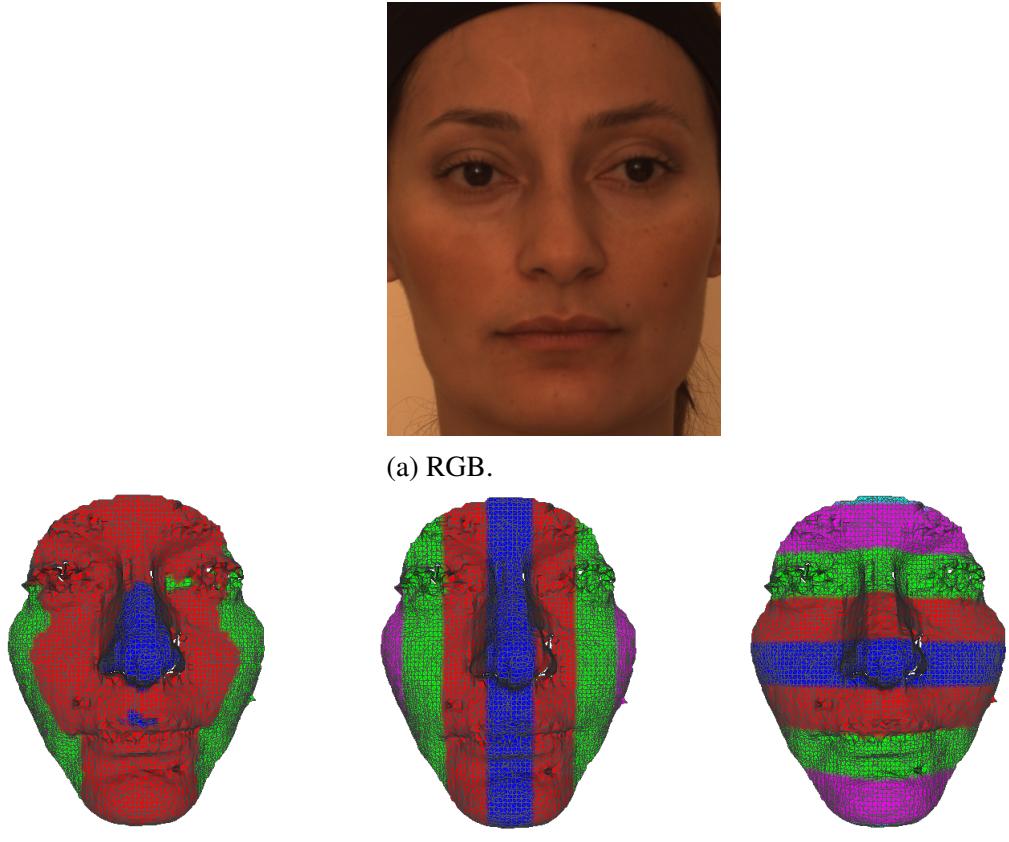


Figura 20 – Indivíduo bs003 extraído da base de dados Bosphorus. Amostra neutra, mulher e atriz.

### 3.4.2 Combinação dos Cortes

Nesta seção é apresentada a motivação para a combinação dos cortes Frontais, Sagitais e Transversais.

Observando as Figuras 21, 22 e 23 é possível perceber que as expressões podem mudar nas fatias de qualquer uma das direções. Os limites da face sofrem mudanças nos três eixos e as fatias variam sensivelmente em alguns casos. Um fato interessante deve ser destacado: os movimentos da face não exigem a interação de todos os músculos desta em uma mesma expressão, ou pelo menos não o fazem de forma muito relevante, a ponto de alterar significativamente todas as fatias. Há uma alteração mais nítida nos cortes Frontais, mas que pode ser percebida nos Sagitais e Transversais também em menor intensidade.

Com esta hipótese, a melhor alternativa é não escolher uma das direções para realizar a extração de atributos, mas utilizar todos eles e identificar se há alguma direção preferencial nos cortes ou se há alguma combinação que tenha uma melhor performance.

Ainda sobre a combinação dos cortes, o número destes por eixo, que já foi apresen-

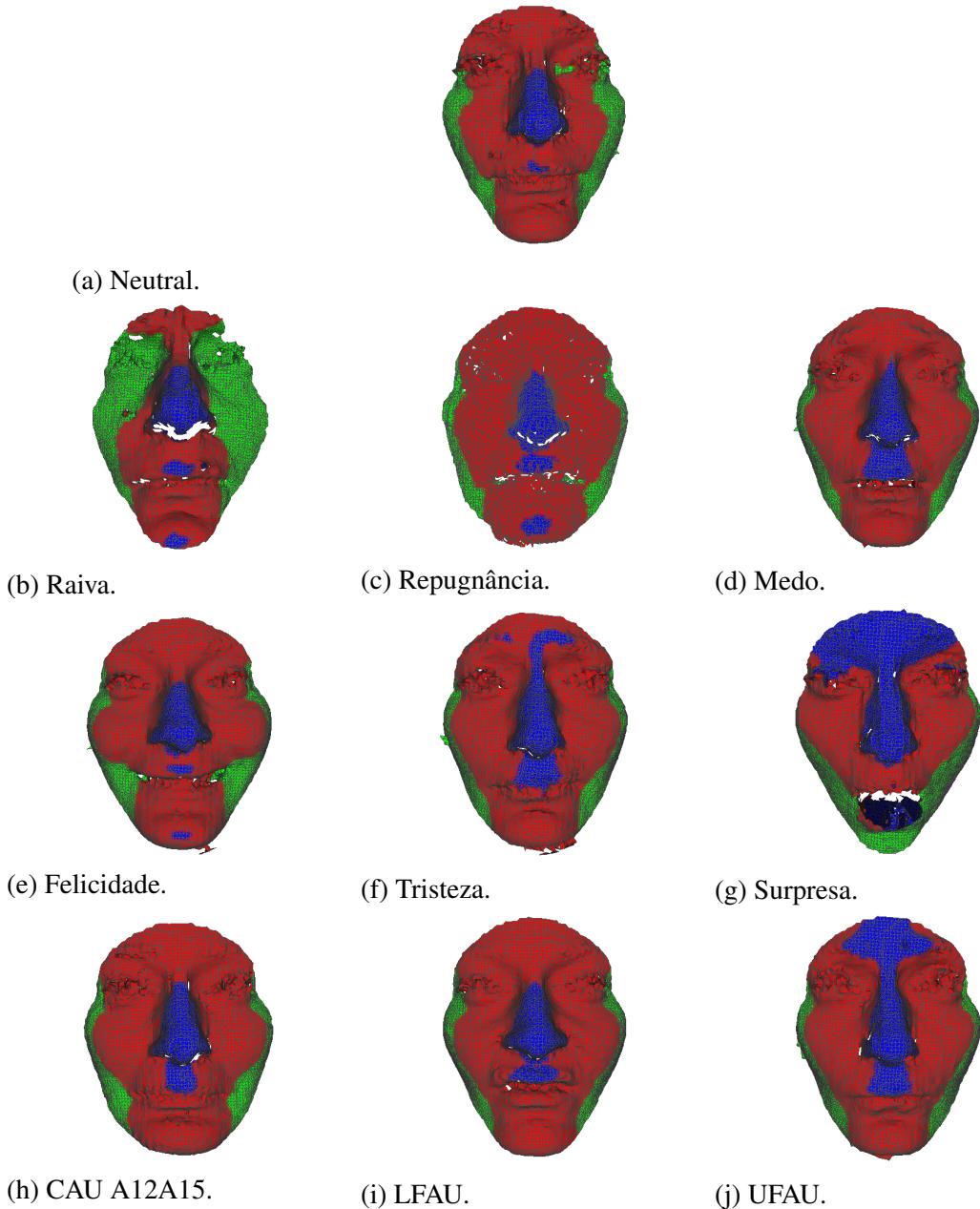


Figura 21 – Cortes Frontais do indivíduo bs003 extraído da base de dados Bosphorus. Amostras: Neutra, Expressões emocionais, LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 mm da ponta do nariz.

tado nas Figuras 17, 18 e 19, abrange os cortes mais relevantes e que possuem, além de uma quantidade mínima de pontos, uma frequência de aparecimento alta na maioria dos indivíduos estudados.

Desta forma, os cortes são combinados na sequência: Frontal, Sagital e Transversal. Os Frontais aparecem na sequência  $[0, 1, 2, 3]$  e os Sagitais e Transversais na sequência  $[0, -1, 1, -2, 2, -3, 3]$ . Considerando que os cortes Frontal, Sagital e Transversal zero sejam padronizados para  $F_0$ ,  $S_0$  e  $T_0$ , respectivamente, a combinação de todos os cortes é definida como:

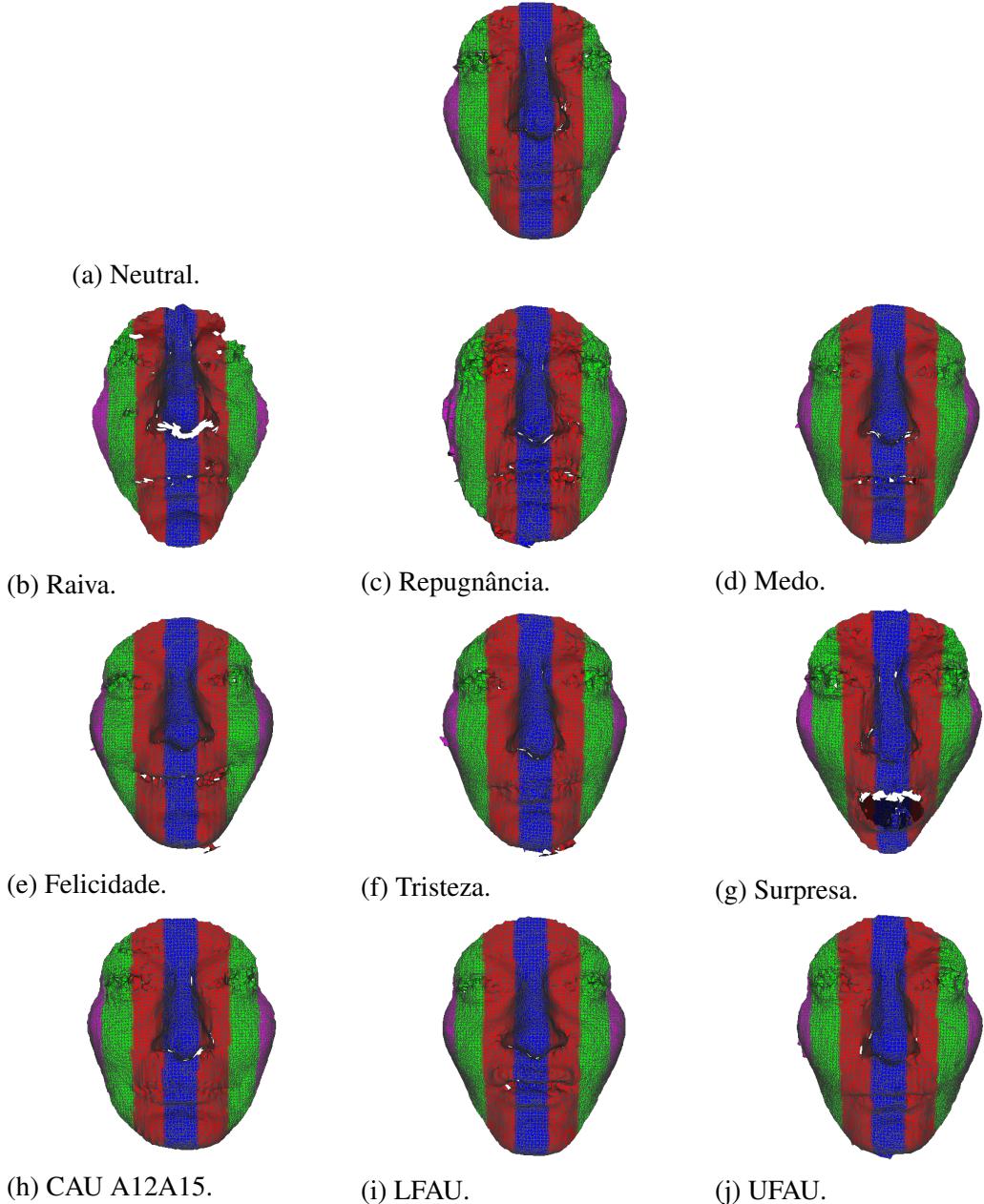


Figura 22 – Cortes Sagitais do indivíduo bs003, extraído da base de dados Bosphorus. Amostras: Neutra e Expressões emocionais, LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 mm da ponta do nariz.

$$[F_0, F_1, F_2, F_3, S_0, S_{-1}, S_1, S_{-2}, S_2, S_{-3}, S_3, T_0, T_{-1}, T_1, T_{-2}, T_2, T_{-3}, T_3].$$

### 3.5 Classificadores

Nesta seção são apresentados os classificadores utilizados neste trabalho para realizar os processos de identificação e verificação das faces, bem como as implementações utilizadas para cada caso.

Trabalhos de reconhecimento facial com imagens 2D e nuvens de pontos 3D utilizam

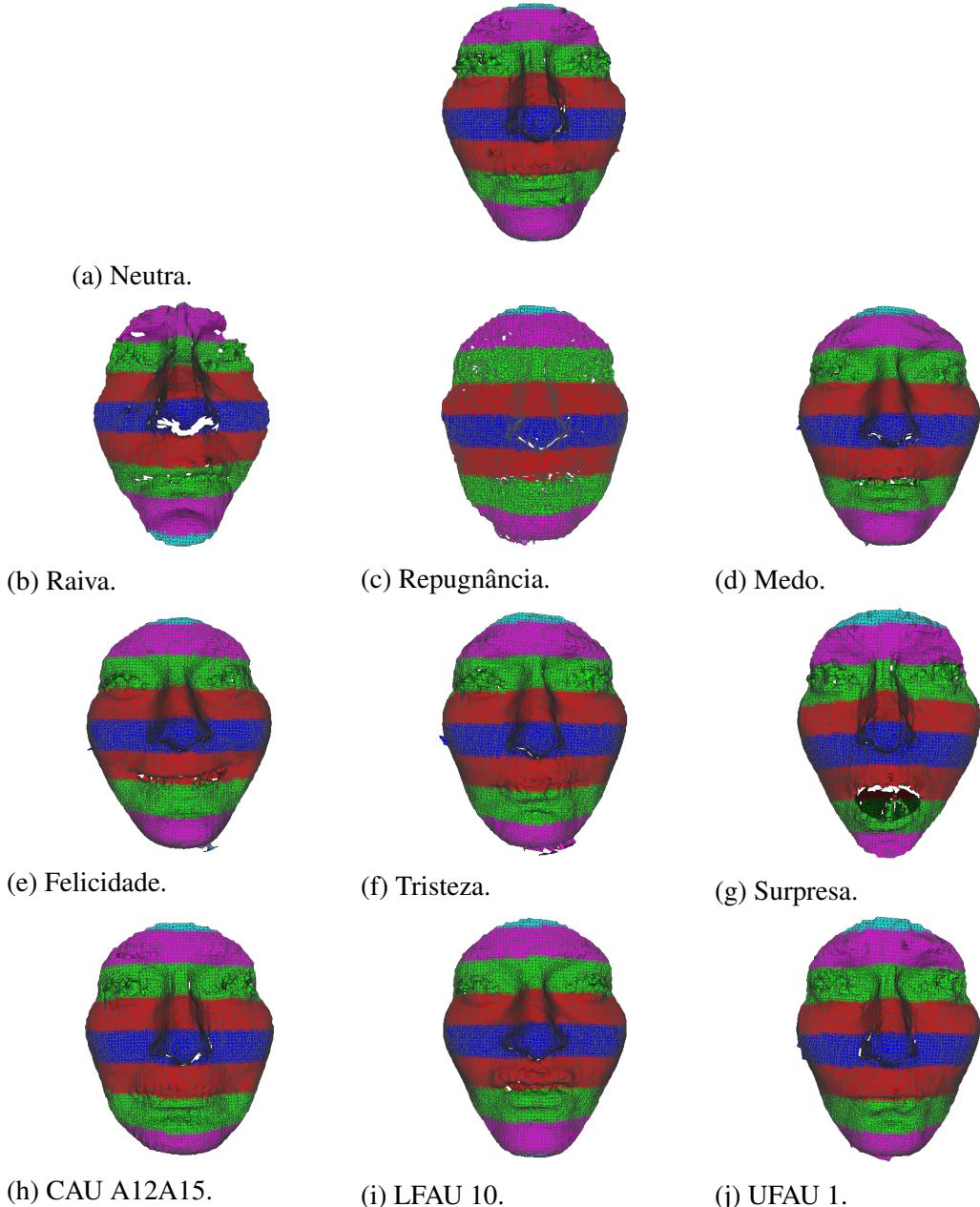


Figura 23 – Cortes Transversais do indivíduo bs003, base de dados Bosphorus, amostras 0 Neutra e Expressões emocionais; LFAU, UFAU e CAU. Nuvens de pontos recortadas a 80 milímetros da ponta do nariz.

diversos classificadores. Dentre os mais comuns para este problema estão o LDA (*Linear Discriminant Analysis*) e SVM (*Support Vector Machines*), como pode ser visto nos trabalhos de Ahdid *et al.* (2017b), Ahdid *et al.* (2017a), Huang *et al.* (2014), Marvadi *et al.* (2015), Wang *et al.* (2015), Cheng *et al.* (2017), Ameur *et al.* (2016) e Abuzneid e Mahmood (2018).

### 3.5.1 LDA

Nos problemas de reconhecimento de padrões, as imagens são geralmente representadas por um vetor de característica  $t$ -dimensional, que pode ser muito grande, comprometendo a robustez e a eficiência na tarefa de reconhecimento. Para resolver o problema de conjuntos de dados de alta dimensão, algumas técnicas de redução de dimensionalidade podem ser empregadas, reduzindo os custos computacionais e evitando *overfitting*.

A Análise Linear Discriminante tem sido comumente usada para muitas aplicações, como o reconhecimento facial. Pode ser usada como uma técnica para redução de dimensionalidade (MOON *et al.*, 2015) e também como um classificador (TZALAVRA *et al.*, 2016). O LDA procura por vetores de base de um subespaço no qual os dados seriam projetados e que maximizam a distinção entre classes (MARTINEZ; KAK, 2001). Para isso, duas medidas são calculadas:  $S_w$ , a matriz de dispersão de *intra-classe*, e  $S_b$ , a matriz de dispersão de *entre-classes*.

$$S_w = \sum_{j=1}^c \sum_{i=1}^{n_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T, \quad (3.16)$$

em que  $c$  é o número de classes,  $n_j$  é o número de amostras da classe  $j$ ,  $x_i^j$  é a  $i$ -éssima amostra da classe  $j$  e  $\mu_j$  é o vetor médio da classe  $j$ .

$$S_b = \sum_j^c (\mu_j - \mu)(\mu_j - \mu)^T, \quad (3.17)$$

em que  $c$  é o número de classes,  $\mu_j$  é o vetor médio da classe  $j$  e  $\mu$  é o vetor médio de todas as classes.

Se  $S_w$  não for singular, então um novo subespaço de recursos pode ser encontrado resolvendo o problema de autovalores generalizados para  $S_w^{-1}S_b$ . Assim, considerando um vetor de característica de dimensão  $t$  e o número de classes  $c$ , no máximo  $c - 1$  autovetores generalizados não zero (discriminantes lineares) podem ser encontrados e pelo menos  $t + c$  amostras são requeridas para garantir que  $S_w$  não se torne singular (MARTINEZ; KAK, 2001).

Diversas abordagens têm sido propostas para resolver esse problema de singularidade, incluindo a aplicação de SVD (*Singular Value Decomposition*) à matriz de dados e a utilização de algoritmos iterativos, como o LSQR, para resolver o LDA como um problema de mínimos quadrados.

### 3.5.2 SVM

O Support Vector Machine (SVM) é originalmente um classificador binário que constrói uma superfície de decisão linear ideal baseada em uma combinação ponderada de elementos do conjunto de treinamento, os chamados vetores de suporte (CORTES; VAPNIK, 1995; MOUSAVIDI *et al.*, 2008). O treinamento é formalmente definido por Cortes e Vapnik (1995) como minimizando a função de erro:

$$\frac{1}{2}w^2 + C \sum_{i=1}^l \xi_i^\sigma \quad (3.18)$$

condicionada à:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1, \dots, l, \quad (3.19)$$

$$\xi_i \geq 0, i = 1, \dots, l. \quad (3.20)$$

onde  $C$  é uma constante e controla o limite entre a complexidade da superfície de decisão e a frequência de erro de classificação no conjunto de treinamento (CORTES; VAPNIK, 1995). Baixos valores de  $C$  tornam a superfície de decisão suave e simples, enquanto valores altos permitem que mais amostras sejam escolhidas como vetores de suporte, criando uma superfície de decisão que tende a separar exatamente classes no conjunto de treinamento.

O *Support Vector Machines* (SVM) também pode ser estendido para construir uma superfície de decisão não linear usando uma função de kernel que projeta o espaço original em um espaço de recurso de alta dimensão. A função do kernel usada neste documento é o kernel da função de base radial (Radial Basis Function (RBF)), como calculado em:

$$K(u, v) = \exp -\gamma|u - v|^2, \quad (3.21)$$

em que  $\gamma$  é uma constante e, na prática, ajusta a influência de cada amostra na definição da superfície de decisão. Valores baixos de  $\gamma$  significam que cada amostra tem uma influência muito grande e, portanto, mesmo amostras distantes podem afetar a definição da superfície de decisão.

Por outro lado, valores altos de  $\gamma$  significam que cada amostra tem uma influência próxima, o que pode forçar a superfície de decisão a se adaptar às amostras mais próximas.

Os parâmetros  $C$  e  $\gamma$  afetam principalmente a capacidade de generalização do classificador e, portanto, seu adequado ajuste é fundamental. Neste ponto há uma questão muito importante que é como configurar os parâmetros  $C$  e  $\gamma$ . Se tomarmos o experimento ROC-I para a base de dados Bosphorus, tem-se dois conjuntos: um de treinamento com 299 amostras neutras e um de testes com 2.902 amostras não neutras.

Os parâmetros são avaliados no conjunto de treinamento e apenas depois analisados nos conjuntos de testes. Na base de dados Bosphorus, que possui 105 classes, há uma distribuição irregular de amostras por classe no conjunto de treinamento, variando de 1 a 4 amostras por classe, ou seja, algumas possuem várias amostras para se extrair uma configuração média melhor e outros apenas uma amostra. Assim, a configuração dos parâmetros pode gerar taxas de acerto diferentes tanto no conjunto de treinamento quanto no conjunto de testes. Analisando apenas o conjunto de treinamento, pode-se encontrar um conjunto de vários pares  $(C, \gamma)$  que encontrem taxas de acerto de 100%, o que não significa que todos tenham taxas de acerto similares junto ao conjunto de testes.

Já para o experimento rank-1, onde há um conjunto de treinamento com 105 amostras, uma por classe, a variação dos valores  $(C, \gamma)$  não afeta consideravelmente as taxas de acerto no treinamento nem nos testes.

### **3.5.3 Escolha dos classificadores**

O classificador SVM possui dois parâmetros de fundamental importância para obter bons resultados de classificação, são o  $C$  e o  $\gamma$ , e ambos devem ser obtidos a partir das amostras de treinamento. Para os experimentos rank-1, onde há apenas uma amostra por classe, não há necessidade de realizar qualquer configuração especial dos parâmetros, pois não há praticamente variações nas taxas de reconhecimento. No caso dos experimentos para ROC I, os parâmetros devem ser configurados de forma a maximizar a taxa de acerto das amostras de teste.

Há duas implementações do SVM a considerar: *one vs all* (One vs All training strategy (OvA)) e *one vs one* (One vs One training strategy (OvO)). Considerando  $N$  o número de classes, a primeira implementação possui  $N$  classificadores binários, um para cada classe, onde o classificador  $W_i$  é configurado considerando como classe positiva as amostras da classe  $i$  e como classe negativa todas as demais amostras. Já a segunda implementação possui  $N(N - 1)/2$

classificadores binários, onde o classificador  $W_{i,j}$  considera como classe positiva as amostras da classe  $i$  e como classe negativa as amostras da classe  $j$ . Não há motivos para realizar o treinamento dentro da mesma classe, logo, o classificador  $W_{i,i}$  não faz sentido e considera-se que  $W_{i,j} = -W_{j,i}$ . Embora as implementações atuais do SVM considerem a configuração OvA tão boa quanto a OvO, os testes realizados neste trabalho, com o número de classes e atributos testados, demonstraram que o segundo ainda é melhor que o primeiro, a despeito do seu tempo de processamento, tanto para a fase de treinamento como para a de testes. Considerando a base de dados Bosphorus, que possui 105 classes, a versão OvO do SVM terá  $105 * 104 / 2 = 5.460$  classificadores. Se considerarmos a base FRGC v2.0, que possui 466 classes, serão  $466 * 465 / 2 = 108.345$  classificadores.

O LDA, utilizado como classificador, possui algumas vantagens em relação ao SVM pois, além de não necessitar de uma fase preliminar de configuração de parâmetros, o tempo de treinamento e principalmente o tempo de classificação é consideravelmente menor, sobretudo quando se avalia o melhor caso do SVM OvO. Em contrapartida, para os experimentos do rank-1 não é possível utilizá-lo, uma vez que há apenas uma amostra por classe para todas as classes e isso o impede de realizar as operações de inversão, ou mesmo pseudo-inversão, das matrizes. A base Bosphorus possui algumas classes com apenas uma amostra neutra, mesmo assim, é possível utilizar LDA pois a opção da pseudo-inversão da matriz é viável neste caso. Uma das implementações do LDA consiste em utilizar *shinkage estimators* da matriz de covariância para os casos onde há um pequeno número de amostras de treinamento, como é o caso deste trabalho.

Como dito anteriormente, há dois tipos de experimentos mais comuns na literatura em relação à base utilizada: ROC I e rank-1. No experimento ROC I tem-se 299 amostras neutras na base de treinamento e 2.603 amostras não-neutras na base de testes. Já no experimento rank-1 tem-se 105 amostras neutras na base de treinamento e duas bases de teste, uma com 194 amostras neutras distintas do treinamento e 2.603 amostras não-neutras.

Martinez (MARTINEZ; KAK, 2001), analisando as diferenças entre PCA e LDA, afirma que para a matriz da Equação 3.16 não se torne singular, deve-se ter  $S > N + A$ , onde  $S$  é o número de amostras,  $N$  é o número de classes e  $A$  o número de atributos. Essa sentença justifica o motivo pelo qual este trabalho usa LDA para os experimentos ROC I e não o faz para os experimentos rank-1.

No Capítulo 4 são apresentados e discutidos os resultados da aplicação do método apresentado.

## 4 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os principais resultados da aplicação do método proposto e apresentado do Capítulo 3. Experimentos padronizados para reconhecimento de faces 3D foram adotados e os resultados obtidos foram comparados aos de trabalhos de referência na literatura especializada neste domínio.

### 4.1 Discussão Preliminar

Várias bases de dados diferentes foram usadas para o reconhecimento facial 3D na literatura recente. A base de dados Bosphorus foi selecionada para este trabalho por ser a que traz as condições mais desafiadoras. Além disso, também foi avaliada a generalização do método proposto no conjunto de dados FRGC v2.0.

A Base de Dados do Bosphorus foi introduzida por Savran *et al.* (2008) e é agora amplamente utilizada no reconhecimento facial 3D. O banco de dados fornece amostras 2D e 3D junto com coordenadas de referência localizadas em regiões de Unidades de Ação (AU) (EKMAN; FRIESEN, 1976). Também fornece a correspondência entre dados 3D e pixels de imagem 2D, possibilitando o trabalho em espaços 2D ou 3D.

A base Bosphorus comprehende 4.666 varreduras faciais não distribuídas uniformemente pelos 105 indivíduos em várias expressões, poses e condições de oclusão. Os indivíduos têm entre 25 e 35 anos, e a maioria deles é caucasiana. Entre eles, 29 atores integram o banco de dados, dando-lhe complexidade adicional, uma vez que podem exagerar as expressões e, assim, exagerar substancialmente na deformação da face (ver Figura 24). Barba, bigode e projeção dos cabelos também estão presentes em alguns indivíduos.

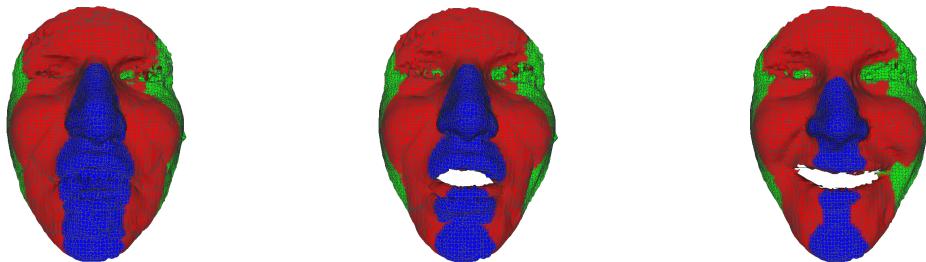


Figura 24 – Cortes frontais de algumas expressões emocionais simples: Tristeza, Surpresa e Felicidade.

As amostras disponíveis no banco de dados de faces 3D Bosphorus podem ser divididas em quatro categorias: as que executam expressões, poses, oclusões ou com rótulo

ignorado. Algumas subcategorias são derivadas das citadas anteriormente e suas composições são as seguintes: 34 Expressões (2.902 amostras); 13 Poses (1.365 amostras); 4 oclusões (381 amostras) e 1 ignorada (18 amostras).

Por sua vez, a base de dados FRGC v2.0 compreende 466 indivíduos coletados em 4.007 amostras (PHILLIPS *et al.*, 2005c) durante o ano letivo de 2003-2004. É composto por 2.410 amostras 3D neutras e 1.597 amostras 3D não neutras de diferentes expressões faciais (repugnância, felicidade, tristeza, surpresa, outras).

Como afirmado por Soltanpour *et al.* (2017), é difícil fazer uma comparação justa entre os métodos na literatura para reconhecimento facial, principalmente porque diferentes experimentos são apresentados em várias situações e para diferentes condições e bancos de dados, ou seja, número de sujeitos, amostras e amostras por expressão variam significativamente. Para contornar isso, os pesquisadores frequentemente adotam números de mérito semelhantes. Por exemplo, a análise das curvas *Receiver Operation Characteristic* (ROC), que consistem em experimentos de verificação de face criados pela plotagem da taxa de verdadeiros positivos (TPR) ou taxa de verificação (VR) contra a taxa de falsos positivos (FPR) em vários ajustes de limiares é um escolha comum para avaliação de desempenho de algoritmos de reconhecimento facial. Procedeu-se da mesma forma e optou-se por realizar dois tipos de experimentos para reconhecimento facial, utilizando os bancos de dados de faces 3D Bosphorus e FRGC v2.0, denominados ROC I, ROC III e rank-1.

No experimento ROC I, o conjunto de treinamento é composto de todas as amostras neutras, enquanto o conjunto de testes é composto de amostras não neutras. Para a base de dados Bosphorus, significa que existem dois conjuntos: (a) Conjunto de treinamento com 299 amostras neutras e (b) Conjunto de teste com 2.603 amostras não neutras. Para FRGC v2.0, esta configuração experimental é : (a) Conjunto de treinamento com 2.399 amostras neutras e (b) Conjunto de testes com 1.608 amostras não neutras.

No experimento rank-1, por sua vez, o conjunto de treinamento compreende uma amostra Neutra por indivíduo, enquanto existem dois conjuntos de testes, um composto de outras amostras Neutras e o outro composto por amostras Não-Neutras. Para a base de dados do Bosphorus, esses três conjuntos são agrupados da seguinte forma: (a) Conjunto de treinamento neutro com 105 amostras neutras, uma por indivíduo; (b) Conjunto de testes neutro ajustado com o restante das 194 Amostras neutras e (c) Conjunto de testes não neutros, ajustado com o restante das 2.603 amostras de expressão. Para o FRGC v2.0, está distribuído da seguinte forma:

(a) o conjunto de treinamento tem 466 amostras Neutras, (b) Conjunto de testes neutros com 1.933 amostras neutras e (c) conjunto de testes não neutro com 1.608 amostras Não-Neutras.

O ROC I é a configuração mais típica do experimento e um dos principais objetivos deste trabalho. Ele permite que um número maior de amostras por indivíduo seja incluído no conjunto de treinamento, ou seja, inclui variabilidade por classe no conjunto de entrada. Na base de dados do Bosphorus existe um pequeno número de amostras neutras. Elas representam cerca de 6,4% do banco de dados inteiro e 10% de todas as amostras de Expressão usadas neste trabalho. Além disso, tem uma distribuição desequilibrada de até 4 amostras por indivíduo e a média é de 2,8 amostras por indivíduo.

Rank-1 é uma configuração mais desafiadora, uma vez que avalia o desempenho do classificador quando apenas uma amostra por classe é apresentada. Com essa restrição, o classificador precisa ter a capacidade de reconhecer indivíduos mesmo quando executam expressões que causam deformações significativas no rosto e apenas com uma amostra de treinamento. Como mencionado anteriormente, o banco de dados do Bosphorus incluiu um número relevante de atores profissionais como indivíduos, a fim de garantir alta expressividade, o que a torna uma base a ser superada. Outra vantagem do Bosphorus é o nível de subdivisão da base em vários subconjuntos que permite que se realize uma análise não apenas de Neutro vs Não Neutro, mas uma análise dentro das diversas subdivisões como Felicidade, Angustia, LFAU, UFAU, CAU, dentre outras.

Tendo definida a configuração de treinamento e teste, o classificador mais adequado deve ser selecionado. Dessa forma, foram selecionados dois para os experimentos propostos: Análise Linear Discriminante (LDA) e *Support Vector Machine* (SVM), pois ambos são freqüentemente referidos na literatura como adequados à cenários com poucas amostras por classe, o que é apropriado para este trabalho.

As Figuras 25 e 26 apresentam o comportamento de todos os subgrupos, da base de dados Bosphorus, analisados de acordo com a combinação dos cortes Frontais, Sagitais e Transversais. É possível perceber, em todos os gráficos, que o desempenho dos cortes Frontais é o pior de todos e que a combinação dele com qualquer um dos outros cortes provoca um decréscimo da taxa de acerto. Os cortes Sagitais sozinhos fornecem os segundos melhores resultados para a grande maioria dos experimentos, mas a combinação Sagital+Transversal (ST) é a que tem a melhor taxa de acerto para o conjunto global dos grupos. Logo a escolha lógica pela combinação ST ser a que deve ser realizada, independente do tipo de expressão da face no

conjunto de testes.

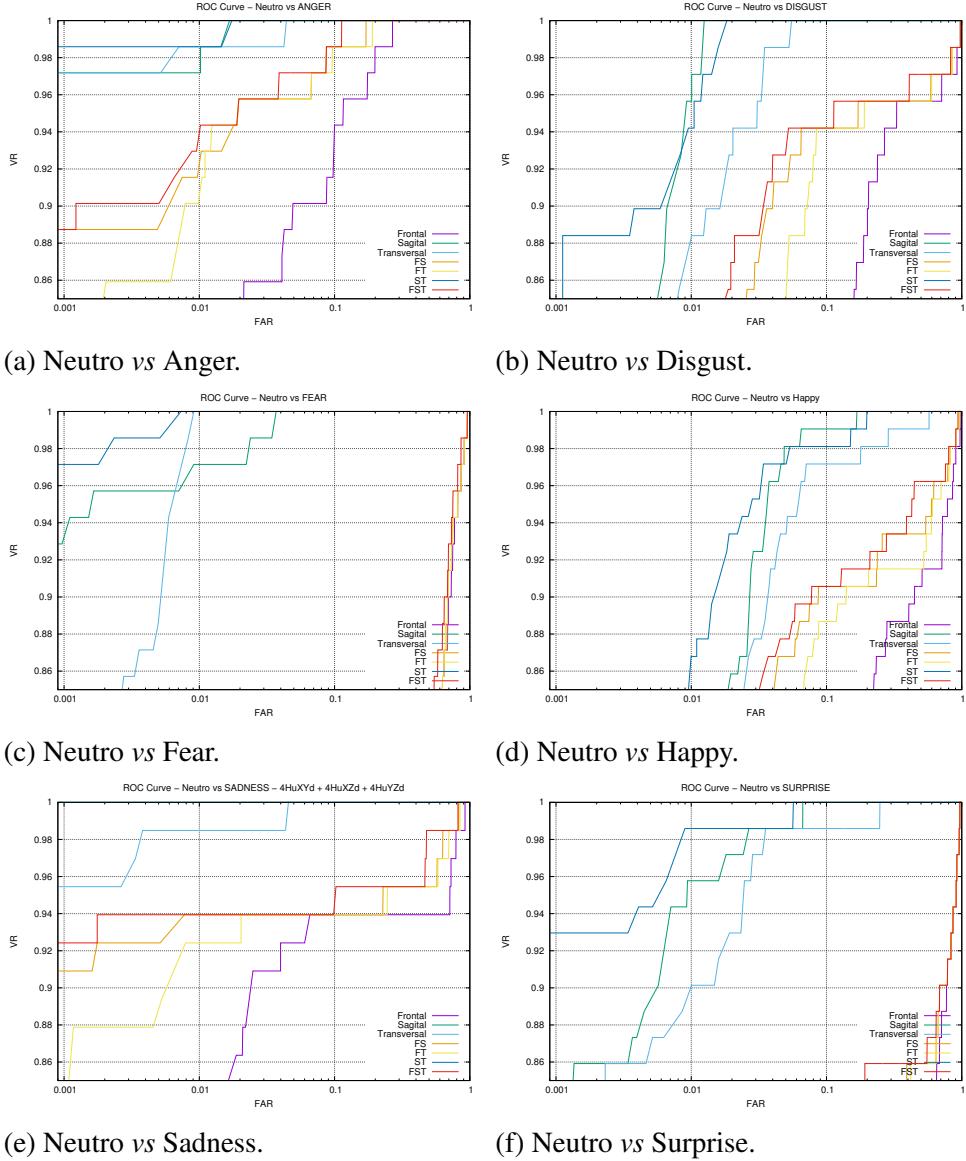


Figura 25 – Combinação dos cortes Frontal, Sagital e Transversal. Bosphorus rank-1. SVM-SVC-rbf. Análise do comportamento das combinações de cortes. Subgrupo das Emoções.

#### 4.1.1 Resultados

O método proposto foi desenvolvido utilizando bibliotecas C / C ++ OpenCV e PCL na plataforma Ubuntu/Linux, e todos os experimentos foram implementados em um PC com a CPU da Intel i7-5500U, 2.4 GHz, 8GB RAM, utilizando apenas um dos quatro núcleos . Foi observado que a correção da pose é o procedimento que demanda mais tempo, 75% do tempo total (4 segundos).

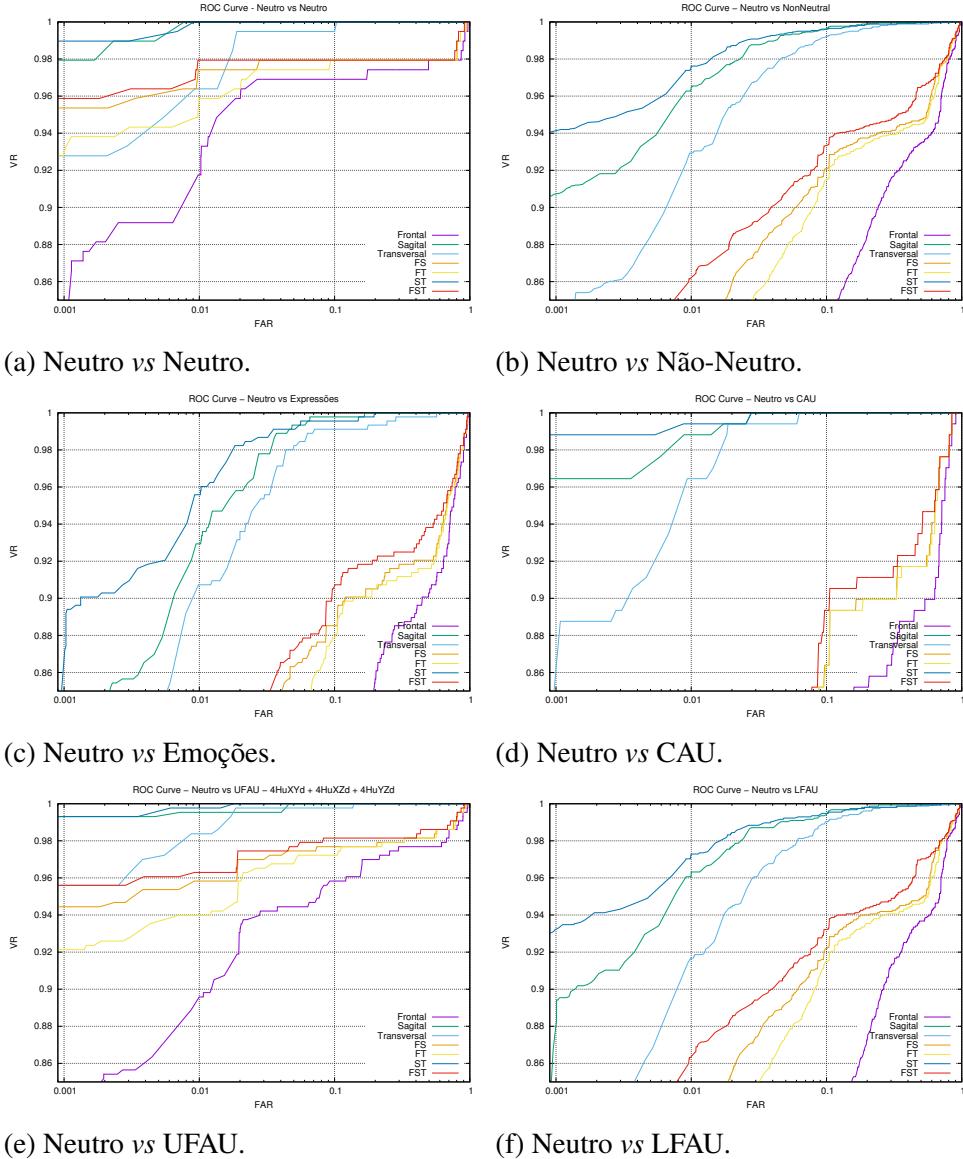


Figura 26 – Combinação dos cortes Frontal, Sagital e Transversal. Bosphorus rank-1. SVM-SVC-rbf. Análise do comportamento das combinações de cortes. Grandes grupos.

O tempo de processamento para cálculo dos momentos de HU e dos momentos 3DIT foi consideravelmente reduzido após os procedimentos de normalização de densidade, com filtro de grade de voxels PCL de tamanho de folha de  $2mm$  para todas as 3 dimensões.

Para realizar o fatiamento triaxial, foram testados os valores [5, 10, 15, 20, 25], em milímetros, para a espessura  $\mathbf{c} = (c_f, c_s, c_t)$ . Os melhores resultados foram obtidos com o triplo  $\mathbf{c} = (20, 20, 20)$ . Valores mais baixos revelaram ser muito sensíveis à rotação, enquanto valores mais altos levaram a uma pequena quantidade de fatias e, portanto, quantidade insuficiente de descritores.

Em seguida, avaliou-se a capacidade descritiva dos momentos de Hu e foi observado que apenas os quatro primeiros são essenciais para a classificação, para qualquer função  $f(x, y)$

Tabela 13 – RR para Neutral *vs* Não neutro, experimento rank-1 com combinação dos cortes ST, utilizando SVM-SVC-rbf.

Moments	4HuXY	4HuXZ	4HuYZ	3DIT
RR (%)	72,30	41,53	48,56	58,47
Moments (distance)	4HuXYd	4HuXZd	4HuYZd	3DITd
RR (%)	93,32	84,90	90,13	90,13

testada. Com essa restrição, o número máximo de atributos foi reduzido para 612. Isso melhorou o cálculo dos descritores, já que os três últimos momentos consomem mais tempo que os primeiros. Assim, definimos 4HuXYd como o conjunto dos primeiros 4 momentos Hu, considerando XY plane e a função  $f(x,y) = 1$  calculada como na Equação 3.4. Terminologia semelhante foi adotada para os outros momentos.

#### 4.1.2 Melhor combinação de cortes e momentos

A partir dos dados apresentados na Tabela 13 é possível observar que recursos calculados de acordo com a função de distância Euclidiana da Equação 3.4 são individualmente melhores que os calculados, para  $f(x,y) = (z - \bar{z})$  na mesma Equação 3.4. Além disso, os recursos HuXY são globalmente melhores e, independentemente da função adotada nos experimentos, o grupo de momentos HuXZ apresentou desempenho pior, o que sugere que a coordenada y é uma das variáveis dependentes que contribui consideravelmente para melhores resultados. Em seguida, também foram testadas combinações de todos os recursos e o melhor arranjo foi 4HuXYd, 4HuXZd e 4HuYZd.

Uma vez que a melhor combinação de momentos foi definida, o próximo passo foi determinar a combinação de fatias que produzisse a melhor taxa de reconhecimento. Foram avaliadas, portanto, as fatias Frontal (F), Sagital (S) e Transversal (T) individualmente e combinadas, formando o conjunto de testes  $[F, S, T, FS, FT, ST, FST]$ .

A Figura 27 apresenta os desempenhos dos atributos 4HuXYd + 4HuXZd + 4HuYZd para todas as combinações possíveis de slicings Frontal (F), Sagital (S) e Transversal (T). O corte frontal sozinho é o pior cenário e, quando combinado a outros, degrada o VR. O melhor cenário é a combinação Sagital-Transversal (ST). Essa tendência se repete nos vários casos investigados.

O comportamento observado na Figura 27 também foi obtido para o ROC I com o classificador SVM-SVC-rbf. Para o ROC I, também foi utilizado o LDA como classificador e a melhor combinação de descritores foi 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd e 4HuYZd. O classificador SVM-SVC-rbf teve um desempenho pior que o LDA.

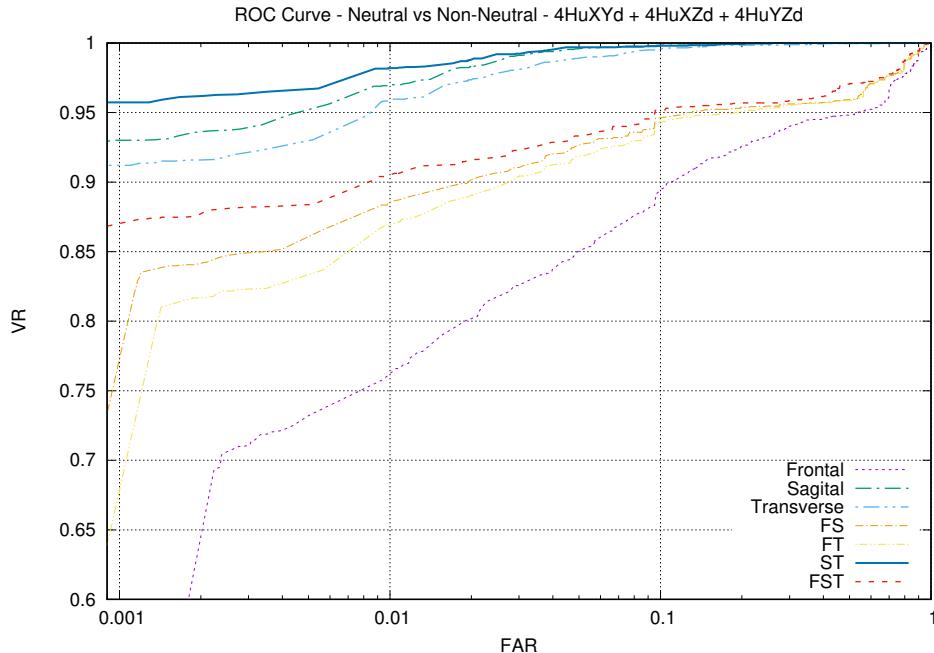


Figura 27 – VR para Neutral vs Não neutro, experimento rank-1 com combinação dos cortes ST, utilizando SVM-SVC-rbf.

#### 4.1.3 Experimentos ROC I, ROC III e rank-1

Inicialmente, os experimento ROC I foram realizados para comparação com os trabalhos que representam o estado da arte. A taxa de verificação (VR) é analisada, uma vez que os trabalhos citados não trazem taxa de reconhecimento (RR). Apresentamos então os resultados para o experimento de rank-1, pois é a mais desafiadora e a que é analisada em trabalhos recentes. Para o experimento rank-1 apenas a taxa de reconhecimento é utilizada como indicador de desempenho. Resultados dos experimentos ROC I e ROC III, na base FRGC v2.0, foram incluídos para demonstrar a generalização do método. O experimento ROC III foi aplicado apenas à base de dados FRGC v2.0, uma vez que não pode ser aplicada na base Bosphorus.

Para o ROC I, foram utilizados os classificadores de LDA. Os resultados obtidos estão listados na Tabela 14 e referem-se ao fatiamento ST com os descritores 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd e 4HuYZd formando um vetor de característica de 308 longos. É importante mencionar que, mesmo que o LDA para o experimento ROC I tenha um número maior de descritores de entrada, internamente ele seleciona um número de atributos igual ao número de classes menos 1 (MARTINEZ; KAK, 2001).

No que diz respeito à base de dados Bosphorus, a Tabela 14 mostra que a taxa de verificação alcançada neste trabalho é superior àquelas relatadas por Elaiwat *et al.* (2015), Liu *et al.* (2013), Ocegueda *et al.* (2013) e Ming (2014). Ming (2014) também negligenciou amostras

Tabela 14 – Taxas de Verificação a 0.1% FAR. Experimento ROC I na base Bosphorus com VR e RR para Neutro vs Não Neutro; e experimento ROC I e ROC III na base FRGC v2.0 com VR para Neutro vs Não Neutro.

Databases	Bosphorus		FRGC v2.0	
	RR (%)	VR (%)	VR (%)	VR (%)
This work (LDA*)	98,7	98,9	97,1	94,7
This work (SVM**)	85,3	87,0	84,2	87,9
Elaiwat et al (ELAIWAT <i>et al.</i> , 2015)	n/a	91,0	98,0	97,8
Ocegueda et al (OCEGUEDA <i>et al.</i> , 2013)	n/a	93,8	98,1	97,9
Ming et al (MING, 2014)	92,0	94,0	n/a	n/a
Liu et al (LIU <i>et al.</i> , 2013)	n/a	95,6	93,1	n/a
Emambakhsh (EMAMBAKHSH; EVANS, 2017)	n/a	n/a	n/a	93,5
Al-Osaimi, Faisal R,(AL-OSAIMI, 2016)	n/a	n/a	97,8	94,1
Berretti et al (BERRETTI <i>et al.</i> , 2014)	n/a	n/a	91,4	86,6

com oclusão e variação de pose, mas incluiu amostras de IGN (apenas 18 amostras). Elaiwat *et al.* (2015) e Liu *et al.* (2013), em vez disso, incluíram amostras com oclusão e apresentaram variações nos experimentos. Ao fazer isso, foram usadas técnicas de preenchimento simétrico (ELAIWAT *et al.*, 2015) (LIU *et al.*, 2013) ou preenchimento de lacunas (LIU *et al.*, 2013). Liu *et al.* (2013) apenas negligenciaram as amostras rotacionadas em 90°. Embora os conjuntos de testes empregados em Elaiwat *et al.* (2015) e Liu *et al.* (2013) sejam muito abrangentes, os dois trabalhos precisam encontrar o plano de simetria facial para obter o preenchimento facial, o que acaba adicionando complexidade à sua solução. Além disso, o Elaiwat *et al.* (2015) usa o ICP para um alinhamento preciso. Nestes experimentos, o tamanho do vetor de característica é de 252 unidades de comprimento para os classificadores de LDA e SVM-SVC-rbf; O LDA foi resolvido pelo método SVD usando o pacote *sklearn* na linguagem python, e os parâmetros SVM foram definidos como  $C = 8$  e  $\gamma = 0,125$ .

Para avaliar a generalização do método proposto, as varreduras do conjunto de dados FRGC v2.0 foram usadas nas mesmas condições dos experimentos relatados até o momento para o banco de dados Bosphorus. Os resultados estão apresentados na Tabela 14. Mais uma vez, o método proposto funcionou bem, com taxas comparáveis às referências recentes importantes e chegando a ser uma das três melhores dentre elas. Foi relatado o experimento ROC I, como descrito anteriormente, e também um experimento de verificação de face ROC III, no qual o treinamento incluiu as varreduras das sessões de Outono de 2003, enquanto o conjunto de testes contém amostras das sessões realizadas na primavera de 2004. Devido a esse lapso temporal entre a aquisição das amostras de teste e as amostras de treinamento, esta experiência foi considerada

a mais difícil do protocolo FRGC, conforme apontado por Berretti *et al.* (2014).

Recapitulando, para a classificação 1, foi utilizado somente o classificador SVM-SVC-rbf, onde os parâmetros foram definidos como  $C = 8$  e  $\gamma = 0.125$ . Esta experiência foi especialmente importante para colocar o método proposto entre aqueles que usam a estratégia de treinamento/teste. Uma grande vantagem dessa comparação é que ela pode ser detalhada, uma vez que a literatura fornece taxas de reconhecimento para o subconjunto de amostras não neutras. Permite avaliar mais especificamente cada método em relação à expressões emocionais, amostras neutras, LFAU, UFAU e CAU.

Os resultados mostrados na Tabela 15 levam em conta apenas os slicings ST com momentos 4HuXYd, 4HuXZd e 4HuYZd, para um total de 168 descritores. Em alguns casos, o método proposto neste trabalho atingiu ou superou os melhores resultados relatados por Berretti *et al.* (2014), Li *et al.* (2015), Ming (2014), Emambakhsh e Evans (2017) e Al-Osaimi (2016). Esses casos estão destacados em negrito. Também para o conjunto de expressões Raiva e Repugnância o método teve um desempenho melhor do que os resultados de Li *et al.* (2015). Também superamos Berretti *et al.* (2014) para amostras neutras, medo, tristeza, surpresa e CAU.

Além dos resultados, alguns aspectos conceituais e práticos devem ser considerados neste momento. Ao contrário do método de Berretti (BERRETTI *et al.*, 2014), que não é direcionado para câmeras de baixo custo ou de baixa resolução, como o Kinect, o método avaliado adaptou-se bem com este sistema de aquisição. Foi realizado o experimento ROC-III na base de dados 3DMAD (ERDOGMUS; MARCEL, 2013), que possui amostras baseadas no Kinect. As taxas de reconhecimento chegaram a 99% utilizando como treinamento as amostras da Parte 1 e como testes as amostras da Parte 2 da base, cujas amostras foram adquiridas com semanas de diferença. Isso sugere o quanto robusto é o método para diferentes sistemas de aquisição de dados. Além disso, como mencionado anteriormente, o presente método tem uma etapa de regularização da densidade da nuvem baseada na grade do voxel e isso é especialmente importante para superar as diferenças na resolução do scanner.

Por fim, Li *et al.* (2015) estenderam a estrutura de correspondência SIFT para dados de malha e propuseram uma nova abordagem usando a combinação refinada de descritores de ponto-chave 3D, denominada Histograma de Quantidades Diferenciais de Superfície Múltiplas (HOMQ) combinando HOG, HOS e HOGS no nível de recursos. Embora Li *et al.* (2015) tenham relatado os melhores desempenhos para conjuntos não neutros, a grande quantidade de pontos-chave necessária resultou em esforços computacionais significativos. Esse método leva

Tabela 15 – RR para categorias distintas de conjuntos de teste na base de dados Bosphorus, experimento rank-1. Ming (2014), Emambakhsh e Evans (2017) e Al-Osaimi (2016)

Descriptors(#)	-	-	130	-	-	168
Key Points(#)	648	145	2	-	-	1
Probes(#)	Li et al	Berreti et al	Ming et al	Emambakhsh et al	Al-Osaimi et al	This work (SVM)
<b>Bosphorus (rank1 RR(%))</b>						
Neutral (194)	100	98,5	94,3	99,0	98,5	<b>99,0</b>
NonNeutral (2603)	98,7	96,6	92,0	n/a	92,4	95,4
Emotions (453)	96,7	91,2	n/a	96,2	n/a	<b>95,2</b>
Anger (71)	97,2	88,7	n/a	94,1	n/a	<b>98,6</b>
Disgust (69)	86,9	81,2	n/a	88,2	n/a	<b>98,6</b>
Fear (70)	98,6	91,4	n/a	98,6	n/a	<b>95,7</b>
Happy (106)	98,1	94,3	n/a	98,1	n/a	87,7
Sad (66)	100	95,5	n/a	96,9	n/a	<b>98,5</b>
Surprise (71)	98,6	94,4	n/a	100	n/a	95,8
LFAU (1549)	98,8	97,5	n/a	n/a	n/a	94,7
UFAU (432)	100	99,1	n/a	n/a	n/a	98,4
CAU (169)	100	96,4	n/a	n/a	n/a	<b>98,8</b>

em média 1,25 minutos em um PC com Intel 950, processador de 3,07 GHz, 8 GB de RAM. O método aqui apresentado, sob condições semelhantes (processador Intel i7-5500U, 2,4 GHz) e empregando bibliotecas OpenCv e PCL para C / C ++, levou menos de 4 segundos para executar todo o processo de reconhecimento.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho é proposta uma técnica para fatiamento tri-axial da nuvem de pontos que representam faces 3D, bem como o uso de momentos 2D e 3D adaptados para realizar o reconhecimento facial. A base de dados Bosphorus foi a principal fonte de dados utilizada para validação do método, tanto pela sua complexidade, quanto por sua completude. O método foi aplicado em outra base de dados (FRGC v2.0), com características parecidas com a base Bosphorus, para demonstrar sua generalização. A técnica demonstrou que os cortes nas direções Sagitais e Transversais combinados fornecem os melhores resultados e que estes estão no nível do estado da arte, igualando-se em alguns e superando-os outros. Além da simplicidade na forma de realizar o fatiamento, utilizando apenas um ponto chave, o tempo total de processamento por amostra está entre os mais rápidos. As bases citadas anteriormente, utilizam sensores de alta resolução, e por isso foi realizado um experimento ROC III na base de dados 3DMAD com ótima taxa de acerto, mostrando que além de funcionar em mais uma base, a técnica não se limita ao uso de sensores de alta resolução.

A combinação dos atributos 4HuXYd, 4HuYZ, 4HuYZd, 4HuXYd, 3DIT e 3DITd atingiram uma taxa de verificação de 98,7% para 0,1% FAR no ROC.

Para o rank 1, usando 4HuXYd, 4HuXZd e 4HuYZd, foram alcançadas as taxas de identificação de 99% e 95.4% para os experimentos Neutro vs Neutro e Neutro vs Não Neutro, respectivamente, o que reforça que o uso da distância da função melhora a caracterização de indivíduos. Taxas de identificação acima de 97% para o Neutro vs Não-Neutro nos subconjuntos CAU, UFAU, Raiva, Repugnância e Tristeza foram obtidas.

A técnica requer somente a identificação de um único ponto-chave para que se possa segmentar a face em 14 diferentes cortes, contendo subconjuntos das nuvens de pontos, nos planos Sagital e Transversal, o que a torna computacionalmente menos complexa do que outras abordagens da mesma natureza.

Os cortes Sagital e Transversal mostraram-se menos sensíveis à variações superficiais mesmo para certos tipos de expressões, enquanto os cortes frontais, na grande maioria dos casos, degradaram a taxa de reconhecimento quando associados aos outros dois planos ortogonais.

Muitos métodos atuais utilizan-se de um grande número de pontos chave na face, baseados ou não nas FAC(*Face Action Units*), outros métodos utilizam curvas sagitais ou superfícies frontais como atributos, entretanto, a combinação de superfícies em mais de uma direção é uma contribuição deste trabalho. Outra contribuição é o uso de momentos de Hu para

nuvens de ponto, uma vez que eles foram criados e são largamente utilizados em imagens 2D. E desta combinação de fatores surgiu mais um método a ser explorado e adaptado no futuro.

Apesar de satisfatório para o objetivo proposto, o método aplicado neste trabalho pode ser aprimorado. Algumas sugestões para trabalhos futuros incluem abordadar as seguintes limitações:

- a) oclusão e auto-oclusão: embora tenham sido negligenciadas na versão atual, a simetria da face humana pode ser explorada para lidar com essa questão;
- b) a dependência do alinhamento inicial: o método pode ser executado mais rapidamente à medida que novas abordagens para a correção da orientação da face sejam disponibilizadas. É importante ressaltar que a essência do método abordado neste trabalho depende da proposta de um novo descritor de atributos; e,
- c) a dependência da densidade das nuvens: os momentos Hu são sensíveis à variação da densidade das nuvens de pontos. Este assunto foi abordado na versão atual, aplicando o algoritmo voxel-grid a fim de equalizar a densidade para todas as amostras. Entretanto, outras estratégias podem ser avaliadas.

## REFERÊNCIAS

- MOUSAVI, Mir Hashem and Faez, Karim and Asghari, Amin. [S.l.: s.n.].
- ABUZNEID, M. A.; MAHMOOD, A. Enhanced human face recognition using lbp descriptor, multi-knn, and back-propagation neural network. **IEEE Access**, v. 6, p. 20641–20651, 2018.
- AHDID, R.; SAID, S.; FAKIR, M.; MANAUT, B.; OUADID, Y.; TAIFI, K. Three dimensional face surface recognition by geodesic distance using jacobi iterations. In: **2017 14th International Conference on Computer Graphics, Imaging and Visualization**. [S.l.: s.n.], 2017. p. 44–48.
- AHDID, R.; TAIFI, K.; SAID, S.; FAKIR, M.; MANAUT, B. Automatic face recognition system using iso-geodesic curves in riemanian manifold. In: **2017 14th International Conference on Computer Graphics, Imaging and Visualization**. [S.l.: s.n.], 2017. p. 73–78.
- AL-OOSAIMI, 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, feb 2016. ISSN 10577149. Disponível em: <<http://ieeexplore.ieee.org/document/7300438/>>.
- ALEHDAGHI, M.; ESFAHANI, M. A.; HARATI, A. Parallel ransac: Speeding up plane extraction in rgbd image sequences using gpu. In: **2015 5th International Conference on Computer and Knowledge Engineering (ICCKE)**. [S.l.: s.n.], 2015. p. 295–300.
- AMEUR, B.; MASMOUDI, S.; DERBEL, A. G.; HAMIDA, A. B. Fusing gabor and lbp feature sets for knn and src-based face recognition. In: **2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)**. [S.l.: s.n.], 2016. p. 453–458.
- BERRETTI, S.; WERGHI, N.; BIMBO, A. del; PALA, P. Selecting stable keypoints and local descriptors for person identification using 3d face scans. **The Visual Computer**, v. 30, n. 11, p. 1275–1292, Nov 2014. ISSN 1432-2315.
- BESL, P. J.; MCKAY, N. D. A method for registration of 3-d shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 2, p. 239–256, Feb 1992. ISSN 0162-8828.
- BEČKA, M.; OKŠA, G.; VAJTERŠIC, M.; GRIGORI, L. On iterative qr pre-processing in the parallel block-jacobi svd algorithm. **Parallel Computing**, v. 36, n. 5, p. 297 – 307, 2010. ISSN 0167-8191. Parallel Matrix Algorithms and Applications.
- BORRMANN, D.; ELSEBERG, J.; LINGEMANN, K.; NÜCHTER, A. The 3d hough transform for plane detection in point clouds: A review and a new accumulator design. **3D Research**, v. 2, n. 2, p. 3, Nov 2011. ISSN 2092-6731.
- CHEN, Y.; MEDIONI, G. Object modeling by registration of multiple range images. In: **Proceedings. 1991 IEEE International Conference on Robotics and Automation**. [S.l.: s.n.], 1991. p. 2724–2729 vol.3.
- CHENG, Z.; SHI, T.; CUI, W.; DONG, Y.; FANG, X. 3d face recognition based on kinect depth data. In: **2017 4th International Conference on Systems and Informatics (ICSAI)**. [S.l.: s.n.], 2017. p. 555–559.

- CORROCHANO, E. B. **Handbook of Geometric Computing.** [S.l.]: Springer, 2005.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.
- DIEDERICH, S.; LENTSCHIG, M.; OVERBECK, T.; WORMANN, D.; HEINDEL, W. Detection of pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs and single-image reporting. **European Radiology**, Springer-Verlag, v. 11, n. 8, p. 1345–1350, aug 2001. ISSN 0938-7994. Disponível em: <<http://link.springer.com/10.1007/s003300000787>>.
- EKMAN, P.; FRIESEN, W. V. Measuring facial movement. **Environmental psychology and nonverbal behavior**, v. 1, n. 1, p. 56–75, 1976. ISSN 1573-3653.
- ELAIWAT, S.; BENNAMOUN, M.; BOUSSAID, F.; EL-SALLAM, A. A Curvelet-based approach for textured 3D face recognition . **Pattern Recognition**, v. 48, n. 4, p. 1235 – 1246, 2015. ISSN 0031-3203.
- EMAMBAKHSH, M.; EVANS, A. Nasal Patches and Curves for Expression-Robust 3D Face Recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 5, p. 995–1007, may 2017. ISSN 01628828.
- ERDOGMUS, N.; MARCEL, S. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In: . [S.l.: s.n.], 2013.
- FISCHLER, M. A.; BOLLES, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. v. 24, p. 381–395, 06 1981.
- FUJIWARA, T.; KAMEGAWA, T.; GOFUKU, A. Plane detection to improve 3d scanning speed using ransac algorithm. In: **2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)**. [S.l.: s.n.], 2013. p. 1863–1869. ISSN 2156-2318.
- GAO, B.; LIU, S.; ZHANG, J.; ZHANG, B. Pose estimation algorithm based on improved ransac with an rgb-d camera. In: **2018 Chinese Control And Decision Conference (CCDC)**. [S.l.: s.n.], 2018. p. 5024–5029.
- GUAN, W.; LI, W.; REN, Y. Point cloud registration based on improved icp algorithm. In: **2018 Chinese Control And Decision Conference (CCDC)**. [S.l.: s.n.], 2018. p. 1461–1465.
- GUENTHER, D.; LEUPERS, R.; ASCHEID, G. A scalable, multimode svd precoding asic based on the cyclic jacobi method. **IEEE Transactions on Circuits and Systems I: Regular Papers**, v. 63, n. 8, p. 1283–1294, Aug 2016. ISSN 1549-8328.
- HARIRI, W.; TABIA, H.; FARAH, N.; BENOUARETH, A.; DECLERCQ, D. 3D face recognition using covariance based descriptors. **Pattern Recognition Letters**, Elsevier Science Inc., New York, NY, USA, v. 78, n. C, p. 1–7, jul 2016. ISSN 01678655. Disponível em: <<https://doi.org/10.1016/j.patrec.2016.03.028><http://linkinghub.elsevier.com/retrieve/pii/S0167865516300320>>.
- HATEM, H.; BEIJI, Z.; MAJEDD, R. A survey of feature base methods for human face detection. **International Journal of Control and Automation**, v. 8, n. 5, p. 61–78, 2015.
- HU, M.-K. Visual pattern recognition by moment invariants. **IRE Transactions on Information Theory**, v. 8, n. 2, p. 179–187, February 1962. ISSN 0096-1000.

- HUANG, H. b.; MU, Z. c.; ZENG, H.; ZHANG, B. q. 3d facial point cloud preprocessing based on skin color detection using svm. In: **2014 10th International Conference on Natural Computation (ICNC)**. [S.l.: s.n.], 2014. p. 524–528. ISSN 2157-9555.
- JAFRI, R.; ARABNIA, H. R. A Survey of Face Recognition Techniques. **Journal of Information Processing Systems**, v. 5, n. 2, p. 41–68, 2009. ISSN 1976-913X.
- KHADHRAOUI, T.; BENZARTI, F.; AMIRI, H. Multimodal hybrid face recognition based on score level fusion using Relevance Vector Machine. In: **2014 IEEE/ACIS 13th International Conference on Computer and Information Science (ICIS)**. IEEE, 2014. p. 211–215. ISBN 978-1-4799-4860-4. Disponível em: <<http://ieeexplore.ieee.org/document/6912136/>>.
- 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, Jun 2015. ISSN 1573-1405.
- LI, X.; KOMULAINEN, J.; ZHAO, G.; YUEN, P.-C.; PIETIKÄINEN, M. Generalized face anti-spoofing by detecting pulse from face videos. In: **2016 23rd International Conference on Pattern Recognition (ICPR)**. [S.l.: s.n.], 2016. p. 4244–4249.
- LIAO, S.; ZHU, X.; LEI, Z.; ZHANG, L.; LI, S. Z. Learning multi-scale block local binary patterns for face recognition. In: LEE, S.-W.; LI, S. Z. (Ed.). **Advances in Biometrics**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 828–837. ISBN 978-3-540-74549-5.
- LIU, P.; WANG, Y.; HUANG, D.; ZHANG, Z.; CHEN, L. Learning the spherical harmonic features for 3-d face recognition. **IEEE Transactions on Image Processing**, v. 22, n. 3, p. 914–925, March 2013. ISSN 1057-7149.
- MARTINEZ, A. M.; KAK, A. C. Pca versus lda. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 2, p. 228–233, Feb 2001. ISSN 0162-8828.
- MARVADI, D.; PAUNWALA, C.; JOSHI, M.; VORA, A. Comparative analysis of 3d face recognition using 2d-pca and 2d-lda approaches. In: **2015 5th Nirma University International Conference on Engineering (NUiCONE)**. [S.l.: s.n.], 2015. p. 1–5.
- MING, Y. Rigid-area orthogonal spectral regression for efficient 3d face recognition. **Neurocomputing**, v. 129, p. 445 – 457, 2014. ISSN 0925-2312.
- MOON, H.-M.; CHOI, D.; KIM, P.; PAN, S. B. Lda-based face recognition using multiple distance training face images with low user cooperation. In: **IEEE. Consumer Electronics (ICCE), 2015 IEEE International Conference on**. [S.l.], 2015. p. 7–8.
- MOUSAVI, M. H.; FAEZ, K.; ASGHARI, A. Three dimensional face recognition using svm classifier. In: **IEEE. Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on**. [S.l.], 2008. p. 208–213.
- OCEGUEDA, O.; FANG, T.; SHAH, S. K.; KAKADIARIS, I. A. 3d face discriminant analysis using gauss-markov posterior marginals. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 35, n. 3, p. 728–739, March 2013. ISSN 0162-8828.
- OCEGUEDA, O.; PASSALIS, G.; THEOHARIS, T.; SHAH, S. K.; KAKADIARIS, I. A. Ur3d-c: Linear dimensionality reduction for efficient 3d face recognition. In: **2011 International Joint Conference on Biometrics (IJCB)**. [S.l.: s.n.], 2011. p. 1–6.

PHILLIPS, P. J.; FLYNN, P. J.; SCRUGGS, T.; BOWYER, K. W.; CHANG, J.; HOFFMAN, K.; MARQUES, J.; MIN, J.; WOREK, W. Overview of the face recognition grand challenge. In: **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**. [S.l.: s.n.], 2005. v. 1, p. 947–954 vol. 1. ISSN 1063-6919.

PHILLIPS, P. J.; FLYNN, P. J.; SCRUGGS, T.; BOWYER, K. W.; CHANG, J.; HOFFMAN, K.; MARQUES, J.; MIN, J.; WOREK, W. Overview of the face recognition grand challenge. In: **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**. [S.l.: s.n.], 2005. v. 1, p. 947–954 vol. 1. ISSN 1063-6919.

PHILLIPS, P. J.; FLYNN, P. J.; SCRUGGS, T.; BOWYER, K. W.; CHANG, J.; HOFFMAN, K.; MARQUES, J.; MIN, J.; WOREK, W. Overview of the face recognition grand challenge. **Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005**, I, p. 947–954, 2005. ISSN 1063-6919.

RATYAL, N. I.; TAJ, I. A.; BAJWA, U. I.; SAJID, M.; BAIG, M. J. A.; BUTT, F. M. 3D face recognition based on region ensemble and hybrid features. In: **2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)**. IEEE, 2016. p. 294–300. ISBN 978-1-5090-1252-7. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7495241>>.

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.

SALAMA, A.; MALEKMOHAMMADI, A.; MOHANNA, S. Single-ended electrical impedance tomography. **The Journal of Engineering**, v. 2018, n. 4, p. 173–185, 2018.

SANTOS, A. M.; de Carvalho Filho, A. O.; SILVA, A. C.; PAIVA, A. C. de; NUNES, R. A.; GATTASS, M. Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. **Engineering Applications of Artificial Intelligence**, Pergamon, v. 36, p. 27–39, nov 2014. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197614001754>>.

SAVRAN, A.; ALYÜZ, N.; DİBEKLIOĞLU, 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.

SCHNABEL, R.; WAHL, R.; KLEIN, R. Efficient ransac for point-cloud shape detection. v. 26, p. 214–226, 06 2007.

SCHUMACHER, D.; SHARMA, R.; GRAGER, J.-C.; SCHRAPP, M. Scatter and beam hardening reduction in industrial computed tomography using photon counting detectors. **Measurement Science and Technology**, v. 29, n. 7, p. 075101, 2018. Disponível em: <<http://stacks.iop.org/0957-0233/29/i=7/a=075101>>.

SIDDQUI, T. A.; BHARADWAJ, S.; DHAMECHA, T. I.; AGARWAL, A.; VATSA, M.; SINGH, R.; RATHA, N. Face anti-spoofing with multifeature videolet aggregation. In: **2016 23rd International Conference on Pattern Recognition (ICPR)**. [S.l.: s.n.], 2016. p. 1035–1040.

- SKARŻYŃSKI Łukasz; SUCHORZEWSKI, J. Mechanical and fracture properties of concrete reinforced with recycled and industrial steel fibers using digital image correlation technique and x-ray micro computed tomography. **Construction and Building Materials**, v. 183, p. 283 – 299, 2018. ISSN 0950-0618. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S095006181831585X>>.
- SOLTANPOUR, S.; BOUFAMA, B.; WU, Q. J. A survey of local feature methods for 3d face recognition. **Pattern Recogn.**, v. 72, n. C, p. 391–406, dez. 2017. ISSN 0031-3203.
- SUK, T.; FLUSSER, J.; BOLDYŠ, J. 3d rotation invariants by complex moments. **Pattern Recognition**, v. 48, n. 11, p. 3516 – 3526, 2015. ISSN 0031-3203.
- THANGARAJ, S.; PERIYASAMY, V.; BALAJI, R. Retinal vessel segmentation using neural network. **IET Image Processing**, v. 12, n. 5, p. 669–678, 2018. ISSN 1751-9659.
- TWIGG, C. D.; KAVCIC-ALESIC, Z. Point cloud glue: Constraining simulations using the procrustes transform. In: **Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation**. Goslar Germany, Germany: Eurographics Association, 2010. (SCA '10), p. 45–54. Disponível em: <<http://dl.acm.org/citation.cfm?id=1921427.1921435>>.
- TZALAVRA, A.; DALAKLEIDI, K.; ZACHARAKI, E. I.; TSIAPARAS, N.; CONSTANTINIDIS, F.; PARAGIOS, N.; NIKITA, K. S. Comparison of multi-resolution analysis patterns for texture classification of breast tumors based on dce-mri. In: SPRINGER. **International Workshop on Machine Learning in Medical Imaging**. [S.l.], 2016. p. 296–304.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: **Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001**. [S.l.: s.n.], 2001. v. 1, p. I-511–I-518 vol.1. ISSN 1063-6919.
- WANG, X.; RUAN, Q.; JIN, Y.; AN, G. 3d face recognition using closest point coordinates and spherical vector norms. In: **6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)**. [S.l.: s.n.], 2015. p. 192–196.
- WANG, X.; ZAMBRENO, J. An fpga implementation of the hestenes-jacobi algorithm for singular value decomposition. In: **2014 IEEE International Parallel Distributed Processing Symposium Workshops**. [S.l.: s.n.], 2014. p. 220–227.
- YAMASHITA, Y.; TAN, J. K.; ISHIKAWA, S. Human motion description and recognition under arbitrary motion direction. In: **2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)**. [S.l.: s.n.], 2017. p. 110–115.
- YIN, L.; WEI, X.; SUN, Y.; WANG, J.; ROSATO, M. J. A 3d facial expression database for facial behavior research. In: **7th International Conference on Automatic Face and Gesture Recognition (FGR06)**. [S.l.: s.n.], 2006. p. 211–216.
- YIN, L.; WEI, X.; SUN, Y.; WANG, J.; ROSATO, M. J. A 3d facial expression database for facial behavior research. In: **Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition**. [S.l.: s.n.], 2006. (FGR '06), p. 211–216. ISBN 0-7695-2503-2.

ZHAO, W.; CHELLAPPA, R.; PHILLIPS, P. J.; ROSENFELD, A. Face recognition: A literature survey. **ACM computing surveys (CSUR)**, ACM, v. 35, n. 4, p. 399–458, 2003.

ZHOU, H.; MIAN, A.; WEI, L.; CREIGHTON, D.; HOSSNY, M.; NAHAVANDI, S. Recent advances on singlemodal and multimodal face recognition: A survey. **IEEE Transactions on Human-Machine Systems**, v. 44, n. 6, p. 701–716, 2014. ISSN 21682291.

**ANEXO A – ARTIGO ACEITO NA IEEE RA-L**

# Tri-axial slicing for 3D face recognition from adapted rotational invariants spatial moments and minimal keypoints dependence

Robson S. Siqueira<sup>1</sup>, Gilderlane R. Alexandre<sup>2</sup>, José M. Soares<sup>3</sup> and George A. P. Thé<sup>3</sup>

**Abstract**—This work presents a multiple slicing model for 3D images of human face, using the Frontal, Sagittal and Transverse orthogonal planes. The definition of the segments depends on just one key point, the nose tip, which makes it simple and independent of the detection of several key points. For facial recognition, attributes based on adapted 2D spatial moments of Hu and 3D spatial Invariant Rotation Moments are extracted from each segment. Tests with the proposed model using the Bosphorus Database for neutral vs non-neutral ROC I experiment, applying Linear Discriminant Analysis as classifier and more than one sample for training, achieved 98.7% of verification rate at 0.1% of false acceptance rate. By using the Support Vector Machine as classifier the rank1 experiment recognition rates of 99% and 95.4% have been achieved for a neutral vs neutral and for a neutral vs non-neutral, respectively. These results approach the state-of-the-art using Bosphorus Database and even surpasses it when Anger and Disgust expressions are evaluated. In addition, we also evaluate the generalization of our method using the FRGC v2.0 database and achieve competitive results, making the technique promising, especially for its simplicity.

**Index Terms**—Computer Vision for Automation, Recognition, Surveillance Systems

## I. INTRODUCTION

THE traditional 2D images have been widely used as data source for Face Recognition (FR) during the past decades and, inspite of that, 2D FR remains with important constraints. The major challenges associated to this task include strong inter-subject facial similarities and intra-subject variations [1]. The latter is greatly affected by illumination changes, variations in pose and expressions, variability of the background and occlusions [2], [3].

3D FR approaches have recently gained popularity as increasingly better 3D sensors have become available. 3D data provide more reliable geometric information and allow the

Manuscript received: February, 24, 2018; Revised May, 19, 2018; Accepted June, 19, 2018.

This paper was recommended for publication by Editor Wan Kyun Chung upon evaluation of the Associate Editor and Reviewers' comments.

This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

<sup>1</sup>Robson da Silva Siqueira is professor with Faculty of Electrical Engineering, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Maracanaú CE Brazil siqueira@ifce.edu.br

<sup>2</sup>Gilderlane Ribeiro Alexandre master student with the Department of Teleinformatics Engineering, Universidade Federal do Ceará, Fortaleza CE Brazil gilderlane.ribeiro@gmail.com

<sup>3</sup>José Marques Soares and George A. P. Thé are PhD professor with the Department of Teleinformatics Engineering, Universidade Federal do Ceará, Fortaleza CE Brazil marques, george.the@ufc.br

Digital Object Identifier (DOI): see top of this page.

extraction of certain features, which are resilient to changes in scale, rotation, and illumination. That capability of preserving geometric information in addition to powerful feature extraction methods allow overcoming degradation conditions, representing a clear advantage over 2D techniques [4].

On the other hand, 3D techniques are not as mature and well-established as 2D approaches are. Furthermore, some information may be more trivial to be extracted from 2D data than from 3D data (e.g., automatic landmark detection). For that reason, hybrid approaches that combine two types of data sources, 2D and 3D, are also employed.

Regarding the feature extraction methods, three main categories are addressed in the literature: holistic, feature-based and hybrid matching methods. Holistic methods find a set of global features from the entire 3D model while feature-based methods focus on the extraction of local features from the face or from regions of the face. Liu et al. [5] employ a holistic method by using spherical harmonic features. Principal Component Analysis (PCA) is also largely employed in this category. Hybrid matching methods combine holistic and feature-based approaches [4].

While holistic methods focus on finding overall similarities of faces, local-feature methods tend to be robust to partial models and occlusions, and are more suitable for matching, identification and verification purposes [4]. 3D local-feature based methods are divided into three groups, according to the descriptors: keypoint-based, curve-based and local surface-based methods.

3D keypoint-based methods rely on the detection of salient points of face and the description of a feature vector. According to [4], although these methods can deal with partial models and occlusions, since they use a large number of keypoints and consequently high dimensional feature vectors, they are computationally demanding. Indeed, recent works using keypoint-based methods compute a large number of keypoints in order to obtain satisfactory recognition rates, as reported in the literature [1], [6], [7], [8].

Curve-based methods, as the one employed in [9], use curves traced on faces according to certain criteria (e.g., contour, profile) and extract geometrical information from different areas of the face. That makes them more robust against facial expressions than keypoint-based methods [4].

Local surface-based methods, in turn, extract geometric information from several regions of the surface face that are invariant to facial expressions and robust to that condition [4]. Examples include Ocegueda et al. in [10], which addresses the

3D FR problem by analyzing probabilistically the relevance of a certain region of the face for the current expression and considering the most discriminative regions for feature extraction.

In [11], Emambakhsh and Evans investigate the effect of the nasal region for expression robust 3D FR. Their method relies on the nose landmarking and performs the feature extraction based on surface normals over patches and curves on the nasal region to finally find the most stable patches under variations of expressions.

In [12], multiple keypoints are employed to derive 3D surface patches. In that paper, Al-Osaimi presents rotation invariant descriptor extraction techniques applicable for FR on 3D static images or 3D videos and does not address the partial occlusions of the local surface.

The approach proposed in the present work offers a novel strategy for the processing of 3D images based on tri-axial slicing with a minimal dependence on keypoints. In a nutshell, this technique includes a unique keypoint detection and the definition of surface regions between three sets of parallel planes, so called Frontal, Sagittal and Transverse. Planes are positioned in space based on the keypoint and attributes are then extracted from each surface region. The main contributions of this work are (1) the proposal of adapted Hu moments, originally applied in the 2D domain, (2) their extraction from surface regions along the three orthogonal axes and combinations (the literature usually considers single axis) and (3) minimal keypoint dependence. As a consequence, we achieved classification performance comparable to [1], in shorter times than those reported in that paper and in many of those referenced in the review [4], as well.

## II. METHOD

### A. Tri-axial Face Slicing

Our method defines well-marked and reproducible regions, based on a unique and easily identifiable keypoint and extracts robust local descriptors to identify individuals using only 3D point clouds. We define regions of the face and descriptors from those regions that are able to identify the individuals even under large variation of facial expressions. We propose the slicing of the face by using planes orthogonal to each other, positioned based on one keypoint, in such manner that considerable surface changes caused by expressions concentrate in some of the regions but are minimal in others.

The intuition behind our method for feature extraction relates to the nature of deformations on the face surface under expression variations. Different patches suffer different deformation profiles along preferable directions while facial expression changes. We have observed that behaviour after estimating deformation as done in the field of nonrigid registration of 3D medical images [13]; from the eigenvalues of the covariance tensor of the 3D data, we proceeded to calculate anisotropy, omnidirectionality and other geometric features, and then investigated how these numbers vary among individuals and among different patches of a given individual. Result of this analysis revealed that some patches are less sensitive to deformation than others, thus confirming our

expectation that patches that do not suffer great deformations along one or more of the three directions are robust enough to yet represent individuals.

The essence of the technique lies in point cloud slicing of a face image along the three spatial axes, giving rise to what is called Frontal, Sagittal and Transverse slices, as shown in Figure 1. The method developed here requires some pre-processing steps: first, the point cloud has to undergo density normalization; this is followed by outliers removal; next by the keypoint detection, then a cropping is carried out based on the keypoint found; finally, a pose correction to align faces.

There are a number of reasons to perform density normalization as the very first pre-processing step. One is to decrease the number of points in a cloud, making the processing in the next steps faster. The other reason, and not less important, relates to the type of descriptors that are to be extracted here, since they can be sensitive to intra-subject and inter-subject density variations. The idea we follow here is to decrease the density of the point clouds until all of them have the same density of points. Ideally, a good acquisition apparatus should provide images with nearly the same point density, however this is not observed in practice. We have seen samples with denser clouds which probably suggests the individual getting closer to the camera in some sessions. Those density variations can occur either for samples of the same subject or for distinct ones. Depending on the type of descriptors employed, the recognition process can be insensitive to that variation or be significantly affected by it. It is important to notice that the decrease of density can not be extreme, so the faces become uncharacterized, nor minimal, so the method does not benefit from it. Our technique uses the Point Cloud Library (PCL) 3D voxel grid algorithm, presented by Rusu and Cousins [14], to perform density normalization with a voxel leaf of 2 millimeters for all 3 dimensions.

Outliers removal is the most common step among methods that perform surface analysis, since outliers may cause peaks that uncharacterize facial surface. In our case, they can add noise to the face descriptors. Although outliers are prevalent in occluded or self-occluded images, in general outliers from any sources can prevent the detection of important keypoints, such as the nose tip.

Nose tip detection is performed by taking the closest point to the sensor (depending on the reference system adopted, it can be also considered the most distant point from the sensor). One reason why outliers removal is important in this case, is that outliers close to the sensor can compete with the actual nose tip.

Before pose correction, a cropping is carried out on the point cloud; starting from the nose tip, it extends up to an Euclidean distance  $d = 80$  mm just enough to comprise the regions of the eyes, nose and mouth. Face cropping is performed by several authors, including [6], [7], [1] to restrict the region of the face that best characterizes the individual.

Pose correction with Iterative Closest Point (ICP) alignment is applied in order to iteratively correct the pose of the 3D faces. The probe faces are aligned with respect to the gallery faces as in Elaiwat et al. [6]. Elaiwat et al. [6] employ ICP algorithm only for the samples with great rotation and use PCA

for the others, since ICP, being iterative, can make the whole procedure slower. In our work, we mitigate this drawback by reducing the cloud density and cropping the face around the nose, as well. We also limited the number of iterations to 100. In addition, we tested 3 possibilities of pose correction, two other methods for pose correction that aimed to find the face normal and align it with z-axis: RANSAC-Plane [15], finding the best plane that fits the face and using its normal for pose correction; and two-sided block-Jacobi SVD method [16] to calculate the face normal. None of them beat ICP. Two-sided block-Jacobi SVD is fast and highly parallelizable [16] while RANSAC-Plane has approximately the same processing time as ICP. The best arrangement we found was first perform a coarse alignment with two-sided block-Jacobi SVD and then apply ICP for fine correction, which minimizes the number of iterations required. The maximum number of iterations was set to 100, since superior values had no considerable effect on alignment.

Pose correction is an important step to mitigate head inclination during image acquisition sessions, thus preventing bad formation and indexation of slices. It allows the keypoint corresponding to the nose tip to be found ; in the following it is referred to as  $\nu = (x_n, y_n, z_n)$ . For each neutral or non-neutral sample of an individual expression, the point  $\nu$  must be coincident or be at a minimal acceptable distance relative to the other samples.

The slicing is carried out in such a way that each slice contains only the points  $p = (x, y, z)$  of the C point cloud lying between two planes of adjacent cut-sections (these are equally spaced along axes). Reference planes for the Frontal slices are those parallel to the  $XY$  plane of Figure 1 and the slice thickness is  $c_f$ ; reference planes for the Sagittal and Transverse slices are, in turn, those parallel to the  $YZ$  and  $XZ$  planes of Figure 1, and the respective thicknesses are  $c_s$  and  $c_t$ . Thus the triple  $\mathbf{c} = (c_f, c_s, c_t)$  defined in millimeters, along with the coordinates of the nose tip,  $x_k, y_k$  and  $z_k$  define the indices  $n$  of Frontal, Sagittal and Transverse slices as shown in Figure 1.

The Frontal, Sagittal and Transverse sections are represented respectively by the sets of points  $F_n, S_n$  and  $T_n, n \in Z$ , as shown in Equations 1, 2 and 3 below:

$$p \in F_n \rightarrow \left\lfloor \frac{|z_k - z|}{c_f} \right\rfloor = n \quad (1)$$

$$p \in S_n \rightarrow \left\lfloor \frac{|x_k - x|}{c_s} \right\rfloor \cdot sign(x_k - x) = n \quad (2)$$

$$p \in T_n \rightarrow \left\lfloor \frac{|y_k - y|}{c_t} \right\rfloor \cdot sign(y_k - y) = n. \quad (3)$$

### B. Extraction of Attributes

The combination of Hu Moments [17] and 3D Invariant Moments [18] can provide robust local descriptors to characterize face, even if there are rotations or translations between samples of a given individual acquired at different time instants. For FR, for example, the distance to the sensor may change significantly, in addition to rotations, even if minimal, around

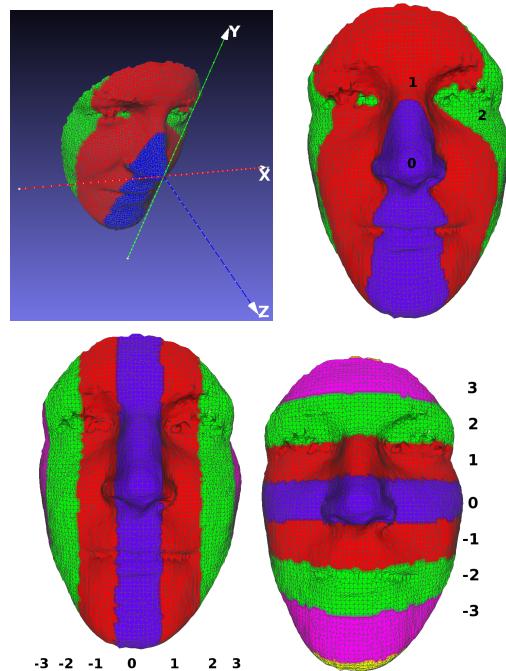


Fig. 1: Nose tip with respect to the  $XYZ$  axes and the slices in the three orthogonal planes, Frontal, Sagittal and Transverse. This example corresponds to a Neutral sample considering  $\mathbf{c} = (c_f, c_s, c_t) = (20, 20, 20)$  in millimeters and a 80-mm wide cropping centered at the nose tip. Here the blue color represents the segment indexed as zero. Note: the visualization above as a surface is only for illustration.

the three orthogonal axes. This requires the used descriptors to be invariant to these transformations. The spatial moments are defined in [17] and consider the  $xy$ -axis as the dependent terms. Equations 4-6 present variations of it for each pair of axis.

$$M_{pq}^{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (4)$$

$$M_{pq}^{XZ} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p z^q f(x, z) dx dz \quad (5)$$

$$M_{pq}^{YZ} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^p z^q f(y, z) dy dz \quad (6)$$

In the algorithm used here, the feature vector contains the Hu moments calculated from each spatial orientation. Thus, in the case of  $M^{XY}$ , the configuration that considers the  $z$ -axis as independent for the calculation of the 7 Hu moments is represented as HuXY. The discrete and centralized version of Equation 4 for the HuXY configuration duly normalized and with the independent term is shown in Equation 7 , where  $x$  and  $y$  are the coordinates of a given pixel,  $F(x, y)$  is its intensity and  $\bar{x}$  and  $\bar{y}$  are the average values along X and Y axis, respectively. It is referred to as central moment of order  $i+j$ .

$$\mu_{ij}^{XY} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j F(x, y) \quad (7)$$

$$f(x, y, z) = 1 \quad (14)$$

The Hu moments implementations for 2D grayscale or binary images make use of the intensity of the pixels as  $F(x, y)$ . Therefore, the function has values ranging between [0, 255] for grayscale images and between [0, 1] for binary images. When adapting this for 3D images, the above assignment loses sense, giving rise to new choices for that function.

We propose 2 possible implementations for  $F(x, y)$ . In Equation 8,  $F(x, y)$  is simply the distance from the coordinate  $z$  to the average value along that coordinate, while in Equation 9,  $F(x, y)$  is the Euclidean distance between the point of coordinates  $(x, y, z)$  and the nose tip  $(x_n, y_n, z_n)$ .

$$F(x, y) = (z - \bar{z}) \quad (8)$$

$$F(x, y) = \sqrt[2]{(x - x_n)^2 + (y - y_n)^2 + (z - z_n)^2} \quad (9)$$

From the above two definitions, we will consider  $Hu_{XY}$  and  $Hu_{XYd}$  as the Hu moments computed in  $XY$  plane that use the functions defined in Equations 8 and 9, 9, respectively. Similarly, we have for planes  $XZ$  e  $YZ$ , the following acronyms:  $Hu_{XZ}$ ,  $Hu_{XZd}$ ,  $Hu_{YZ}$  and  $Hu_{YZd}$ .

Hu moments undergo a normalization to be endowed with scale invariance, which is a very important property for 2D methods. In [17], the scale invariant moments of order  $i+j$  are presented as  $\psi_{ij}$  and calculated as Equation 10.

$$\psi_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}} = \frac{\mu_{ij}}{\sqrt[2]{\mu_{00}^{(2+i+j)}}}, (i+j) \geq 2 \quad (10)$$

Since we are dealing with 3D point clouds of faces, scale transformations are not feasible. Instead, we deal with another case: density variation. For that reason, the 7 Hu moments proposed in [17] and computed based on  $\psi_{ij}$  from Equation 10 are here adapted and computed based on  $\psi_{ij}$  from Equation 11.

$$\psi_{ij} = \mu_{ij}^{XY} \quad (11)$$

An example for the first Hu moment is expressed in Equation 12. All the others follow the same idea. The complete list of moments can be reviewed in [17].

$$\Psi_1 = \psi_{20} + \psi_{02} \quad (12)$$

The 3D Invariant moments [18], which were also employed as descriptors in our approach, are based on an expanded version of Equation 4 to three dimensions. Thus, they are invariant to translation and rotation in 3-D. Its continuous version can be reviewed in Equation 13.

$$M_{pqr} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q z^r f(x, y, z) dx dy dz \quad (13)$$

Similarly to Equations 8 and 9, we present two options for  $f(x, y, z)$  in Equations 14 and 15, where  $(x_n, y_n, z_n)$  are the nose tip coordinates.

$$f(x, y, z) = \sqrt[2]{(x - x_n)^2 + (y - y_n)^2 + (z - z_n)^2} \quad (15)$$

The seven Hu moments and the three 3DIT moments are extracted from each of the slices. The total number of descriptors depends on the value set in  $c$  for the intervals that cut the three orthogonal axes.

In summary, we adopted 6 forms to compute Hu moments: 3 forms with regard to the possible pairs of dependent axes  $XY$ ,  $XZ$  and  $YZ$  and 2 possible functions for each pair. As for the 3DIT moments, there are 2 forms of calculation: one for each function. Therefore, we end up with 8 distinct groups of moments, each group of Hu moments with 7 moments and each group of 3DIT moments with 3 moments. Altogether, the feature vector comprises 48 measures. This number is dependent on parameter  $c$ , because the thickness of a slice ultimately defines the amount of them. As we used  $c = (20, 20, 20)$ , 4 frontal, 7 sagittal and 7 transversal slices were assigned 48 descriptors each, amounting 864 descriptors per face point cloud.

### III. PRELIMINARY DISCUSSION

Several different databases have been used for the purpose of 3D facial recognition in the recent literature. We justify our choice of primarily working with Bosphorus 3D face database since it appears to be the one bringing the most challenging conditions. Additionally, we also evaluated the generalization of our method on FRGC v2.0 dataset.

Bosphorus Database was introduced by Savran et al 2008 [19] and is now widely used 3D face recognition. The database provides 2D and 3D samples along with landmark coordinates placed on regions of Action Units (AU) [20]. It also provides the correspondence between 3D data and 2D image pixels, making it possible to work either in 2D or 3D spaces.

The database comprises 4666 face scans not equally distributed over 105 individuals in various expressions, poses and occlusion conditions. The subjects are aged between 25 and 35, and most of them are Caucasian. Among them, 29 actors integrate the database, giving it additional complexity since they can exaggerate expressions and thus push face deformation to high levels (see Figure 2). Beard/mustache as well as short facial hair are present in some subjects.

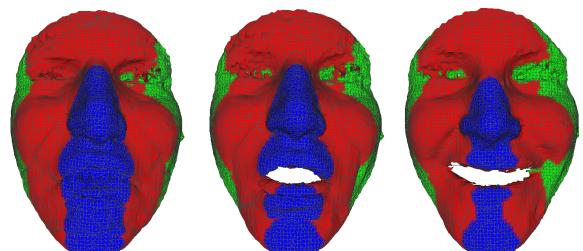


Fig. 2: Frontal slices from some emotional expressions samples: Anger, Fear, Disgust, Sad, Surprise and Happy.

The samples available in Bosphorus 3D face database can be divided in four wide categories: the ones performing expressions, poses, occlusions or with ignored label. Some subcategories are also derived from those and their descriptions are as follows: 34 Expressions (2902 samples); 13 Poses (1365 samples); 4 Occlusions (381 samples) and 1 Ignored (18 samples).

The FRGC v2.0 dataset, instead, comprises 466 subjects collected in 4,007 subject sessions [21] during the 2003-2004 academic year. It consists of 2,410 3D neutral samples and 1,597 3D non-neutral samples of different facial expressions (disgust, happy, sadness, surprise, other).

As stated by Soltanpour et. al. [4], it is difficult to make a fair comparison between methods in the literature for FR, mainly because different experiments are presented in various situations and for different conditions and databases (i.e., number of subjects, samples and samples per expression vary significantly). To circumvent this, researchers frequently adopt similar figures-of-merit. For instance, the analysis of Receiver Operation Characteristic (ROC) curves, which consist of face verification experiments created by plotting the true positive rate (TPR) or verification rate (VR) against the false positive rate (FPR) at various threshold settings is a common choice for performance evaluation of FR algorithms. We proceeded similarly and decided to perform 2 kinds of experiments for FR, using Bosphorus 3D face database and FRGC v2.0, named ROC I and rank-1 experiments.

In ROC I experiment, the gallery set is composed of all Neutral samples while the probe set is composed of Non-Neutral samples. For Bosphorus database it means there are two sets: (a) Neutral gallery set with 299 samples and (b) Non-Neutral probe set with the rest 2603 samples. For FRGC v2.0, this experiment setup is (a) Neutral gallery set with 466 samples (b) Neutral probe set with 1,944 samples and (c) Non-Neutral probe set with 1,597 samples.

In rank-1 experiment, in turn, the gallery set comprises one Neutral sample per subject while there are two probe sets, one composed of the other Neutral samples and the other composed of Non-Neutral samples. For Bosphorus database those 3 sets are grouped as follows: (a) Neutral gallery set with 105 Neutral samples, one per subject; (b) Neutral probe set with the rest 194 Neutral samples and (c) Non-Neutral probe set with the rest 2603 expression samples. For FRGC v2.0, the gallery set has 466 Neutral samples, while the probe set has 3,451 Non-Neutral samples.

ROC I is the most typical experiment setup and the primary purpose of this paper. It allows a larger number of samples per subject to be included in the training set i.e., it includes variability per class in the input set. In Bosphorus 3D face database there is a small number of neutral samples. They represent around 6.4% of the whole database and 10% of all the Expression samples used in our work. Moreover, it has an unbalanced distribution of up to 4 samples per subject and the average is 2.8 samples per subject.

Rank-1 is a more challenging experiment setup, since it evaluates the classifier performance when only one sample per class is presented. With that restriction, the classifier has its ability to recognize individuals measured even when

they perform expressions that cause significant deformations on face. As mentioned before in this paper, Bosphorus 3D face database included a relevant number of professional actors/actresses as subjects in order to guarantee high expressiveness.

Having defined the training and testing setup, the most suitable classifier should be selected. We selected two classifiers to perform our experiments: Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), since both are frequently referred to in the literature as adequate to scenarios with few samples per class, which is specifically our case.

Linear Discriminant Analysis (LDA) has been commonly used for many applications such as FR as a dimensionality reduction technique [22] and as a classifier [23]. LDA is used in the present work as a classifier itself, and not for dimensionality reduction purposes prior to classification. LDA searches for basis vectors of a subspace onto which the data would be projected and that maximize the distinction between classes [24] and reduces the number of features down to the number of classes minus one. Various approaches have been proposed to solve this singularity problem, including applying Singular Value Decomposition (SVD) or eigenvalue decomposition to the data matrix and using iterative algorithms, such as LSQR to solve LDA as a least squares problem. The last one was the approach we have adopted in this work. But there is a constraint [24] while comparing PCA and LDA performances; Martinez [24] states that, in order the classifier not to become singular, the number of samples must be superior than the sum of number of classes and number of features. In summary, the number of samples must be, at least double the number of classes minus one. This is the reason why in this work LDA is used in ROC I , but not in rank-1 experiments.

Support Vector Machine (SVM) is originally a binary classifier that constructs an optimal linear decision surface based on a weighted combination of elements of the training set, the so called support-vectors [25], [26]. Training is formally defined by Cortes and Vapnik [25] as minimizing an error function which depends on two parameters ( $C, \gamma$ ), where  $C$  is a constant and controls the trade-off between complexity of decision surface and frequency of misclassification in the training set [25]. Low values of  $C$  make the decision surface smooth and simple, while high values allow that more samples to be chosen as support vectors, creating a decision surface that tends to exactly separate classes in the training set. SVM can also be extended to construct a nonlinear decision surface by using a kernel function that projects the original space onto a high dimensional feature space. The kernel function used in this paper is the Radial Basis Function (RBF) kernel, where  $\gamma$  is a constant and in practice, adjusts how the decision surface is affected by each sample. Low values of  $\gamma$  mean that each sample has a far influence and therefore even far samples may affect the definition of the decision surface. Conversely, high values of  $\gamma$  mean that each sample has a close influence, which may force the decision surface to adapt to the closest samples. Parameters  $C$  and  $\gamma$  mainly influence the classifier's ability of generalization and therefore their appropriate tuning is fundamental. Unlike LDA classifier, no restrictions are imposed by the SVM in ROC I and rank-1

TABLE I: RR for Neutral vs NonNeutral rank-1 experiment with ST slices combination, using SVM-SVC-rbf.

Moments	4HuXY	4HuXZ	4HuYZ	3DIT
RR (%)	72.30	41.53	48.56	58.47
Moments (distance)	4HuXYd	4HuXZd	4HuYZd	3DITd
RR (%)	93.32	84.90	90.13	90.13

experiments. For what concerns the decision function shape, the implementation is twofold: *one vs all* (OvA) and *one vs one* (OvO); we use OvA in our experiments.

#### IV. RESULTS AND DISCUSSION

The proposed method was developed using C/C++ OpenCV and PCL libraries on Ubuntu/Linux platform, and all the experiments are implemented on a PC with the CPU by Intel i7-5500U, 2.4 GHz, 8GB RAM, using just one of the four cores. We found that the pose correction is the most time-consuming part, taking nearly up to 3 seconds (out of 4 seconds in total).

The computation of the HU moments and the 3DIT moments had time demand considerably decreased after the density normalization procedures, with PCL voxel grid filter of 2mm leaf size for all the 3 dimensions.

To perform the tri-axial slicing, we tested the values [5, 10, 15, 20, 25] in millimeters for the thickness  $c = (c_f, c_s, c_t)$ . The best results were achieved with the triple  $c = (20, 20, 20)$ . Lower values revealed to be very sensitive to rotation while higher values led to small quantity of slices and therefore insufficient amount of descriptors.

Next, we evaluated the descriptive capacity of the Hu moments and concluded that only the first 4 are essential for classification, for any function  $f(x, y)$  tested. With that restriction, the maximum number of attributes was reduced down to 612. This improved the computation of descriptors, since the 3 last moments are more time-consuming than the first ones. Thus, we define 4HuXYd as the set of the first 4 Hu moments, taking in XY plane and function  $f(x, y)$  computed as in Equation 9. Similar terminology was adopted for the other moments.

##### A. Best combination of slices and moments

It is shown in Table I that features calculated according to the distance function of Equation 9 are individually better than those calculated from Equation 8. Moreover, the features HuXY are globally better and, regardless the function adopted in our experiments, the group of moments HuXZ performed worse, which suggests that the coordinate  $y$  as one of the dependent variables which considerably contributes to better results. Next, we also tested combinations of all the features and found that the best arrangement is 4HuXYd, 4HuXZd and 4HuYZd.

Once the best combination of moments has been defined, the next step would be to determine the combination of slices that yields the best recognition rate. We evaluated Frontal (F), Sagittal (S) and Transversal (T) slices individually and combined, forming the test set  $[F, S, T, FS, FT, ST, FST]$ .

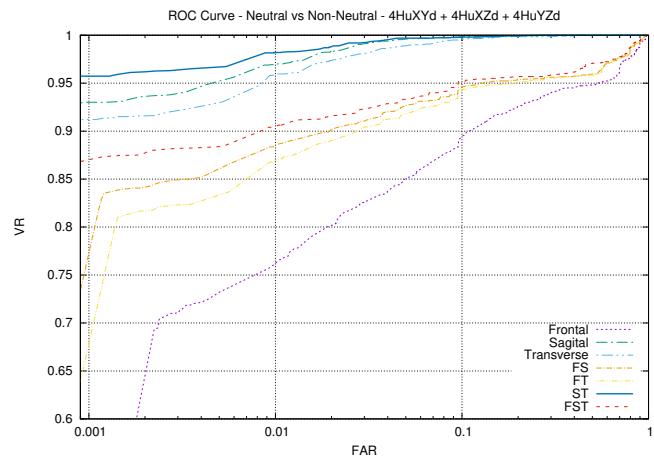


Fig. 3: VR for Neutral vs Non-Neutral rank-1 experiment with different slice combinations, using SVM-SVC-rbf.

Figure 3 presents the performances for 4HuXYd + 4HuXZd + 4HuYZd attributes for all the possible combinations of Frontal (F), Sagittal (S) and Transverse (T) slicings. Frontal slicing alone is the worst scenario and, when combined to others, it degrades the VR. The best scenario is the combination Sagittal-Transverse (ST). This trend repeats in the various cases investigated.

The behavior observed in Figure 3 was also noticed for ROC I experiment with SVM-SVC-rbf classifier. For ROC I, we also employed LDA as classifier and the best combination of descriptors was 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd and 4HuYZd. The SVM-SVC-rbf classifier performed worse than LDA.

##### B. ROC I and rank-1 Experiments

In this subsection, we discuss the best experimental results for face recognition achieved in this work and compare them to the literature. Initially, results for ROC I are brought. To make fair the comparison the verification rate (VR) is analyzed. We then present results for rank 1 experiment, since it appears very often in recent works. In this case, recognition rate (RR) is used as performance indicator.

For ROC I, we utilized LDA classifiers. Our results are listed in Table II and refer to the ST slicing with 3DIT, 3DITF, 4HuXY, 4HuYZ, 4HuXYd and 4HuYZd descriptors forming a 308-long feature vector. It is important to mention that, even though LDA for ROC I experiment has a larger number of input descriptors, internally it selects a number of attributes equal to the number of classes minus 1, as explained in [24].

For what concerns Bosphorus database, Table II shows that the verification rate achieved in this work is superior to the ones reported by Elaiwat et al [6], Liu et al [5], Ocegueda et al [10] and Ming et al [27]. Ming et al [27] also neglected samples with occlusion and pose variation, but included IGN samples (only 18 samples). Elaiwat [6] and Liu [5], instead, included samples with occlusion and pose variation in the experiments; on doing this, symmetric filling [6][5] or hole filling [5] techniques were used. Liu [5] only neglected 90°

TABLE II: VR values at 0.1% FAR. Bosphorus ROC I experiment with VR and RR for Neutral vs NonNeutral, and FRGC v2.0 ROC I experiment with VR for Neutral vs NonNeutral and ROC III experiment with VR.

Databases	Bosphorus		FRGC v2.0	
Experiments	ROC I		ROC I	ROC III
References	RR (%)	VR (%)	VR (%)	VR (%)
This work (LDA*)	98.7	98.9	97.1	94.7
This work (SVM**)	85.3	87.0	84.2	87.9
Elaiwat et al [6]	n/a	91.0	98.0	97.8
Ocegueda et al [10]	n/a	93.8	98.1	97.9
Ming et al [27]	92.0	94.0	n/a	n/a
Liu et al [5]	n/a	95.6	93.1	n/a
Emambakhsh [11]	n/a	n/a	n/a	93.5
Al-Osaimi, Faisal R.[12]	n/a	n/a	97.8	94.1
Berretti et al [7]	n/a	n/a	91.4	86.6

rotated samples. Although the testing sets employed in Elaiwat [6] e Liu [5] are very comprehensive, both works need to find the facial symmetry plane to achieve face completion, which ultimately adds complexity to their solution. Moreover, [6] uses ICP for fine alignment. In these experiments, the feature vector size is 252 units-long for LDA and SVM-SVC-rbf classifiers; LDA was solved by SVD method using the pack *sklearn* in python language, and SVM parameters were set  $C = 8$  and  $\gamma = 0.125$ .

To assess the generalization of the proposed method, scans from the FRGC v2.0 dataset were used in the pretty same conditions of the experiments reported so far for the Bosphorus database; results are reported in Table II. Once again, the proposed method performs well, with rates comparable to important recent references and reaching the top 3 among those. Here we report ROC I experiment (as described earlier) and also a ROC III face verification experiment, in which the training includes the scans of the Fall (Autumn) 2003 sessions, whereas the probe scans contains shots from the Spring 2004 sessions. Due to this temporal lapse between the acquisition of probe and training scans, this experiment is regarded as the most difficult one of the FRGC protocol as pointed out by Berretti et al[7].

As mentioned earlier, for rank 1 we only use SVM-SVC-rbf classifier where parameters were set  $C = 8$  and  $\gamma = 0.125$ . This experiment is especially important to place the proposed method among those using training/testing strategy. A major advantage of this comparison is that it can be detailed, since the literature provides recognition rates for the subset of non-neutral samples. It allows to evaluate more specifically each method in regard to emotional expressions, neutral, LFAU, UFAU and CAU samples.

The results shown in Table III take into account only the ST slicings with moments 4HuXYd, 4HuXZd and 4HuYZd, for a total of 168 descriptors. In some cases, our method reaches or surpasses the best results reported by Berretti et al. [7], Li et al. [1], Ming et al [27], Emambakhsh and Evans [11] and Al-Osaimi and Faisal [12]. Those cases are highlighted in bold. Also for the set of expressions Anger and Disgust our method performed better than Li et al results. We also beat Berretti et al. [7] for Neutral, Fear, Sadness, Surprise and CAU samples.

TABLE III: RR for different probe categories on Bosphorus rank1 experiment. NOTE: Ming et al [27], Emambakhsh and Evans [11] and Al-Osaimi and Faisal [12]

Descriptors(#)	-	-	130	-	-	168
Key Points(#)	648	145	2	-	-	1
Probes(#)	Li et al [1]	Berretti et al [7]	[27]	[11]	[12]	This work (SVM)
<b>Bosphorus</b> (rank1 RR(%))						
Neutral (194)	100	98.5	94.3	99.0	98.5	<b>99.0</b>
NonNeutral (2603)	98.7	96.6	92.0	n/a	92.4	95.4
Emotions (453)	96.7	91.2	n/a	96.2	n/a	<b>95.2</b>
Anger (71)	97.2	88.7	n/a	94.1	n/a	<b>98.6</b>
Disgust (69)	86.9	81.2	n/a	88.2	n/a	<b>98.6</b>
Fear (70)	98.6	91.4	n/a	98.6	n/a	<b>95.7</b>
Happy (106)	98.1	94.3	n/a	98.1	n/a	87.7
Sad (66)	100	95.5	n/a	96.9	n/a	<b>98.5</b>
Surprise (71)	98.6	94.4	n/a	100	n/a	95.8
LFAU (1549)	98.8	97.5	n/a	n/a	n/a	94.7
UFAU (432)	100	99.1	n/a	n/a	n/a	98.4
CAU (169)	100	96.4	n/a	n/a	n/a	<b>98.8</b>

In addition to the results, some conceptual and practical aspects should be considered at this point. Unlike Berretti's method, which is not targeted for Kinect or low-resolution low-cost cameras, our method does well with this acquisition system. Indeed, in previous investigation, our method did well when using the 3DMAD [28] database of samples based on Kinect; using Multi Layer Perceptron Neural Network (MLP-NN) as classifier, we found recognition rates equivalent to those obtained for Bosphorus in neutral vs neutral experiment, which is consistent to the fact that 3DMAD only contains neutral expressions. This suggests how robust the method is to different data acquisition systems. Furthermore, as mentioned earlier in this paper, the present method has a cloud density regularization step based on voxel grid; this is especially important to overcome differences in the scanner resolution.

Finally, Li et al [1] extended the SIFT-like matching framework to mesh data and proposed a novel approach using fine-grained matching of 3D keypoint descriptors, called Histogram of Multiple surface differential Quantities (HOMQ) by combining HOG, HOS, and HOGs at feature level. Although Li reported the best performances for non-neutral set, the high quantity of required keypoints resulted in significant computational efforts; his method takes on average 1.25 minutes on a PC with Intel 950, 3.07 GHz processor, 8GB RAM. The method here introduced, running under similar conditions (Intel i7-5500U processor, 2.4 GHz) and employing OpenCv and PCL libraries for C/C++ and takes less than 4 seconds to perform the entire recognition process.

## V. CONCLUSION AND FUTURE WORKS

In this work a technique for 3D face recognition relying on adapted Hu and 3D invariant moments of tri-axial point cloud slices was introduced. The combination of 4HuXY, 4HuXY, 4HuYZd, 4HuXYd, 3DIT and 3DITd attributes reached verification rate of 98.7% for 0.1% FAR in ROC I experiment

using LDA classifier. For rank-1, using 4HuXYd, 4HuXZd and 4HuYZd, experiment the identification rates of 99% and 95.4% for the Neutral vs Neutral and Neutral vs Non-Neutral experiments, respectively, which reinforces that the use of function distance improves the characterization of individuals. Identification rates above 97% for the Neutral vs Non-Neutral in the subsets CAU, UFAU, Anger, Disgust and Sad **were achieved**. The technique only needs to identify a single key point to cut the face into 14 different cloud points in Sagittal and Transversal planes, which makes it computationally less complex than other approaches of the same nature. The Sagittal and Transverse slices showed to be less sensitive to surface variations even for certain types of expressions, while the Frontal slices, in the vast majority of cases, degraded the recognition rate when associated with the other two orthogonal planes.

The method introduced here can be additionally improved and made more general if the following limitations are addressed: a) occlusion and self-occlusion: although were neglected in the current version, symmetry of human face can be exploited for dealing with this issue; b) the dependence on initial alignment: the method can be made faster as new approaches for the face orientation correction are made available. We emphasize that the essence of our method relies on the proposal of a new feature descriptor; and, finally, c) the dependence on cloud density: Hu moments are sensitive to density variation of point clouds. We addressed this issue in the current version by applying the voxel-grid algorithm in order to equalize the density for all samples, however other strategies may be tried.

## REFERENCES

- [1] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "Towards 3d face recognition in the real: A registration-free approach using fine-grained matching of 3d keypoint descriptors," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128–142, Jun 2015.
- [2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.
- [3] H. Hatem, Z. Beiji, and R. Majeed, "A survey of feature base methods for human face detection," *International Journal of Control and Automation*, vol. 8, no. 5, pp. 61–78, 2015.
- [4] S. Soltanpour, B. Boufama, and Q. Jonathan Wu, "A survey of local feature methods for 3d face recognition," *Pattern Recognit.*, vol. 72, no. C, pp. 391–406, Dec. 2017.
- [5] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-d face recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 914–925, March 2013.
- [6] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, "A Curvelet-based approach for textured 3D face recognition," *Pattern Recognition*, vol. 48, no. 4, pp. 1235 – 1246, 2015.
- [7] S. Berretti, N. Werghi, A. del Bimbo, and P. Pala, "Selecting stable keypoints and local descriptors for person identification using 3d face scans," *The Visual Computer*, vol. 30, no. 11, pp. 1275–1292, Nov 2014.
- [8] J. Gao and A. N. Evans, "Expression robust 3D face landmarking using thresholded surface normals," *Pattern Recognition*, vol. 78, pp. 120–132, jun 2018.
- [9] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013.
- [10] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "3d face discriminant analysis using gauss-markov posterior marginals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 728–739, March 2013.
- [11] M. Emambakhsh and A. Evans, "Nasal Patches and Curves for Expression-Robust 3D Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 995–1007, may 2017.
- [12] F. R. Al-Osaimi, "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*, vol. 25, no. 2, pp. 658–672, feb 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7300438/>
- [13] E. B. Corrochano, *Handbook of Geometric Computing*. Springer, 2005.
- [14] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1–4.
- [15] D. Borrman, J. Elseberg, K. Lingemann, and A. Nüchter, "The 3d hough transform for plane detection in point clouds: A review and a new accumulator design," *3D Research*, vol. 2, no. 2, p. 3, Nov 2011.
- [16] M. Beka, G. Oka, M. Vajteric, and L. Grigori, "On iterative qr pre-processing in the parallel block-jacobi svd algorithm," *Parallel Computing*, vol. 36, no. 5, pp. 297 – 307, 2010, parallel Matrix Algorithms and Applications.
- [17] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962.
- [18] T. Suk, J. Flusser, and J. Boldy, "3d rotation invariants by complex moments," *Pattern Recognition*, vol. 48, no. 11, pp. 3516 – 3526, 2015.
- [19] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökerk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *Biometrics and Identity Management*, B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 47–56.
- [20] P. Ekman and W. V. Friesen, "Measuring facial movement," *Environmental psychology and nonverbal behavior*, vol. 1, no. 1, pp. 56–75, 1976.
- [21] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. I, pp. 947–954, 2005.
- [22] H.-M. Moon, D. Choi, P. Kim, and S. B. Pan, "Lda-based face recognition using multiple distance training face images with low user cooperation," in *Consumer Electronics (ICCE), 2015 IEEE International Conference on*. IEEE, 2015, pp. 7–8.
- [23] A. Tsalavra, K. Dalakleidi, E. I. Zacharaki, N. Tsiparas, F. Constantinidis, N. Paragios, and K. S. Nikita, "Comparison of multi-resolution analysis patterns for texture classification of breast tumors based on dee-mri," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2016, pp. 296–304.
- [24] A. M. Martinez and A. C. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, Feb 2001.
- [25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [26] M. H. Mousavi, K. Faiez, and A. Asghari, "Three dimensional face recognition using svm classifier," in *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on*. IEEE, 2008, pp. 208–213.
- [27] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3d face recognition," *Neurocomputing*, vol. 129, pp. 445 – 457, 2014.
- [28] N. Erdogmus and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," 2013.