

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Uma Abordagem para a Deformação de Faces  
em Vídeos visando à Caricaturização

Vinícius Fernandes de Sousa

Campina Grande, Paraíba, Brasil

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Uma Abordagem para a Deformação de Faces  
em Vídeos visando à Caricaturização

Vinícius Fernandes de Sousa

Dissertação submetida à Coordenação do Curso de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Campina Grande -  
Campus I como parte dos requisitos necessários para obtenção do grau de  
Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação  
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes  
José Eustáquio Rangel de Queiroz  
(Orientadores)

Campina Grande, Paraíba, Brasil

©Vinícius Fernandes de Sousa, Novembro de 2021

S725a      Sousa, Vinícius Fernandes de.  
              Uma abordagem para a deformação de faces em vídeos visando  
              à caricaturização / Vinícius Fernandes de Sousa. – Campina  
              Grande, 2021.  
              120 f. : il. color.

              Dissertação (Mestrado em Ciência da Computação) –  
              Universidade Federal de Campina Grande, Centro de Engenharia  
              Elétrica e Informática, 2021.  
              "Orientação: Prof. Dr. Herman Martins Gomes, Prof. Dr.  
              Eustáquio Rangel de Queiroz."  
              Referências.

              1. Processamento de Imagens. 2. Vídeos Digitais. 3.  
              Deformações. 4. Faces Humanas. 5. Caricatura. I. Gomes, Herman  
              Martins. II. Queiroz, Eustáquio Rangel de. III. Título.

CDU 004.932(043)

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECÁRIA MARIA ANTONIA DE SOUSA CRB 15/398



MINISTÉRIO DA EDUCAÇÃO  
**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**  
POS-GRADUACAO CIENCIAS DA COMPUTACAO  
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

## FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

**VINÍCIUS FERNANDES DE SOUSA**

UMA ABORDAGEM PARA A DEFORMAÇÃO DE FACES EM VÍDEOS VISANDO À CARICATURIZAÇÃO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Mestre em Ciência da Computação

Aprovada em: 21/10/2021

Prof. Dr. HERMAN MARTINS GOMES, UFCG, Orientador

Prof. Dr. JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, Orientador

Profa. Dra. JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, UFCG, Examinadora Interna

Prof. Dr. EDMAR CANDEIA GURJÃO, UFCG, Examinador Externo



Documento assinado eletronicamente por **HERMAN MARTINS GOMES, PROFESSOR 3 GRAU**, em 21/10/2021, às 17:16, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ, Usuário Externo**, em 21/10/2021, às 17:53, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).

Documento assinado eletronicamente por **EDMAR CANDEIA GURJAO, PROFESSOR 3 GRAU**, em



22/10/2021, às 07:50, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **JOSEANA MACEDO FECHINE, PROFESSOR DO MAGISTERIO SUPERIOR**, em 22/10/2021, às 08:01, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **1821101** e o código CRC **C6094F55**.

## Resumo

Caricatura facial pode ser definida como a representação de um indivíduo por meio do exagero das características mais distintas da face e a simplificação das características mais comuns. As caricaturas geradas por computadores, por sua vez, são desenvolvidas com a intenção de produzir caricaturas de maneira automática ou semiautomática. Isto deve-se, em grande parte, ao crescente avanço nas áreas de Computação Gráfica e Visão Computacional. Neste contexto, o presente estudo tem como principal objetivo a proposição de uma abordagem para a geração automática de distorções faciais em vídeos digitais que lembrem caricaturas faciais. O estudo teve como principais desafios a concepção de um modelo para a detecção dos pontos da face em vídeos, a obtenção de regras gerais que satisfaçam visualmente à geração automática de caricaturas, como também, o tratamento dos possíveis problemas provenientes da geração dos vídeos distorcidos. Neste estudo, foi concebida uma abordagem adaptativa para a da detecção dos pontos fiduciais em vídeos, foi elaborada uma estratégia para que a coerência temporal das deformações se mantivesse. Após a comparação da abordagem adaptativa, desenvolvida neste estudo, com outras abordagens, foi possível observar uma menor taxa de erro na localização dos pontos fiduciais da face. Também foi possível determinar uma configuração que melhor atendeu aos critérios elencados neste estudo, a partir de uma análise dos resultados obtidos por meio de um questionário de sondagem da opinião de um conjunto de voluntários.

**Palavras-chave:** Vídeos digitais; Deformações; Faces Humanas; Caricatura; Processamento de Imagens.

## **Abstract**

Facial caricature can be defined as the representation of an individual by exaggerating the most distinct features of the face and simplifying the most common features. Computer-generated cartoons, in turn, are developed with the intention of producing automatic or semi-automatic cartoons. This is due, in large part, to the growing advancement in the areas of Computer Graphics and Computer Vision. Within this context, the present study has as its main goal the proposition of an approach for the automatic generation of facial distortions in digital videos that resemble facial caricatures. The main challenges of the study were the design of a model for the detection of facial points in videos, obtaining general rules that visually satisfy the automatic generation of caricatures, as well as the treatment of possible problems related to the generation of the distorted videos. In this study, an adaptive approach was designed to detect fiducial points in videos, and a strategy was developed so that the temporal coherence of deformations was maintained. After comparing the adaptive approach developed in this study with other approaches, it was possible to observe a lower rate of error in locating the fiducial points of the face. It was also possible to determine a configuration that best met the criteria listed in this study, based on an analysis of the results obtained through a questionnaire to survey the opinions of a set of volunteers.

**Key words:** Digital Videos; Deformations; Human Faces; Caricature; Image Processing.

## **Agradecimentos**

Primeiramente, gostaria de agradecer à minha esposa Michele e minha filha Alice, pelo apoio, dedicação e carinho recebidos não apenas durante o período de mestrado, mas ao longo de nossas vidas. Agradeço também aos meus pais, José Carlos e Natanilda, por sempre acreditarem em mim e proverem todo suporte necessário durante toda minha vida de estudante para que eu pudesse chegar até aqui. Gostaria de agradecer enormemente aos meus orientadores, Herman Martins Gomes e José Eustáquio Rangel, pela paciência, suporte, incentivo e orientações que me ajudaram a realizar esse trabalho. Aos colegas e professores da UFCG, e a todos os funcionários que influenciaram, de forma direta ou não, a conclusão dessa fase em minha vida. A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização da Pesquisa . . . . .	1
1.2	Motivações . . . . .	2
1.3	Objetivos e Relevância . . . . .	4
1.4	Questões de Pesquisa . . . . .	5
1.5	Estrutura da Dissertação . . . . .	5
<b>2</b>	<b>Fundamentação</b>	<b>6</b>
2.1	Detecção de Pontos Fiduciais em Faces . . . . .	6
2.2	Definição das Características a serem Deformadas . . . . .	10
2.3	Transformação da Imagem . . . . .	12
2.4	Caricatura . . . . .	15
2.5	Considerações Finais . . . . .	17
<b>3</b>	<b>Pesquisas Relacionadas</b>	<b>18</b>
3.1	Breve Histórico da Geração Automática de Caricaturas . . . . .	18
3.2	Detecção da Face . . . . .	22
3.3	Detecção de Pontos Fiduciais . . . . .	24
3.4	Regras de Definição das Características Relevantes . . . . .	26
3.5	Considerações Finais . . . . .	28
<b>4</b>	<b>Abordagem Proposta</b>	<b>30</b>
4.1	Visão Geral do Modelo Proposto . . . . .	30
4.2	Detecção dos Pontos Fiduciais no Vídeo . . . . .	31
4.3	Regras de Definição das Características Relevantes . . . . .	34

4.4	Métodos de Deformação da Face Contida nos Vídeos . . . . .	36
4.5	Detalhes de Projeto e Implementação . . . . .	38
4.6	Metodologia Adotada para a Avaliação com Usuários . . . . .	40
4.7	Considerações Finais . . . . .	44
<b>5</b>	<b>Resultados e Discussões</b>	<b>45</b>
5.1	Avaliação dos Algoritmos de Detecção de Pontos Fiduciais . . . . .	45
5.1.1	Banco de Imagens . . . . .	46
5.1.2	Definição da Métrica . . . . .	47
5.1.3	Validação Experimental . . . . .	48
5.1.4	Resultados Obtidos . . . . .	48
5.2	Avaliação do Algoritmo Adaptativo . . . . .	53
5.2.1	Base de Dados . . . . .	54
5.2.2	Métrica de Erro . . . . .	54
5.2.3	Abordagens Avaliadas . . . . .	55
5.2.4	Apresentação e Discussão dos Resultados Experimentais . . . . .	57
5.3	Avaliação da Abordagem Proposta Com Usuários (QDPU) . . . . .	66
5.3.1	Resultados da Análise do Questionário de Delineamento de Perfil do Usuário . . . . .	68
5.3.2	Resultados da Análise do Questionário para a Sondagem da Opinião do Usuário (QSOU) . . . . .	72
5.4	Considerações Finais . . . . .	79
<b>6</b>	<b>Considerações Finais</b>	<b>81</b>
6.1	Síntese da Dissertação . . . . .	81
6.2	Contribuições do Estudo Realizado . . . . .	84
6.3	Estudos Futuros . . . . .	84
<b>A</b>	<b>Questionário de Delineamento do Perfil do Usuário</b>	<b>93</b>
<b>B</b>	<b>Questionário de Sondagem da Opinião do Usuário sobre os resultados de uma Abordagem para Geração Automática de Caricaturas</b>	<b>96</b>

---

<b>C E-mail de Recrutamento dos Usuários</b>	<b>111</b>
--	------------

<b>D Exemplos dos Quadros Deformados Utilizados no Experimento com a Avaliação com Usuários</b>	<b>112</b>
---	------------

# Lista de Figuras

2.1	Exemplos das variações existentes no processo de detecção dos pontos fiduciais faciais . . . . .	7
2.2	Pontos fiduciais da face, extraídos a partir da técnica AAM. . . . .	8
2.3	Busca dos pontos fiduciais da face utilizando ASM . . . . .	9
2.4	Exemplo de método de deformação interativo . . . . .	10
2.5	Distâncias proporcionais utilizadas para a deformação facial . . . . .	11
2.6	Exemplos do uso das malhas triangulares em um conjunto de pontos e suas diferentes possibilidades de triangulações. . . . .	14
2.7	Exemplos de caricaturas. . . . .	16
3.1	(A) Face original, (B) 50% de exagero, (C) 100% de exagero, (D) 140% de exagero, (E) 160% de exagero . . . . .	19
3.2	(A) Imagens originais, (B) Imagens caricaturadas pelo sistema proposto . .	21
3.3	(A) Imagens originais, ((B)(C)) Imagens caricaturadas pelo sistema proposto	23
3.4	Faces de entrada e suas respectivas caricaturas produzidas pelo WarpGAN .	27
3.5	<i>Frames</i> de vídeos gerados a partir de alterações das expressões faciais . . .	27
4.1	Processo geral para gerar uma caricatura facial. . . . .	31
4.2	Fluxograma do processo de captura dos pontos fiduciais. . . . .	32
4.3	Distribuição dos pontos fiduciais ao redor da face. . . . .	32
4.4	Proporções Faciais. . . . .	35
4.5	Diferentes Níveis de Distorções. . . . .	36
4.6	Exemplo do uso da rotação dos pontos fiduciais para aprimorar a transformação espacial dos pontos . . . . .	38

---

4.7	Exemplo do uso de pontos âncoras para realizar a deformação espacial dos <i>pixels</i> . . . . .	39
4.8	Quadros ilustrativos dos vídeos utilizados no experimento . . . . .	41
4.9	Exemplo de parte do questionário de delineamento de perfil do usuário . . . . .	43
4.10	Exemplos de quadros com os diferentes níveis de deformação utilizados. (a) Nível de deformação 1; (b) Nível de deformação 2 e (c) Nível de deformação 3	43
5.1	Boxplot dos erros na fase de validação . . . . .	51
5.2	Modelo do PDM de 21 pontos fiduciais detectados de uma face frontal . . . . .	56
5.3	Boxplots do RMSE das abordagens avaliadas no Experimento 1 . . . . .	58
5.4	Histogramas de valores RMSE para o Experimento 1. (a) MED, (b) MFO, (c) PFO, (d) PDM, (e) AAD . . . . .	59
5.5	Resultados visuais das abordagens analisadas no Experimento 1, em uma sequência de quadros de um vídeo da base de dados 300-vw. (a) MED, (b) MFO, (c) PFO, (d) PDM, (e) AAD . . . . .	61
5.6	Boxplots de estatísticas RMSE para o Experimento 2 . . . . .	63
5.7	Histogramas de valores RMSE para o Experimento 2. (a) CLM, (b) DLB, (c) FAL, (d) OPF e (e) AAD . . . . .	64
5.8	Resultados visuais das abordagens analisadas no Experimento 2, em uma sequência de quadros de um vídeo da base de dados 300-vw. (a) AAD, (b) CLM, (c) DLB, (d) OPF, (e) FAL . . . . .	67
5.9	Distribuição da Escolaridade dos Participantes . . . . .	69
5.10	Distribuição do Gênero dos Participantes . . . . .	70
5.11	Distribuição do Uso de Corretivos Oculares dos Participantes . . . . .	70
5.12	Distribuição do Costume dos Usuários de Observar Caricaturas nos Meios de Comunicação . . . . .	71
5.13	Distribuição da Frequência de Observações de Caricaturas nos Meios de Comunicação dos Participantes no Dia a Dia. Seguindo as seguintes âncoras semânticas: 1 - Muito Baixa; 2 - Baixa; 3 - Nem Baixa nem Alta; 4 - Alta; 5 - Muito Alta . . . . .	71

5.14	Distribuição da Facilidade de Identificação de uma Caricatura. Seguindo as seguintes âncoras semânticas: 1 - Muito Baixa; 2 - Baixa; 3 - Nem Baixa nem Alta; 4 - Alta; 5 - Muito Alta . . . . .	72
5.15	<i>Boxplots</i> das avaliações dos usuários para cada questão. (a) Questão 1; (b) Questão 2 e (c) Questão 3 . . . . .	73
5.16	Histogramas das avaliações dos usuários para cada processamento da Questão 1. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4 . . . . .	76
5.17	Histogramas das avaliações dos usuários para cada processamento da Questão 2. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4 . . . . .	77
5.18	Histogramas das avaliações dos usuários para cada processamento da Questão 3. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4 . . . . .	78
D.1	Exemplos de Quadros Utilizando o Processamento 1 - Processamento sem a utilização do algoritmo de adaptativo de detecção dos pontos fiduciais e nível 3 de deformação da imagem. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. . . . .	113
D.1	Exemplos de Quadros Utilizando o Processamento 1 - Processamento sem a utilização do algoritmo de adaptativo de detecção dos pontos fiduciais e nível 3 de deformação da imagem. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. . . . .	114
D.2	Exemplos de Quadros Utilizando o Processamento 2 - Processamento com a utilização do algoritmo adaptativo e nível 3 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. . . . .	115
D.2	Exemplos de Quadros Utilizando o Processamento 2 - Processamento com a utilização do algoritmo adaptativo e nível 3 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. . . . .	116

- 
- D.3 Exemplos de Quadros Utilizando o Processamento 3 - Processamento com a utilização do algoritmo adaptativo e nível 2 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. 117
- D.3 Exemplos de Quadros Utilizando o Processamento 3 - Processamento com a utilização do algoritmo adaptativo e nível 2 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. 118
- D.4 Exemplos de Quadros Utilizando o Processamento 4 - Processamento com a utilização do algoritmo adaptativo e nível 1 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. 119
- D.4 Exemplos de Quadros Utilizando o Processamento 4 - Processamento com a utilização do algoritmo adaptativo e nível 1 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento. 120

# Lista de Tabelas

3.1	Sumário das principais abordagens para geração automática de caricaturas .	29
4.1	Afirmativas apresentadas ao usuário parar a avaliação dos vídeos gerados .	42
5.1	Erro médio na fase de de treinamento para os 5 modelos analisados de cada algoritmo . . . . .	49
5.2	Média, desvio padrão e coeficiente de variação na fase de treinamento . . . .	50
5.3	Erro associado à detecção dos pontos fiduciais na fase de validação . . . . .	50
5.4	Comparação entre as técnicas utilizando o tese de Tukey . . . . .	52
5.5	Grupos de similaridade obtidos por meio do teste de Tukey-Kramer . . . . .	53
5.6	Lista de vídeos da base 300-vw selecionados para o processo de avaliação .	55
5.7	Estatísticas RMSE do Experimento 1 . . . . .	57
5.8	Teste de normalidade Shapiro-Wilk para distribuições RMSE de abordagens avaliadas no Experimento1 . . . . .	58
5.9	Teste de Kruskal-Wallis para avaliar diferenças estatísticas entre as aborda- .	59
	gens no Experimento 1	
5.10	Resultados do teste <i>post-hoc</i> de Dunn para identificar pares de métodos no .	60
	Experimento 1 que têm diferenças estatisticamente significativas.	
5.11	Estatísticas RMSE obtidas no Experimento 2 . . . . .	62
5.12	Resultados do teste de Shapiro Wilk para distribuições RMSE de aborda- .	63
	gens avaliadas no Experimento 2	
5.13	Resultado do teste de Kruskal para experimento 2 . . . . .	65
5.14	Teste <i>post-hoc</i> de Dunn para comparação entre as abordagens do Experi- .	65
	mento 2.	
5.15	Valores médios obtidos a partir da avaliação dos usuários . . . . .	73
5.16	Resultados do teste de Shapiro Wilk para a Questão 1 . . . . .	74
5.17	Resultados do teste de Shapiro Wilk para a Questão 2 . . . . .	74

---

5.18	Resultados do teste de Shapiro Wilk para a Questão 3 . . . . .	75
5.19	Teste de Kruskal-Wallis para avaliar as diferenças estatísticas entre os processamentos abordados . . . . .	75
5.20	Resultados do teste <i>post-hoc</i> de Dunn para identificar pares de processamentos na Questão 1 que têm diferenças estatisticamente significativas . . . . .	79
5.21	Resultados do teste <i>post-hoc</i> de Dunn para identificar pares de processamentos na Questão 2 que têm diferenças estatisticamente significativas . . . . .	79
5.22	Resultados do teste <i>post-hoc</i> de Dunn para identificar pares de processamentos na Questão 3 que têm diferenças estatisticamente significativas . . . . .	80

# Capítulo 1

## Introdução

Neste capítulo, é apresentada uma introdução ao problema de pesquisa abordado neste estudo. Inicialmente, na Seção 1.1, é apresentada uma contextualização da pesquisa. Na Seção 1.2, são explicitadas as motivações. Em seguida, na Seção 1.3, são apresentados os objetivos desta pesquisa. Na Seção 1.4, são formuladas as questões de pesquisa que norteiam os resultados reportados nesta dissertação. Finalmente, na Seção 1.5, a estrutura da dissertação é apresentada.

### 1.1 Contextualização da Pesquisa

No mundo da arte, a caricatura é tradicionalmente descrita como uma representação exagerada do rosto de um indivíduo, conforme percebido e visualizado por um artista ou caricaturista. A caricatura tem como objetivo chamar a atenção para as características faciais de um indivíduo ou enfatizá-las, em especial, aquelas que as tornam singularmente identificáveis como um indivíduo.

O desenho, por si próprio, é uma linguagem universalmente compreendida e um esboço simples pode transmitir muito mais informações do que, talvez, uma página inteira de texto (RICHMOND, 2011). Neste contexto, a caricatura pode ser considerada como uma minibiografia ilustrativa do assunto. Artistas e caricaturistas normalmente procuram criar uma caricatura que não apenas produza uma semelhança divertida do tema, mas também expresse uma parte de sua vida naquele momento atual.

O termo *caricatura* é derivado do italiano *caricare* que, com a conotação de representar

graficamente, significa desenhar exagerando certos detalhes (*gravare con un peso eccessivo*). Esta forma de arte recebeu esse nome pois a principal intenção de um caricaturista é “carregar/inserir” tanto exagero sobre um assunto quanto possível, utilizando-se de um efeito tão cômico e agradável que capta a personalidade do sujeito (BRENNAN, 1982).

A palavra caricatura também já foi sinônimo de algo ridículo e grotesco. No entanto, é importante notar que uma caricatura normalmente não visa degradar a aparência da pessoa, distorcendo suas características menos favoráveis, mas, em vez disso, o foco deve ser exagerar a face de tal forma que proporcione entretenimento e diversão para o alvo.

Por muito tempo, os caricaturistas têm permanecido em locais populares de cidades turísticas e movimentadas, atraindo os transeuntes e vendendo suas criações como lembranças. Mais recentemente, caricaturistas *in loco* tornaram-se cada vez mais populares em festas e casamentos, que até então, era um meio inexplorado de caricaturização baseada em entretenimento.

Novas ideias e desenvolvimentos são introduzidos o tempo todo e a tecnologia continua avançando e, com as pessoas se tornando cada vez mais exigentes quanto às suas expectativas. Portanto, melhorias contínuas em todas as áreas da sociedade são necessárias para manter o interesse e a demanda do público. Neste contexto, passou a ser mais comum o emprego de ferramentas de software pelos caricaturistas, a exemplo do *Adobe Photoshop*. Tais ferramentas, entre outras vantagens, auxiliam o aprimoramento das criações desenhadas a mão e aceleram a produtividade do artista e a disseminação de suas obras.

## 1.2 Motivações

Há três elementos essenciais que uma caricatura não deve deixar de conter. A *semelhança*, que torna reconhecíveis e identificáveis características daquilo que é representado pela caricatura (coisa ou indivíduo). Outro elemento é o *exagero*, segundo o qual detalhes daquilo que é representado (coisa ou indivíduo) são excessivamente valorizados, a fim de realçar na representação gráfica características originais para as quais se quer chamar a atenção, diferentemente do que ocorre em um retrato. Por fim, é necessário que a caricatura tenha algo a dizer sobre o alvo a ser caricaturado. Pode ser algo sobre a situação da imagem, da personalidade do indivíduo caricaturado, por meio de expressões ou linguagens corporais ou

algum aspecto que manifeste alguma emoção à imagem (RICHMOND, 2011).

O processo de elaboração de uma caricatura envolve duas etapas básicas: a observação e o exagero (SADIMON SURIATI et al., 2010). A *observação* é considerada uma das etapas mais importantes na elaboração da caricatura. Os caricaturistas faciais precisam entender o que veem na face do indivíduo, observar quais os traços faciais que são mais diferentes daqueles apresentados pela maioria dos indivíduos e extrair esses dados para sua interpretação pessoal.

De acordo com especialistas, os seres humanos gravam uma face média em seu cérebro, que seria o resultado da média de todas as faces até então visualizadas (CLARKE; CHEN; MORA, 2011). Tal face média atuaria como uma referência para a determinação das características mais distinguíveis de uma face. No entanto, os caricaturistas consideram a diferença não só a partir da face média, mas também dos elementos visuais relacionados à característica a ser deformada (RICHMOND, 2011).

Com o advento da Internet e da Tecnologia Móvel, a caricatura também começou a ser empregada como um veículo de comunicação social e entretenimento, a partir do qual o indivíduo a ser caricaturado pode proteger sua identidade e imagem real de outros, para fins de segurança, mas ainda permitir que suas características faciais básicas sejam reconhecíveis.

Uma caricatura é definida como digital quando seu processo de criação ocorre a partir do uso de editores gráficos de software, e.g., *Macromedia Flash*, *Adobe Photoshop*, *Adobe Illustrator*. Caricaturas geradas por computadores são o resultado do uso de técnicas automáticas ou semiautomáticas de Computação Gráfica, normalmente com a finalidade de imitar as caricaturas produzidas por um artista utilizando papel e lápis (AKLEMAN, 2005).

O processo da geração automática de caricaturas por computadores pode ser dividido em duas categorias principais: (1) abordagem centrada no ser humano, que se utiliza de palavras previamente estabelecidas para incorporar distorções à imagem caricaturada; e (2) abordagem focada no processamento de imagens, que é fundamentada no emprego de imagens de faces como entrada para o processo de caricaturização (SADIMON SURIATI et al., 2010).

As caricaturas faciais podem ser geradas a partir das descrições verbais das características faciais e das impressões sobre a face analisada (CLARKE; CHEN; MORA, 2011). Em tais abordagens, pode-se fazer uso da teoria *fuzzy* para traduzir o significado do termo linguístico e atribuir um valor ao parâmetro de cada característica facial. O usuário, por sua vez, utiliza

palavras previamente definidas para descrever a caricatura desejada. Caso seu resultado não esteja de acordo com aquele esperado, alterações/correções podem ser incorporadas aos termos linguísticos pré-definidos. Os problemas que são encontrados nestas abordagens dizem respeito aos significados ambíguos dos termos linguísticos, à similaridade entre as características faciais e os termos pré-definidos, ao número limitado de expressões definidas e, por fim, aos significados dos termos referentes ao valor de exagero a ser conferido às características faciais.

Por sua vez, as caricaturas faciais baseadas no processamento de imagem, utilizam imagens 2D da face como dados de entrada para a geração das caricaturas (DU, X. et al., 2015). Tenta-se imitar a forma como os caricaturistas atuam, observando-se a imagem do rosto ou um indivíduo real.

Baseando-se no recente avanço tecnológico na geração automática de caricaturas a partir de imagens estáticas, verifica-se a necessidade de investigações que focalizem a geração automática de faces humanas caricaturadas em vídeos. A mera aplicação a vídeos de técnicas originalmente concebidas para imagens estáticas não garante a coerência temporal entre os quadros deformados, tampouco a manutenção da regularidade e estilo das caricaturas faciais ao longo de todos os quadros do vídeo considerado.

### **1.3 Objetivos e Relevância**

O objetivo geral desta pesquisa é **propor uma abordagem para a geração automática de distorções faciais em vídeos digitais que se assemelhe às distorções realizadas em caricaturas faciais.**

Para atingi-lo, identificou-se previamente a necessidade de atingir uma série de objetivos específicos, assim expressos:

- Investigar técnicas que permitam aprimorar o processo de detecção dos pontos da face em vídeos;
- Pesquisar regras de deformação que melhor se adaptem ao processo de geração automática de caricaturas em vídeos;

- Investigar possíveis problemas no processo de geração de vídeos caricaturados, objetivando-se a proposição de uma abordagem temporalmente coerente;
- Desenvolver um sistema de software que implemente a abordagem proposta;
- Validar a abordagem proposta a partir de um estudo de caso experimental.

## 1.4 Questões de Pesquisa

Visando-se a atingir os objetivos pré-definidos no presente estudo as seguintes questões de pesquisa (QP) foram formuladas:

- **QP1:** É possível conceber uma abordagem que produza deformações de faces humanas em vídeos visando à caricaturização, as quais apresentem coerência temporal e sejam bem avaliadas por observadores humanos?
- **QP2:** Qual nível de deformação é mais adequado para um conjunto de vídeos contendo faces deformadas como resultado da abordagem proposta?

## 1.5 Estrutura da Dissertação

O presente documento foi estruturado em seis capítulos. Além do que foi tratado neste capítulo, este documento contém mais cinco capítulos. No Capítulo 2, apresenta-se uma fundamentação teórica dos principais tópicos tratados nesta pesquisa. No Capítulo 3, são descritas e comentadas pesquisas relacionadas identificadas durante a revisão da literatura da área. No Capítulo 4, descreve-se a abordagem proposta para a geração automática de caricaturas. No Capítulo 5, são apresentados e discutidos os experimentos realizados e seus resultados. Finalmente, no Capítulo 6, são apresentadas as considerações finais deste estudo, as contribuições e recomendações para pesquisas futuras passíveis de derivação desta pesquisa.

# Capítulo 2

## Fundamentação

Neste capítulo, apresentam-se os conceitos necessários para a compreensão da pesquisa realizada. Na Seção 2.1, são apresentadas as técnicas de detecção de pontos fiduciais em faces. Na Seção 2.2, exibem-se os métodos de definição das características a serem deformadas. Na Seção 2.3, são apresentados os principais conceitos para a transformação espacial da imagem. Por fim, na Seção 2.4, são aprofundadas as definições de uma caricatura.

### 2.1 Detecção de Pontos Fiduciais em Faces

*Pontos fiduciais* são pontos de controle sobre um objeto que definem regiões características com propriedades que sejam de interesse ou relevância para o processo de detecção intentado (RIBEIRO et al., 2012).

Detectar pontos de interesse (*landmarks*) em faces é uma etapa essencial em sistemas que envolvem tarefas de reconhecimento facial, interação homem-máquina, animação facial e análise de expressões faciais. Quanto mais preciso for o detector de pontos de interesse, melhor será a qualidade do sistema. No entanto, segundo Sohail e Bhattacharya (2008) e Valstar et al. (2010), muitas dificuldades são encontradas ao desenvolver métodos para a detecção automática de pontos de interesse faciais, tais como: variação interpessoal (e.g., sexo e raça), intrapessoal (e.g., mudanças de pose e expressão) e inconsistência das condições de aquisição (e.g., iluminação e resolução da imagem). Na Figura 2.1, apresentam-se exemplos das variações existentes no processo de detecção dos pontos fiduciais da face.

Figura 2.1 Exemplos das variações existentes no processo de detecção dos pontos fiduciais faciais



Fonte: (SAGONAS; TZIMIROPOULOS et al., 2013)

O processo de detecção dos pontos de interesse da face é um problema de difícil resolução, para o qual as abordagens mais modernas emprestam ideias de um grande número de áreas afins, incluindo Visão Computacional, Geometria Computacional, Aprendizado de Máquina e Processamento de Imagens. Este problema de pesquisa tem sido estudado há mais de duas décadas, mas só recentemente foi possível o desenvolvimento de abordagens robustas e computacionalmente viáveis.

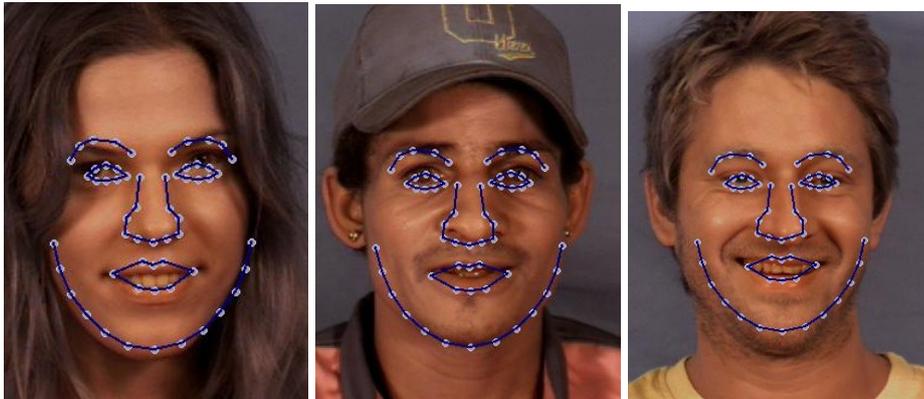
O rastreamento dos pontos da face foi popularizado pela primeira vez no início e meados dos anos 90, com o advento de modelos de forma ativa (*Active Shape Model* - ASM) por Cootes, Baldock e Graham (2000). Desde então, muitas pesquisas têm se dedicado a resolver o difícil problema do rastreamento genérico de faces. Um marco importante foi a extensão do ASM para modelos de aparência ativos (*Active Appearance Model* - AAM) em 1998, proposto por Cootes, Edwards e Taylor (1998). Em outra vertente de pesquisa, Blanz e Vetter (2003) focalizaram a modelagem de morfismo tridimensional (*Morphable Model*- 3DMM) que, assim como AAM, não só permite modelar as texturas da imagem, mas também possibilita a representação dos modelos com alta densidade de dados 3D obtidos a partir de varreduras a laser de faces.

A detecção de pontos fiduciais pode ser realizada manualmente ou automaticamente. Nesta etapa, são utilizadas diversas técnicas de processamento de imagem, tais como o pré-processamento, a detecção, localização e segmentação da área e a extração de suas características.

No contexto de pontos fiduciais da face, Monzo et al. (2010) compararam diversas propostas para a extração dos pontos fiduciais da face e analisaram sua influência na resolução dos problemas de reconhecimento facial. Para realizar a comparação, os autores utilizaram o mesmo número de pontos fiduciais e o mesmo tipo de descritores de orientação do gradiente (*Histogram of oriented gradients* - HOG) para cada proposta analisada. A localização precisa dos pontos fiduciais da face tem impacto positivo nas taxas de reconhecimento facial. Na Figura 2.2, são ilustrados exemplos dos resultados obtidos a partir da detecção dos pontos fiduciais da face a partir da técnica AAM. O modelo do AAM pode ser descrito pela Equação 2.1, no qual,  $\bar{x}$  representa a média da forma,  $P_s$  é um conjunto de modos ortogonais de variação e  $b_s$  é um conjuntos de parâmetros de forma.

$$x = \bar{x} + P_s b_s, \quad (2.1)$$

Figura 2.2 Pontos fiduciais da face, extraídos a partir da técnica AAM.



Fonte: Autoria própria

O Modelo Ativo de Forma (*Active Shape Model* - ASM), proposto por Cootes, Baldock e Graham (2000) consiste em um modelo estatístico da forma dos objetos. No ASM, a forma de um objeto é representada por um conjunto de pontos que são definidos pelas ordenadas e abscissas do plano da imagem. O modelo é empregado para descrever a forma e suas variações típicas a partir da variação da posição de cada ponto fiducial nas imagens de treinamento. A partir de um dado conjunto  $s$  de pontos  $x_i$ , alinhados com as coordenadas do quadro de vídeo, é possível gerar novos exemplares similares ao conjunto treinado. Para tanto, aplica-se

a Análise de Componentes Principais (*Principal Component Analysis* - PCA), de forma que se pode aproximar qualquer conjunto de treinamento  $x$ , conforme a Equação 2.2,

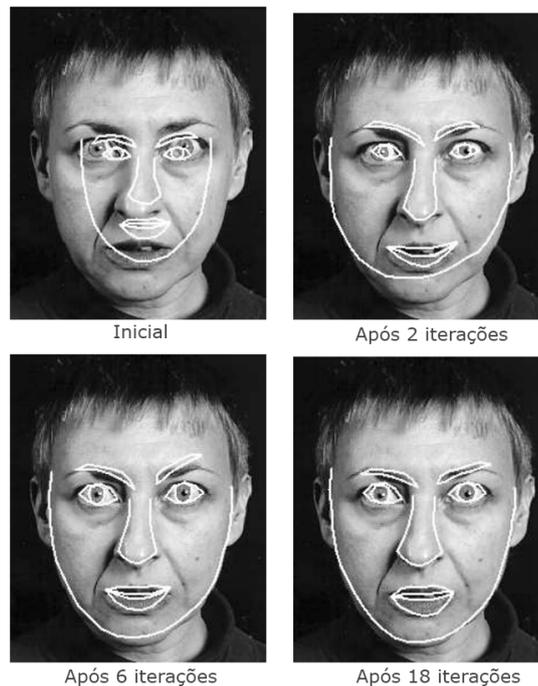
$$x \approx \bar{x} + Pb, \quad (2.2)$$

em que  $P = (P_1|P_2|\dots|P_t)$  contém  $t$  autovetores da matriz de covariância,  $b$  é um vetor  $t$  dimensional e  $\bar{x}$  é a média dos conjuntos de pontos, descrito pela Equação 2.3

$$b = P^T(x - \bar{x}). \quad (2.3)$$

Na Figura 2.3, é possível observar o processo de busca dos pontos fiduciais da face em um determinado período de iterações do algoritmo. Ao iniciar, o algoritmo apresenta um modelo de pontos genérico conforme treinado anteriormente e com o passar das iterações, o algoritmo ajusta o modelo de pontos de acordo com as características faciais apresentadas.

Figura 2.3 Busca dos pontos fiduciais da face utilizando ASM



Fonte: Adaptado de Cootes, Baldock e Graham (2000)

## 2.2 Definição das Características a serem Deformadas

Após todos os pontos fiduciais da face serem extraídos da imagem da face de entrada, faz-se necessário verificar e determinar quais as características irão ser exageradas.

Conforme Sadimon Suriati et al. (2010), as técnicas de exagero das características da face para a geração de caricaturas podem ser divididas em três grupos, a saber: (i) interativas; (ii) fundamentadas em regras; e (iii) fundamentadas no treinamento.

Nas técnicas *interativas*, o nível de exagero e seu formato são determinados interativamente pelo usuário. O usuário pode arrastar os pontos fiduciais da face para aumentar ou reduzir as dimensões entre pontos, com o intuito de gerar a caricatura. Clarke, Chen e Mora (2011) propuseram uma ferramenta que permitia ao usuário definir estilos de deformação para realizar o exagero. Na Figura 2.4, é apresentado um exemplo da utilização do método de deformação interativa, a partir do qual o usuário informa ao sistema as características que deverão ser deformadas (ZHAO et al., 2019).

Figura 2.4 Exemplo de método de deformação interativo



Fonte: Zhao et al. (2019)

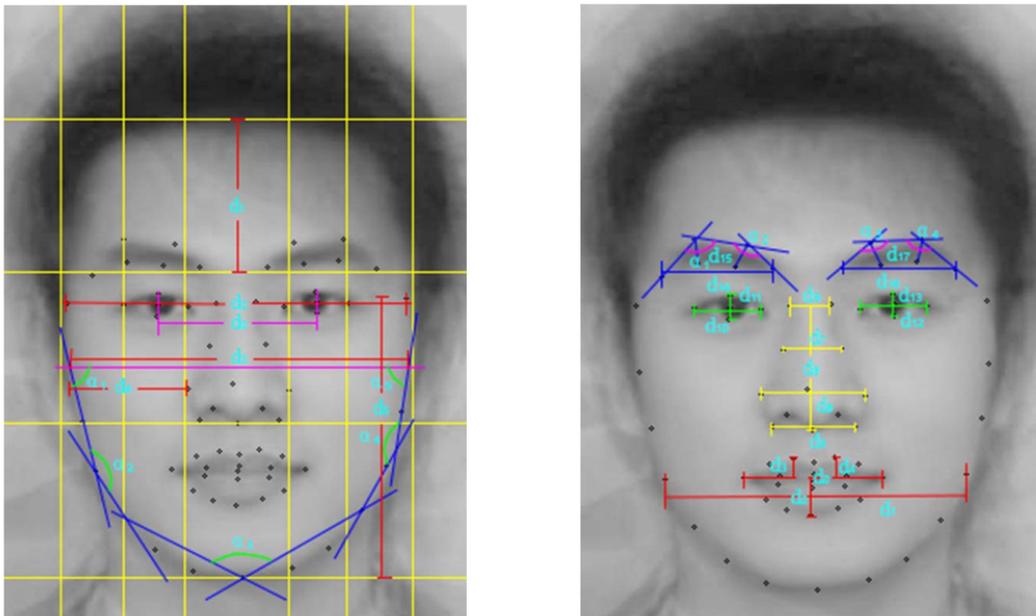
Para a representação das características a serem deformadas, cada ponto da face é associado a um estilo de deformação que corresponde a alguma propriedade material, e.g., a borracha para definir características mais deformáveis ou o aço para características mais rígidas. Entretanto, estes estudos requerem um alto grau de conhecimento do usuário, o que limita a sua aplicabilidade ao uso no cotidiano.

As *técnicas fundamentadas em regras* tentam formalizar, automática ou semiautomáticamente, as etapas do processo de geração da caricatura. De acordo com a teoria e as regras para desenhar uma caricatura, a imagem de entrada necessita ser comparada a alguma face, com o intuito de determinar as características discrimináveis (RICHMOND, 2011). A face de

referência utilizada pode ser uma face média ou um modelo padrão de face.

Em seu estudo, Xiaorong Du et al. (2015) propuseram, o princípio do contraste para a produção da deformação, evidenciando as características. Por exemplo, caso a face seja longa, é deformada a altura da face e sua largura é reduzida para realçar a deformação. Por sua vez, Li e Miao (2013) empregaram regras fixas para produzir distorções de faces em vídeos. Han, Song e Yoon (2010) empregaram dados de imagens previamente selecionadas para servir de regra e adequar a distorção com a caricatura da imagem. Na Figura 2.5, é apresentado um exemplo da utilização das regras fixas para servir de parâmetro para a detecção das características a serem exageradas.

Figura 2.5 Distâncias proporcionais utilizadas para a deformação facial



Fonte: Adaptado de Li e Miao (2013)

Vários autores utilizam a noção de exagero a partir da diferença da média - EDFM (*exaggerating the difference from the mean*), formalizada na Equação 2.4, em que os parâmetros  $Q$ ,  $P$ ,  $b$  e  $S$  representam, respectivamente, a caricatura gerada, a imagem de entrada, a taxa de exagero e a face média.

$$Q = P \times b(P - S). \quad (2.4)$$

As técnicas fundamentais no treinamento são usualmente utilizadas para capturar o estilo

de um artista ou para simular a criatividade do caricaturista. Não é simples codificar na forma de algoritmos o estilo particular de um artista ao exagerar determinadas características faciais. Esta família de técnicas necessita de um extenso banco de dados para treinamento, de modo a conter os pares de imagens, sendo a imagem original e a imagem correspondente de uma caricatura desenhada por um artista (SADIMON SURIATI et al., 2010). Liang et al. (2002) propuseram um sistema de geração de caricaturas fundamentado em exemplos de treinamento, em que é utilizada a regressão por mínimos quadrados parciais (PLS) para classificar 92 pares de imagens de face caricaturadas em 28 protótipos. Um modelo linear local é utilizado para estimar as características distintas da face a ser exagerada e seu nível de exagero. O principal problema desta técnica reside na obtenção de uma base de dados que contenha, antecipadamente, todas as caricaturas a serem utilizadas na etapa de treinamento.

## 2.3 Transformação da Imagem

Após a etapa de detecção das características a serem deformadas, a próxima etapa é a transformação da imagem original na imagem deformada. Nesta etapa, são utilizadas técnicas destinadas à transformação da imagem facial de entrada em uma imagem facial caricaturada de saída. Esta é a última etapa da geração automática de caricaturas. A imagem facial de entrada é transformada na caricatura facial desejada a partir de técnicas de deformação (*morphing*) da imagem. Existem duas técnicas básicas de deformação de imagens: a transformação de campo e a transformação de malha (WANG; LI et al., 2013).

O método de *transformação do campo* é mais fácil de ser aplicado do que o método de *transformação da malha*, embora exija usualmente maior tempo de processamento. A deformação será definida como uma transformação  $W : UU'$  tal que sua restrição ao domínio  $s_i$ ,  $W|_{s_i}$ , coincide com a transformação  $W_i : s_i d_i$ . Em outras palavras: as transformações  $W_i$  são estendidas para se obter a deformação  $W$ .

Malhas regulares, além de simplificarem a manipulação de dados na implementação, também facilitam a compreensão do processo pelo usuário. Em qualquer um dos casos, o usuário é responsável por ajustar as duas malhas para que indiquem a transformação desejada entre elementos correspondentes das malhas. Existem casos particulares de malhas que são especialmente relevantes. Malhas poligonais, por exemplo, são mais fácil entendimento pelos

usuários, sendo empregadas frequentemente com graus de interpolação mais elevados.

O uso de malhas de *splines*<sup>1</sup> sugere uma transformação naturalmente suave. Apesar das malhas de *splines* permitirem uma implementação extremamente eficiente, este tipo de especificação é às vezes restritivo. Posicionar os pontos da malha sobre as características da imagem relevantes ao contexto da deformação nem sempre constitui uma tarefa trivial. Além disso, a regularidade da malha gerada nem sempre é adequadamente compatível com a estrutura original da imagem associada. (DAVIS et al., 1997).

Um caso importante de deformação fundamentada em características reside na especificação fundamentada em pontos, na qual cada característica é indicada por um ponto na imagem. Um modo simples de especificar uma deformação genérica em uma imagem é empregar malhas triangulares. Deformações descritas desta forma podem ser produzidas por um algoritmo que utiliza um sistema de coordenadas em cada triângulo, no qual, qualquer triângulo possa ser mapeado em qualquer outro triângulo a partir de uma transformação linear (ZOPE; ZOPE, 2017).

Se as malhas triangulares tiverem as mesmas propriedades combinatórias, esta transformação pode ser estendida para uma transformação global linear por partes que implementa a mudança de coordenadas. Na Figura 2.6, são apresentados exemplos das malhas geradas a partir de um conjunto de pontos determinados e suas possibilidades de divisões triangulares.

O uso de *splines* para especificar as malhas regulares que definem o sistema de coordenadas apresenta diversas vantagens. Em primeiro lugar, o uso de uma estrutura regular, ao invés de uma topologia arbitrária, simplifica significativamente tanto as estruturas de dados quanto as funções de mapeamento.

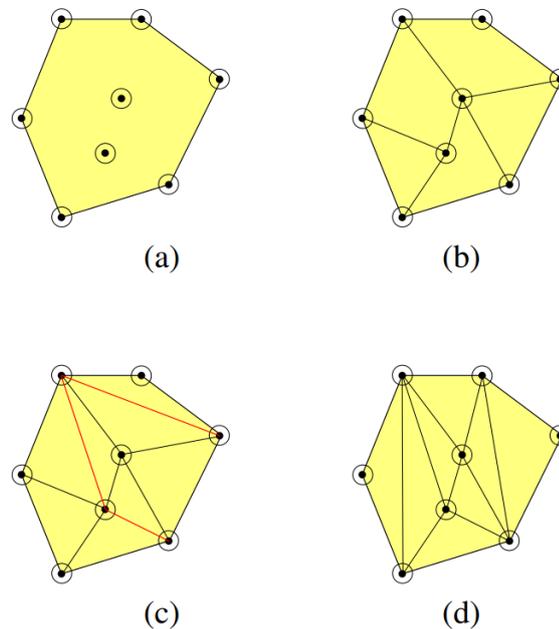
Além disso, pode ser demonstrado que este mapeamento é separável, i.e., pode ser realizado em dois passos independentes, a saber: (i) um horizontal, que produz uma imagem intermediária horizontalmente deformada; e (ii) outro vertical, que usa esta imagem intermediária para produzir a saída final. Isto transforma o problema de deformar uma imagem 2D em um problema unidimensional, reduzindo de modo significativo a complexidade dos cálculos, especialmente para *anti-aliasing*<sup>2</sup> (MACEDO; APOLINÁRIO, 2018).

---

<sup>1</sup>*Splines* são funções de regressão segmentada, unidas por pontos chamados de nós, com capacidade de melhorar a curvatura dos modelos, melhorando assim o ajuste de funções.

<sup>2</sup>Método de redução de serrilhamento, que é o efeito em forma de serra que se cria ao desenhar uma reta inclinada em uma tela de computador.

Figura 2.6 Exemplos do uso das malhas triangulares em um conjunto de pontos e suas diferentes possibilidades de triangulações.



Fonte: (MONDAL; BISWAS, 2020)

A técnica de deformação *por campo* fundamenta-se no emprego de “campos de influência” ao redor das principais características da imagem. Essas características são definidas em seus estados relaxado e deformado mediante o emprego de segmentos de reta orientados na abordagem proposta por Beier e Neely (1992). Neste caso, cada segmento de reta posicionado sobre uma característica da imagem define um sistema de coordenadas  $u - v$ , no qual  $u$  representa a posição relativa ao longo do segmento, e  $v$  representa a distância ortogonal absoluta até o segmento.

O uso de mais de um segmento de reta produz um sistema de coordenadas conflitante, cujas coordenadas precisam ser combinadas de uma forma suave e controlada. Beier e Neely (1992) propuseram a atribuição de um peso a cada segmento: mais forte para pontos sobre o segmento e decaindo à medida que os pontos distanciam-se deste. Cada pixel da saída é mapeado de volta para o sistema de coordenadas de origem de cada segmento, produzindo o mapeamento de um conjunto de pontos distintos. Os deslocamentos do pixel de saída para cada um dos pontos mapeados são empregados em uma média ponderada para calcular o deslocamento final, fazendo-se uso do peso  $w$  definido pela Equação 2.4

$$w = (l^p/a + d)^b, \quad (2.5)$$

em que,  $l$  representa o comprimento de um segmento,  $d$  representa a distância do pixel de saída ao segmento, e  $a$ ,  $b$  e  $p$  são constantes usadas para ajustar a deformação.

## 2.4 Caricatura

A palavra Caricatura, do italiano *caricare*, significa “carregar, no sentido de exagerar, aumentar algo em proporção”. A caricatura está presente em jornais, revistas, blogs, convites de casamento, festas de aniversário, charges, tirinhas, entre outros suportes ou gêneros discursivos textuais. Uma caricatura visa a atenção ou a ênfase das características faciais de um indivíduo que os tornam exclusivamente identificáveis como pertencentes a um indivíduo específico. Desenhar uma caricatura não é meramente realçar as orelhas de uma pessoa deixando-as maiores ou seu nariz menor, mas em vez disso, fornecer um meio de expressar a personalidade de um sujeito por meio da arte. O desenho é uma linguagem universalmente compreendida e um simples esboço pode muitas vezes passar muito mais informações do que talvez uma página inteira de texto.

O objetivo por trás da criação de caricaturas pode variar, desde empregá-las como ferramentas úteis para comunicar uma mensagem política até criá-las puramente para fins de entretenimento. Para um artista, é muito fácil desenhar uma caricatura de maneira favorável ou depreciativa. Como resultado, caricaturas podem expressar diversos tipos de informações visuais em um formato que seja facilmente acessível e compreensível por uma grande quantidade de pessoas.

Por esse motivo, políticos e celebridades são frequentemente retratados na forma de caricaturas em revistas e jornais populares. Nesta situação, o artista pode usar informações atuais para influenciar o tipo de caricatura que eles criam. Ao se fazer uso de caricaturas dessa forma muitas vezes é possível se ter uma grande influência sobre os sentimentos do público-alvo, portanto, qualquer revista ou jornal que queira expressar uma determinada mensagem sobre um indivíduo pode fazê-lo de forma muito eficaz usando uma caricatura. Por outro lado, as caricaturas baseadas em entretenimento são geralmente usadas para expressar mensagens menos controversas e são mais voltadas à simples geração de uma imagem humorística que

proporcione um efeito agradável. Na Figura 2.7, são apresentados exemplos de caricaturas habitualmente encontradas no cotidiano das pessoas.

Figura 2.7 Exemplos de caricaturas.



Fonte: (HUO et al., 2017).

Caricatura digital ou gerada por computador é uma arte relativamente complexa e ainda é um problema não integralmente solucionado pela comunidade de pesquisa de Computação Gráfica. Esta área está se tornando cada vez mais atrativa para os pesquisadores, sendo que várias propostas para a criação automática de caricaturas digitais foram apresentadas nos últimos anos (CAO; LIAO; YUAN, 2018).

Uma das principais dificuldades associadas à geração de caricaturas por computador é que a maioria dos métodos existentes depende de um grau considerável de interação com o usuário para funcionar com eficácia (HILL et al., 2019). Muitos sistemas não são automatizados e exigem que o usuário mova manualmente partes do rosto com a finalidade de criar uma caricatura. Embora esta possa ser considerada uma tarefa divertida e agradável para o usuário, o problema com esses sistemas é que pressupõem que o usuário tenha alguma habilidade artística e conhecimento sobre o que são caricaturas e como são criadas.

O usuário também deve ser capaz de concluir com precisão quais características exagerar e quais ignorar. Tal tarefa pode não ser simples para muitos artistas. Há sistemas totalmente automatizados que superam esta limitação, no entanto, podem criar uma nova limitação porque, ao se removerem todas as entradas do usuário para o processo de caricaturização, o “toque pessoal” de talento artístico individual que está presente em caricaturas desenhadas à mão também é perdido (HAN; HOU et al., 2018).

## **2.5 Considerações Finais**

Neste capítulo, foram apresentados os principais conceitos relacionados às técnicas de geração automática de caricaturas faciais. Foram descritos conceitos relativos à detecção de pontos fiduciais, métodos de definição das características a serem transformadas e técnicas de transformação espacial de imagens.

No capítulo seguinte, são apresentados os principais estudos relacionados a esta pesquisa.

# Capítulo 3

## Pesquisas Relacionadas

O crescente avanço nas áreas de Computação Gráfica e Visão Computacional impulsionou o progresso de estudos nas áreas de processamento de imagens não-fotorrealísticas. Em particular, na área de geração automática de caricaturas, diversos estudos objetivam simular a geração automática de caricaturas assim como imitar o modo como os caricaturistas produzem suas obras (LI; MIAO, 2013).

De modo simplificado, o processo de criação automática de uma caricatura consiste em, primeiramente, extrair os pontos fiduciais da face da imagem original, em seguida encontrar quais características da face são distintas em relação a uma face de referência e, por fim, produzir a versão caricaturada da face original, de acordo com suas características mais distintas.

Esta seção foi estruturada em quatro seções. Na Seção 3.1, é apresentado um breve histórico da geração de caricaturas de forma automática. Na Seção 3.2, apresentam-se as principais técnicas de detecção da face. Na Seção 3.3, são apresentadas as técnicas utilizadas para a detecção de pontos fiduciais. Na Seção 3.4, são apresentadas principais técnicas para a escolha das características mais relevantes. Por fim, na Seção 3.5, são apresentadas as considerações finais deste capítulo.

### 3.1 Breve Histórico da Geração Automática de Caricaturas

Nos últimos anos, o tema de geração de caricaturas por computador tem atraído uma quantidade crescente de pesquisas. Os estudos existentes podem ser divididos em duas categorias

distintas de abordagens, a saber: (i) *interativas*; e (ii) de *geração automática*. As abordagens *interativas* são concebidas para oferecer ao usuário elevado controle do processo. Os sistemas que implementam essas abordagens atuam tão somente como uma ferramenta para fornecer como resultado uma caricatura da face. As abordagens automáticas, no entanto, são muito mais desafiadoras tecnicamente, uma vez que seu objetivo é atuar como um sistema que requeira o mínimo de interação com o usuário, mas que ainda possa produzir resultados agradáveis.

Brennan (1982), apresentou a primeira contribuição significativa nesta área, ao desenvolver um sistema interativo destinado ao delineamento de caricaturas. A autora introduziu a ideia de exagerar as diferenças entre uma face-alvo e uma face média. Seu sistema permitia ao usuário a especificação de menos de 200 pontos fiduciais em uma imagem a ser caricaturada e, a partir dos pontos, um esboço da imagem (formado por linhas) era gerado a partir da interpolação de uma curva entre os pontos fiduciais. O resultado era comparado a uma imagem de referência pelo sistema e o usuário podia, então, interagir com o esboço delineado por meio de um controle deslizante do grau de exagero a ser produzido. Este estudo precursor é considerado como um dos mais importantes na área de geração automática de caricaturas por computador, uma vez que forneceu relevante contribuição técnica. Na Figura 3.1, apresentam-se exemplos de resultados da face original como também as caricaturas geradas com diferentes níveis de exagero.

Figura 3.1 (A) Face original, (B) 50% de exagero, (C) 100% de exagero, (D) 140% de exagero, (E) 160% de exagero



Fonte: Frowd et al. (2007)

Akleman (2005) sugeriram uma abordagem de deformação fundamentada em uma estrutura denominada *simplex*, disponibilizando ao usuário uma ferramenta interativa para

deformar manualmente as características faciais. Segundo os autores, um *simplex* é definido como sendo um conjunto de  $d + 1$  pontos cuja envoltória convexa tem dimensão  $d$ . No contexto da geração de caricaturas, tal envoltória pode ser constituída por pontos, linhas ou triângulos, definindo as características faciais. Naquela pesquisa, a deformação é descrita por pares *simplex* que incluem um *simplex* de origem e outro de destino para a imagem. Para cada um deles, é definido um vetor de translação que descreve como o *simplex* de origem muda de tamanho e posição para alcançar a forma de destino. Uma função de ponderação também é definida para indicar o grau de exagero para cada fonte *simplex*, baseando-se na distância da origem até o destino. Combinando-se todos esses fatores, consegue-se, então, especificar a distorção de cada característica facial para a geração da caricatura.

Clarke, Chen e Mora (2011), afirmou que suas tentativas anteriores de geração de caricaturas interativas (incluindo suas próprias obras) haviam falhado devido à combinação de diversos fatores. O primeiro fator apontado pela autora foi que algumas abordagens destinadas ao delineamento dos alvos exploravam o fato de que, para a maioria dos indivíduos comuns, isso seria aceito como uma caricatura, dispensando a necessidade de qualquer outra manipulação posterior. Enquanto tais sistemas realmente fornecem um meio para exagero e manipulação, a sucessão de tais exageros tornava-se irrelevante, uma vez que o usuário já se convencera de que o resultado apresentado anteriormente já seria uma caricatura. O segundo fator apontado foi que a maioria dos indivíduos não entendia as regras para a criação de uma caricatura, exceto pelo seu aspecto lúdico. A autora afirmou que muitos sistemas produziam deformações incorretas que não eram, na verdade, caricaturas, mas para indivíduos comuns eram perfeitamente aceitáveis por sua natureza cômica. A autora também comentou que é relativamente fácil desenvolver um sistema que possa deformar um rosto, muito embora o real desafio seria desenvolver um sistema que pudesse criar uma caricatura visualmente precisa e agradável.

Uma abordagem fundamentada em exemplos foi proposta por Baltrušaitis, Robinson e Morency (2012), segundo a qual o estilo do caricaturista é “apreendido” pelo sistema a partir do emprego de mínimos quadrados parciais (*Partial least squares regression* - PLS). O sistema empregava um conjunto de dados de treinamento que consistia em 92 imagens faciais e suas correspondentes caricaturas desenhadas à mão pelo mesmo artista. Essas imagens eram, então, empregadas para aprender um modelo de exagero da forma, alinhando-os e identificando

tendências comuns no exagero. Esse Modelo de Exagero de Forma podia, então, ser aplicado a uma imagem, com o objetivo de gerar uma nova caricatura a partir do PLS. O PLS era empregado para exagerar espacialmente a forma analisada e, assim, produzir um conjunto de protótipos de distorção definidos por certas formas (protótipos) de face. A imagem era, então, classificada para se adequar a um desses protótipos de distorção, sendo então, linearmente exagerada pela tendência do protótipo definido. Na Figura 3.2, apresentam-se os resultados que o autor obteve utilizando essa abordagem.

Figura 3.2 (A) Imagens originais, (B) Imagens caricaturadas pelo sistema proposto



Fonte: Baltrušaitis, Robinson e Morency (2012)

Blanz e Vetter (2003) apresentaram um breve esboço de uma técnica, que identifica que é a abordagem padrão adotada pela maioria dos caricaturistas, que ele chamou de “Exagero da Diferença da Média” ou EDFM, nem sempre funciona bem para a caricaturização computadorizada. Como alternativa, os autores propõem uma nova abordagem que não só considera como uma característica difere em tamanho da média, mas também leva em consideração sua variação quando comparada a outras características. Por exemplo, uma boca que é 2cm maior que a média pode parecer razoavelmente normal. No entanto, um olho que é 2cm maior que a média pode parecer muito distinto. A abordagem tradicional EDFM, no entanto, exageraria estas duas características da mesma maneira, sem levar em conta a variância em relação ao rosto como um todo. A abordagem de Blanz e Vetter buscou resolver esse problema aplicando diferentes fatores de exagero para cada característica de acordo com a escala normalizada de diferenças.

Dal, Dal e Abraham (2011) propuseram um método de mapeamento de descrição gramatical generalizada para um modelo de subcaricatura. No processo de descrição das caracte-

rísticas faciais, uma estrutura de árvore foi empregada para representar as características e traduzir descrições verbais de tais características em descrições gramaticais correspondentes.

Liu et al. (2005) propuseram uma abordagem de aprendizado de mapeamento para a geração automática de caricaturas usando PCA (Análise de Componentes Principais - *Principal Component Analysis*). Os componentes principais das características faciais identificados a partir do PCA durante uma fase de treinamento do sistema. Uma imagem de origem e uma caricatura desenhada à mão são então inseridas no sistema, e se utiliza o aprendizado de máquina via SVR (*Support Vector Regression*) para treinar um modelo de mapeamento que determina e seleciona as informações importantes para a caricaturização. O mapeamento relaciona os componentes de entrada da imagem de origem para os componentes de saída da caricatura, aprendendo uma função de regressão que pode combinar os dois conjuntos de dados juntos. A SVR foi escolhida para realizar esta tarefa por seu desempenho quando se lida com regressão não-linear de alta dimensão. Na Figura 3.3, apresentam-se os resultados obtidos a partir da abordagem proposta por (LIU et al., 2005).

Chunhua Du et al. (2008) propuseram a geração de caricaturas propuseram a geração de caricaturas de descrição de palavras. Os autores calcularam a semântica das palavras descritoras da face original, usando um conjunto de conceitos nebulosos (*fuzzy*) e ajustado a face média no banco de dados a partir de uma análise semântica.

Feng e Ge (2012) propuseram um método usando formas primitivas combinadas com diferentes pesos para deformar uma face, ajustando-as posteriormente em função de modelos de desenhos animados contidos no dicionário de renderização. Os autores empregaram arte, regras psicológicas e uma máscara de desenho animado para exagerar as características do rosto do retrato da face frontal. Fujiwara et al. (2001) também estudaram o problema de caricaturização automática, mensurando faces de entrada a partir de um modelo de referência e exagerando aquelas características que mais diferiam mais do modelo de referência.

## 3.2 Detecção da Face

A detecção de faces é uma das tarefas visuais que o sistema visual/cérebro humano pode realizar sem esforço. No entanto, esta não é uma tarefa fácil quando se utilizam técnicas de visão computacional. Dada uma imagem estática ou ou dinâmica (sequência de quadros de

Figura 3.3 (A) Imagens originais, ((B)(C)) Imagens caricaturadas pelo sistema proposto



Fonte: Liu et al. (2005)

vídeo), o objetivo é detectar um número desconhecido de faces. A solução para o problema envolve segmentação, extração e verificação de faces e, possivelmente, características faciais em um *background* não controlado. Para solucionar este problema, uma grande variedade de técnicas foram propostas, variando de algoritmos simples baseados em borda até abordagens de alto nível utilizando métodos avançados de reconhecimento de padrões.

Estas abordagens podem ser divididas em três categorias. Dado um problema típico de detecção de faces em uma cena com muitos elementos visuais, a primeira categoria, referente à análise de baixo nível, lida inicialmente com a segmentação das características visuais

usando propriedades em nível de pixel, tais como escala de cinza e cor. Por causa da natureza de baixo nível, as características geradas a partir desta análise são ambíguas. Na categoria de análise de características, as características visuais são organizadas em um conceito mais global de rosto, incluindo atributos faciais, mediante informações da geometria do rosto. Por meio de um processo de análise, as ambiguidades entre características são reduzidas e as localizações do rosto e das características faciais são determinadas. A terceira categoria envolve o uso de modelos de forma ativos (*Active Shape Models*). Esses modelos vão desde *snakes*, propostos no final dos anos 1980, até os modelos mais recentes de distribuição de pontos (*Point Distribution Models* - PDM), os quais foram desenvolvidos com o propósito de extração de características complexas e não rígidas, tais como a pupila ocular e os contornos labiais.

### 3.3 Detecção de Pontos Fiduciais

Na pesquisa apresentada por Cootes, Edwards, Taylor et al. (1999), os autores empregaram AAM (*Active Appearance Model*) no processo de detecção dos pontos fiduciais. AAM fundamenta-se na combinação de um modelo estatístico de forma com um modelo de escala de cinza representando a aparência do objeto. Tal combinação alia as coordenadas de pontos fiduciais encontrados por modelos de forma com uma representação da textura de regiões locais no entorno dos pontos, a qual representa a aparência do objeto. O modelo de aparência pode representar a variação da forma e da textura apresentada no conjunto de treinamento. A forma de um objeto pode ser representada como um vetor  $x$  e a textura (ou níveis de cinza) pode ser representada como um vetor  $g$ , de acordo com as Equações 3.1 e 3.2, respectivamente.

$$x = \bar{x} + Q_s c, \quad (3.1)$$

$$g = \bar{g} + Q_g c, \quad (3.2)$$

em que,  $\bar{x}$  representa a forma média,  $\bar{g}$  representa a textura média e  $Q_s$ ,  $Q_g$  são as matrizes que descrevem os modelos de variação derivados do conjunto de treinamento.

Na pesquisa de Milborrow e Nicolls (2014), os autores propõem para seu estudo de localização de pontos fiduciais um modelo baseado no *Active Shape Model* de Tim Cootes,

o STASM. Este modelo utiliza uma forma simplificada método SIFT proposto por Brown e Lowe (2007) para descrever os modelos correspondentes, substituindo os perfis de uma dimensão utilizados no modelo clássico. Além disso, é utilizada uma regressão multivariada adaptativa para que a detecção ao redor dos pontos fiduciais seja mais eficiente.

A aplicação de software OpenFace, desenvolvida por Baltru, Robinson, Morency et al. (2016), adota campos de redes neurais condicionais locais (*Conditional Local Neural Fields - CLNF*) para a detecção e o rastreamento de pontos fiduciais da face. Os dois principais componentes que são utilizados no CLNF são: o modelo de distribuição de pontos (*Point Distribution Model - PDM*), que captura as variações da forma dos pontos fiduciais e os *patch experts* que capturam as variações locais de aparência para cada ponto fiducial.

O CLM (*Constrained Local Model*), proposto por Baltrušaitis, Robinson e Morency (2012), baseia-se em técnicas fundamentadas em modelos de distribuição de pontos (*Point Distribution Model - PDM*). Para modelar a aparência, não é necessária a utilização de toda a face, mas apenas os locais ao redor dos pontos fiduciais de interesse (similar à técnica ASM). Isto permite que o sistema seja mais generalista pois não é necessária toda a complexidade de um modelo de aparência da face por completo. O modelo CLM pode ser descrito pelos parâmetros  $p = [s, R, q, t]$ , que podem variar para se obter várias instâncias do modelo, em que  $s$  representa o fator de escala,  $R$  é a rotação do objeto,  $t$  é a translação em 2D, e  $q$  é um vetor com a variação da forma. O modelo de distribuição de pontos utilizado no CLM está formalizado na Equação 3.3.

$$x_i = s \times R(\bar{x}_i + \Phi_i q) + t, \quad (3.3)$$

em que,  $x_i = (x, y)$  define a localização do ponto  $i$  na imagem,  $\bar{x}_i$  representa o valor médio do elemento  $i$  do PDM, e  $\Phi_i q$  denota um vetor com o autovalor  $i$  obtido a partir do conjunto de treinamento.

Para realizar a comparação dos algoritmos de detecção de pontos fiduciais no contexto de imagens de raio-x cefalométricas, Wang, Huang et al. (2015) fizeram uso de um banco de imagens contendo cefalogramas de 300 pacientes entre seis e sessenta anos, manualmente marcadas por dois médicos experientes. A avaliação quantitativa foi realizada com base na comparação dos resultados apresentados pelos diversos algoritmos analisados com os dados originais marcados pelo médico.

Ainda no contexto de imagens médicas, Riegler et al. (2015) compararam algoritmos fundamentados em redes neurais convolucionais treinadas com dados sintéticos para a detecção de pontos fiduciais da anatomia humana.

No contexto de pontos fiduciais da face, Monzo et al. (2010) compararam diversas propostas para a extração dos pontos fiduciais da face e analisaram sua influência na resolução de problemas de reconhecimento facial. Para realizar a comparação, os autores utilizam o mesmo número de pontos fiduciais e o mesmo tipo de descritores HOG (*Histograms of Oriented Gradients*) para cada proposta analisada.

No estudo realizado por Sagonas, Tzimiropoulos et al. (2013), é apresentado um conjunto de diretivas que formam, segundo o autor, o primeiro padrão de referência para a avaliação de pontos fiduciais e que foi utilizado no primeiro desafio de localização de pontos fiduciais da face, realizado na conferência internacional de Visão Computacional de 2013, em Sydney (SAGONAS; ANTONAKOS et al., 2016).

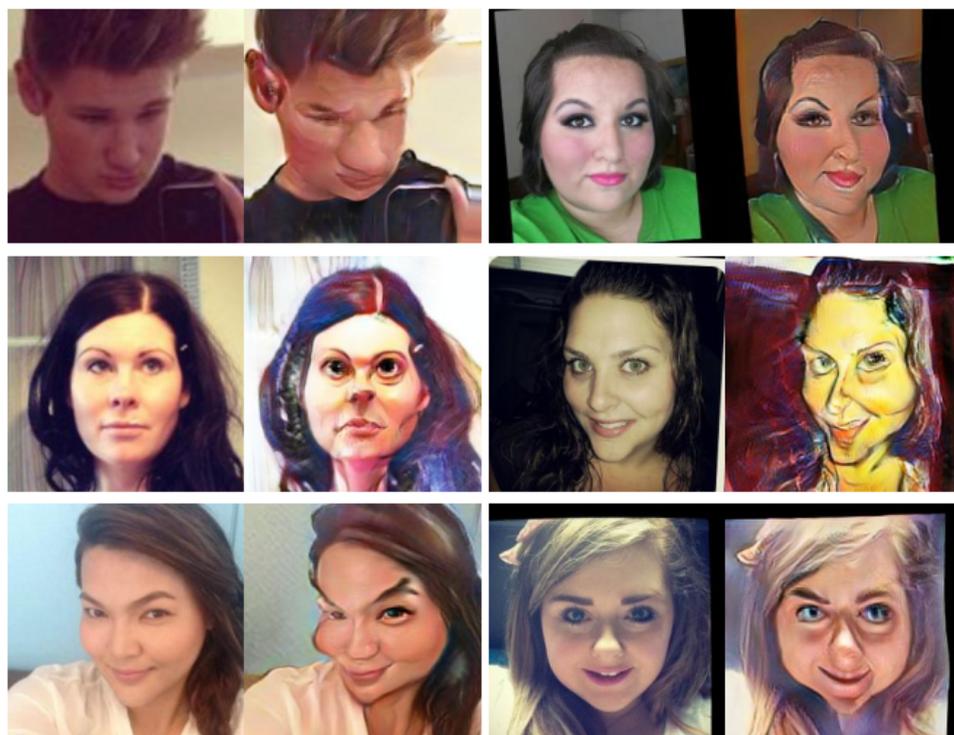
### 3.4 Regras de Definição das Características Relevantes

Shi, Deb e Jain (2019) propuseram uma rede neural (WarpGAN) para a geração de caricaturas a partir de uma imagem digital de entrada. O processo de deformação se dá a partir do módulo de distorção, encarregado de identificar os pontos de controle e, a partir desses pontos, distorcer a imagem para gerar uma caricatura. Na Figura 3.4, apresentam-se os resultados obtidos após a aplicação do modelo WarpGAN a algumas imagens de teste.

É importante observar que os estudos supracitados focalizam a geração automática de caricaturas no contexto de imagens estáticas. A utilização de tais técnicas, conforme descritas anteriormente, no contexto de vídeos digitais usualmente produzem artefatos indesejáveis entre os quadros, tendo em vista não considerarem aspectos de coerência temporal entre os quadros do vídeo. Objetivando solucionar estes problemas, estudos como aquele de Yang et al. (2012), por exemplo, levam em consideração o fluxo de informação temporal entre os quadros, a fim de que a deformação realizada torne-se temporalmente coerente.

No contexto da geração automática de vídeos, Tulyakov et al. (2018) propuseram uma rede adversarial generativa decomposta por movimento e conteúdo (*MoCoGAN - Motion and Content decomposed Generative Adversarial Network*). Por meio da aplicação deste modelo

Figura 3.4 Faces de entrada e suas respectivas caricaturas produzidas pelo WarpGAN



Fonte: Shi, Deb e Jain (2019)

de rede neural, é possível gerar vídeos de faces humanas com seis diferentes expressões faciais: raiva, medo, nojo, felicidade, tristeza e surpresa. Na Figura 3.5, apresentam-se os quadros das expressões faciais que vão sendo alteradas com o decorrer do vídeo, iniciando com felicidade, passando por medo e finalizando com surpresa.

Figura 3.5 *Frames* de vídeos gerados a partir de alterações das expressões faciais

Fonte: Tulyakov et al. (2018)

Para reduzir os ruídos entre os quadros e proporcionar uma coerência temporal satisfatória, o estudo de Zhao et al. (2019) propõe que as deformações faciais sejam transferidas do espaço 2D para o espaço 3D, de forma que possibilite que as modificações realizadas em um quadro se propaguem por todo o vídeo.

## 3.5 Considerações Finais

Neste capítulo, foram apresentadas as principais pesquisas na área de geração automática de caricaturas. Foram apresentadas as principais técnicas utilizadas para a detecção da face, técnicas utilizadas para a detecção dos pontos fiduciais e diferentes abordagens para a escolha das características faciais a serem deformadas. No Quadro 3.1, é apresentado um sumário com as principais características de abordagens relacionadas apresentadas anteriormente, em que também é possível comparar com a abordagem proposta neste estudo.

Nesta Tabela, é possível observar que a abordagem proposta apresenta distorções em vídeos com o auxílio de regras estatísticas. A abordagem proposta também tem a característica de utilizar uma estabilização temporal dos pontos fiduciais detectados (conforme apresentada na Seção 4.2). A abordagem proposta também faz uso de estimativa de pose 3D das faces, para que as distorções realizadas a partir das regras estatísticas sejam mais precisas (conforme discutido na Seção 4.4). No capítulo seguinte, são apresentadas as metodologias utilizadas neste estudo. Também é detalhada a abordagem proposta para a geração automática de caricaturas faciais em vídeos.

Quadro 3.1 Sumário das principais abordagens para geração automática de caricaturas

<b>Estudos Observados</b>	<b>Distorções em Vídeos</b>	<b>Estilização das Imagens</b>	<b>Regras de Deformação</b>	<b>Deteção de Pontos Fiduciais</b>
<b>(LIU et al., 2005)</b>			Regras Estatísticas	x
<b>Chunhua Du et al. (2008)</b>		x	Descrição Semântica	
<b>Baltrušaitis, Robinson e Morency (2012)</b>			Regras Estatísticas	x
<b>Tulyakov et al. (2018)</b>	x		GAN	x
<b>Shi, Deb e Jain (2019)</b>		x	GAN	x
<b>Zhao et al. (2019)</b>			Regras Fixas	x
<b>Abordagem Proposta</b>	x		Regras Estatísticas	x

Fonte: Autoria própria

# Capítulo 4

## Abordagem Proposta

Neste capítulo, é descrita a abordagem proposta para a geração automática de caricaturas faciais em vídeos. Na Seção 4.1, é apresentada uma visão geral da abordagem proposta neste estudo. Na Seção 4.2, são detalhados os métodos de detecção de pontos fiduciais da face. Na Seção 4.3, são definidas as regras de definição das características mais relevantes da face. Na Seção 4.4, são apresentadas as metodologias para realizar as deformações das faces presentes nos quadros do vídeo. Na Seção 4.5, são apresentados os detalhes do projeto desenvolvido e as principais ferramentas utilizadas para a implementação da abordagem proposta. Por fim, na Seção 4.7, são apresentadas as considerações finais do capítulo.

### 4.1 Visão Geral do Modelo Proposto

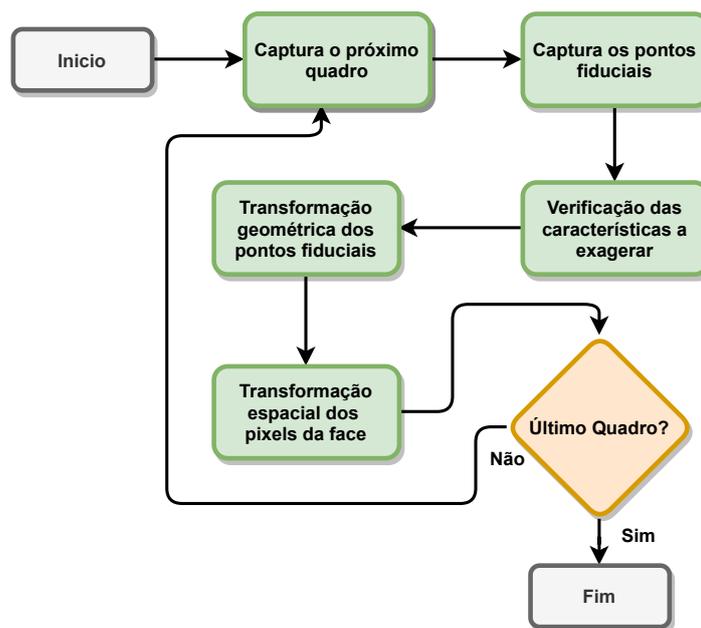
Na Figura 4.1, apresentam-se as etapas que compõem a abordagem proposta para a criação automática das caricaturas, cuja entrada é o vídeo a ser caricaturado e em cuja saída obtém-se o vídeo contendo a face caricaturada.

Na primeira etapa, o primeiro quadro do vídeo é capturado e analisado. Nesta etapa, também ocorre o pré-processamento na imagem, utilizando-se técnicas destinadas à redução de ruídos e à melhoria da iluminação. Em seguida, a face é detectada a partir das técnicas de histogramas de gradientes orientados (*Histogram of Oriented Gradient - HOG*) (NIGAM; SINGH; MISRA, 2018) e, após esta etapa, os pontos fiduciais da face são detectados.

Em seguida, por meio de regras fixas, são determinadas quais características faciais serão exageradas. Com base nessas características definidas, os pontos fiduciais detectados são

transformados para produzir o exagero. De posse das novas coordenadas transformadas, ocorre o processo de distorção (*morphing*) dos *pixels* da imagem, produzindo como saída um novo quadro do vídeo contendo a face caricaturada. O processo geral é repetido até que todos os quadros do vídeo tenham sido processados. Nas próximas seções, são apresentados detalhamentos de cada uma das etapas do processo proposto.

Figura 4.1 Processo geral para gerar uma caricatura facial.



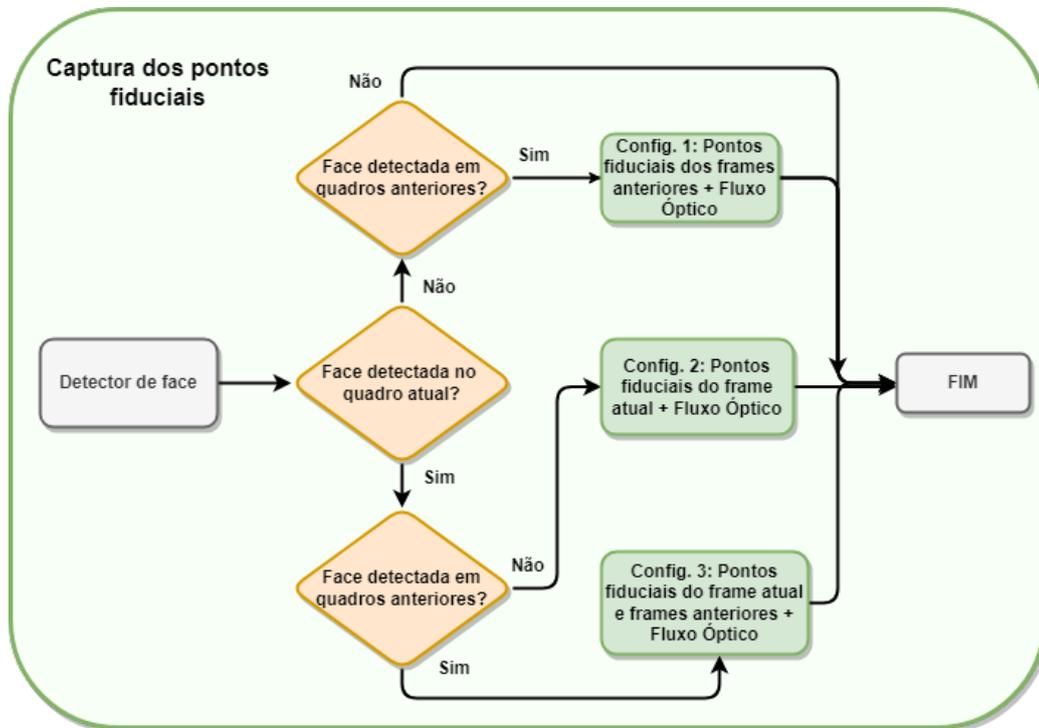
Fonte: Autoria própria

## 4.2 Detecção dos Pontos Fiduciais no Vídeo

Para o processo específico da detecção dos pontos fiduciais no vídeo, foi utilizada uma abordagem adaptativa, conforme o diagrama contido na Figura 4.2. Para obter os pontos fiduciais, três estratégias foram adotadas. A primeira obtém informações de um detector de pontos fiduciais. O detector empregado foi o OpenFace (BALTRUSAITIS et al., 2018), o qual retorna 68 pontos distribuídos por toda a face, conforme a ilustração contida na Figura 4.3.

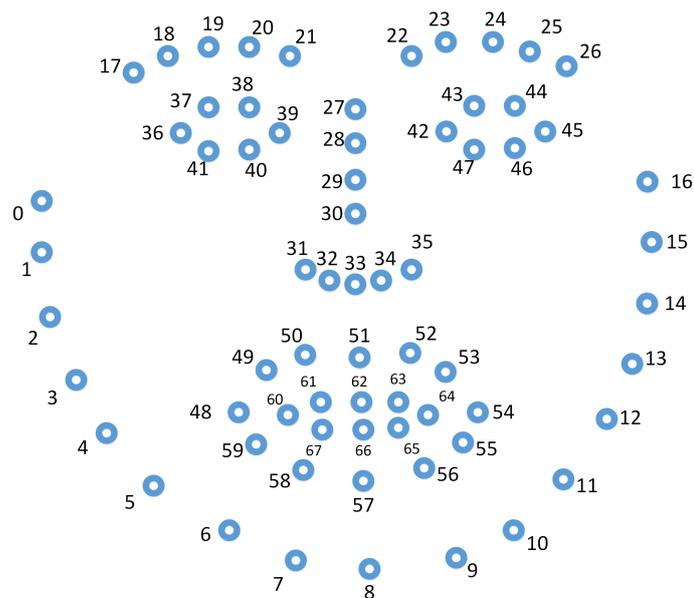
A segunda estratégia é dada pelo cálculo da mediana dos pontos fiduciais detectados em uma sequência de quadros. Finalmente, a terceira estratégia consiste na extração do fluxo óptico, conforme descrita por Ranftl, Alonso-Fernandez e Karlsson (2015).

Figura 4.2 Fluxograma do processo de captura dos pontos fiduciais.



Fonte: Autoria Própria

Figura 4.3 Distribuição dos pontos fiduciais ao redor da face.



Fonte: Baltrusaitis et al. (2018)

Três possibilidades de configuração foram projetadas para auxiliar durante o processo de detecção dos pontos fiduciais no vídeo. Para a definição dos pesos utilizados para cada configuração, tais pesos foram variados sistematicamente no intervalo [0.0 e 1.0], em passos de 0.1 e, em seguida, foi computado o erro de detecção a partir de um pequeno conjunto de dados de calibração contendo coordenadas de pontos fiduciais faciais anotadas. É importante observar que, a soma dos pesos apresentados nas Equações 4.1, 4.2 e 4.3 tem que totalizar sempre 100%. O conjunto de dados utilizado para a calibração é um subconjunto de 20 vídeos do banco de dados 300vw (UŘIČÁŘ et al., 2016).

Para a primeira configuração (processo rotulado como Config. 1 no fluxograma da Figura 4.2), a localização do ponto fiducial ajustado deve ser ponderada como  $w_1 = 0,7$  da mediana das localizações dos pontos dos quadros anteriores próximos e peso de  $0,3(1 - w_1)$  das coordenadas representando a localização prevista por um algoritmo de fluxo óptico. Na Equação 4.1, formaliza-se esta ideia, em que  $f(x, y)$  representa uma mediana dos pontos fiduciais dos quadros próximos e  $P_{flw}(x, y)$  representa o ponto retornado pelo algoritmo de fluxo óptico.

$$P_1(x', y') = w_1 f(x, y) + ((1 - w_1) P_{flw}(x, y)). \quad (4.1)$$

Para a segunda configuração (processo denominado Config. 2 no fluxograma da Figura 4.2), um esquema de ponderação semelhante foi adotado, mas desta vez as coordenadas foram calculadas com pesos diferentes:  $w_1 = 0,9$  das coordenadas do detector de pontos fiduciais e peso de  $0,1(1 - w_1)$  das coordenadas retornadas pelo fluxo óptico. Isso é formalizado na Equação 4.2, em que  $P_d(x, y)$  representa o ponto fiducial retornado pela abordagem OpenFace e  $P_{flw}(x, y)$  representa o ponto previsto pelo algoritmo de fluxo óptico.

$$P_2(x', y') = w_1 P_d(x, y) + ((1 - w_1) P_{flw}(x, y)). \quad (4.2)$$

Para a última configuração (processo denominado Config. 3 no fluxograma da Figura 4.2), é especificado genericamente na Equação 4.3, na qual  $P_d(x, y)$  representa o ponto fiducial retornado pela abordagem OpenFace,  $f(x, y)$  representa a mediana dos pontos fiduciais de quadros adjacentes e  $P_{flw}(x, y)$  representa o ponto previsto pelo algoritmo de fluxo óptico, conforme mencionado anteriormente, os pesos  $w_1$ ,  $w_2$  e  $w_3$  seguem uma restrição matemática em que a soma dos pesos tem como resultado igual a 1 (100%). Após os testes realizados,

obteve-se os melhores resultados para os pesos da Equação 4.3:  $w_1 = 0,7$  para as coordenadas do detector,  $w_2 = 0,2$  para as informações medianas dos quadros adjacentes próximos e  $w_3 = 0,1$  para as coordenadas produzidas pelo algoritmo de fluxo óptico, mais detalhes de como foi realizado o experimento pode ser encontrada na Seção 5.2. A Equação 4.4 apresenta a utilização dos pesos encontrados a partir da Equação genérica.

$$P_3(x', y') = w_1 P_d(x, y) + w_2 f(x, y) + w_3 P_{flw}(x, y). \quad (4.3)$$

$$P_3(x', y') = 0,7 P_d(x, y) + 0,2 f(x, y) + 0,1 P_{flw}(x, y). \quad (4.4)$$

Na primeira etapa do fluxograma da Figura 4.2, é recebido o quadro do vídeo a ser analisado. Após esta etapa, é verificada a possibilidade de se detectar a face no quadro atual, sendo também verificada a existência de quadros anteriores contendo faces. A detecção facial da biblioteca Dlib (KING, 2009) para esta tarefa.

A lógica da abordagem proposta pode ser descrita como segue. Se uma face for detectada no quadro atual e houver quadros com faces previamente detectadas, a Config. 3 é usada. Como alternativa, se uma face for detectada no quadro atual, mas não houver quadros anteriores próximos com faces previamente detectadas, a Config. 2 é usada. Se uma face não for detectada no quadro atual, mas houver quadros próximos anteriores com faces detectadas, a Config. 1 é usada. Presume-se que devem haver rostos em todo o vídeo, mas, eventualmente, o detector de face pode falhar, o que justifica a necessidade da Config. 1. Finalmente, se uma face não for detectada no quadro atual, não houver quadros anteriores próximos com faces detectadas, então não há mais nada a fazer e o algoritmo passa para o próximo quadro. Este processo é realizado até que sejam percorridos todos os quadros do vídeo.

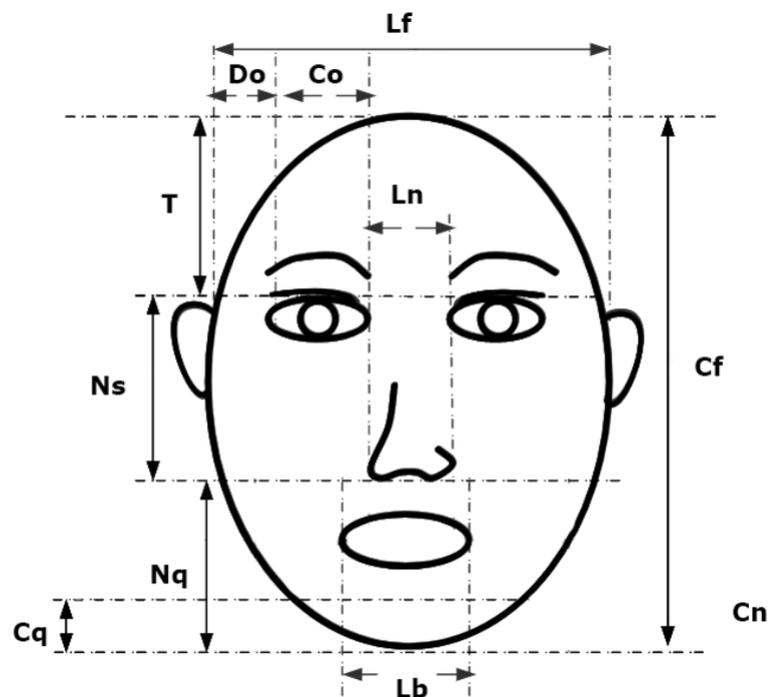
### 4.3 Regras de Definição das Características Relevantes

Para que seja possível gerar deformações faciais, faz-se necessária uma análise das características faciais, com a finalidade de pré-definir os padrões de deformação. As seguintes medições e proporções da face, obtidas de acordo com os estudos de Xiaorong Du et al. (2015), foram utilizadas como referência nesta pesquisa:

- Comprimento e largura da face, (Cf : Lf)(5 : 4);
- Altura do nariz a queixo, altura do nariz sobrancelha e altura da testa, (Nq: Ns: T) (1: 1: 1);
- Comprimento do queixo: Altura do nariz até o queixo, (Cq: Nq)(1: 3);
- Comprimento do olho até a sobrancelha: Comprimento da face, (Co: Cf)(1:10);
- Comprimento do olho: Largura do nariz: Largura da boca, (Co: Ln: Lb)(2: 2: 3);
- Comprimento do olho: Largura da face, (Co: Lf)(2: 9);
- Distância entre os cantos dos olhos: Largura da face, (Do: Lf)(2: 9);
- Largura do nariz: Comprimento do Nariz, (Ln: Cn)(7: 10).

Na Figura 4.4, representam-se, de forma gráfica, estas proporções faciais utilizadas.

Figura 4.4 Proporções Faciais.



Fonte: Adaptado de Xiaorong Du et al. (2015)

A partir dos pontos fiduciais detectados de uma determinada face, é possível extrair um conjunto de características  $T$ . Estas características estão associadas às características faciais,

tais como nariz, queixo, olhos, boca, etc. Com base neste conjunto  $T$  é possível verificar quais dessas características  $i$  se distinguem daquelas correspondentes em uma face média  $X_i$ . Em outras palavras, para exagerar a característica  $i$  associada a um conjunto de características  $T$ , calcula-se a diferença  $D_i$  a partir da Equação 4.5

$$D_i = X_i - T_i. \quad (4.5)$$

A diferença entre as características pode abranger números positivos, para características que precisam ser exageradas e números negativos, para características que precisam ser reduzidas. Em seguida, os módulos das diferenças de cada característica,  $|D|$ , são ordenados de forma decrescente, para que, as três principais características escolhidas sejam deformadas, levando-se em consideração o seu sinal. Nas Figuras 4.5(a), (b) e (c), é possível observar a variação na imagens com a aplicação de diferentes níveis de deformações.

Figura 4.5 Diferentes Níveis de Distorções.



Fonte: Autoria própria

## 4.4 Métodos de Deformação da Face Contida nos Vídeos

Após a definição das características que serão deformadas, faz-se necessário realizar a transformação geométrica dos pontos fiduciais pertencentes a cada característica definida.

A partir dos pontos fiduciais detectados, é possível definir as características faciais com base em um conjunto de pontos numerados, e.g., o conjunto de pontos pertencentes à sequência de 36 à 47 são pontos fiduciais correspondentes aos olhos da face, como pode ser observado na Figura 4.3. Com base nos valores da comparação com uma face média, esses conjuntos de pontos serão expandidos (caso o resultado de comparação com a face média seja positivo) ou contraídos (caso o resultado de comparação com a face média seja negativo).

Entretanto, para que seja possível realizar esta transformação geométrica dos pontos faz-se necessário que os pontos fiduciais da face estejam na posição frontal, o que normalmente não é possível de se obter em uma sequência de quadros em movimento.

Para solucionar este problema, torna-se necessário primeiramente rotacionar os pontos fiduciais detectados para uma pose frontal, para em seguida, realizar-se a expansão ou retração dos pontos referentes às características definidas e por fim, os pontos distorcidos são remapeados para a pose original da face. Esta etapa do processamento é explicada a seguir.

A rotação de um ângulo  $\theta$  em torno dos eixos  $x$ ,  $y$  e  $z$  é expressa pelas Equações 4.6, 4.7 e 4.8, respectivamente.

$$P_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}, \quad (4.6)$$

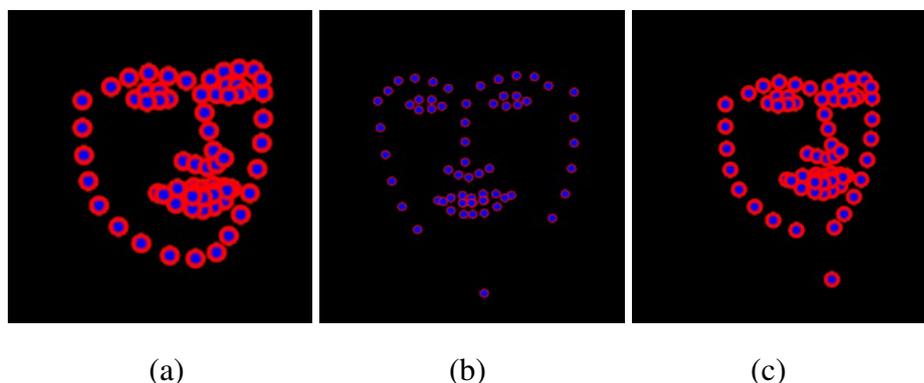
$$P_y(\theta) = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}, \quad (4.7)$$

$$P_z(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (4.8)$$

As Equações 4.6, 4.7 e 4.8 são empregadas para rotacionar os pontos da face para a posição frontal e, após realizar a deformação dos pontos, é utilizada a matriz de rotação inversa para retornar os novos pontos para a rotação original.

Na Figura 4.6, apresenta-se um exemplo da rotação dos pontos fiduciais e da transformação geométrica exagerada de um dos pontos do queixo. Na Figura 4.6(a), é possível observar os pontos da face com certo grau de rotação. Na Figura 4.6(b), se apresentam os pontos faciais rotacionados para frente e um ponto correspondente ao contorno do queixo alterado. Por fim, na Figura 4.6(c), se apresentam os pontos transformados e rotacionados de volta para a orientação original.

Figura 4.6 Exemplo do uso da rotação dos pontos fiduciais para aprimorar a transformação espacial dos pontos

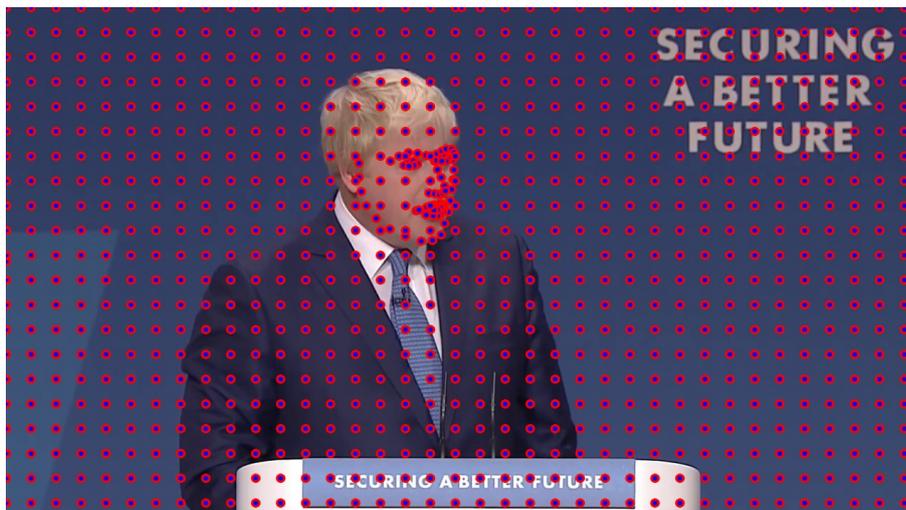


Fonte: Autoria própria

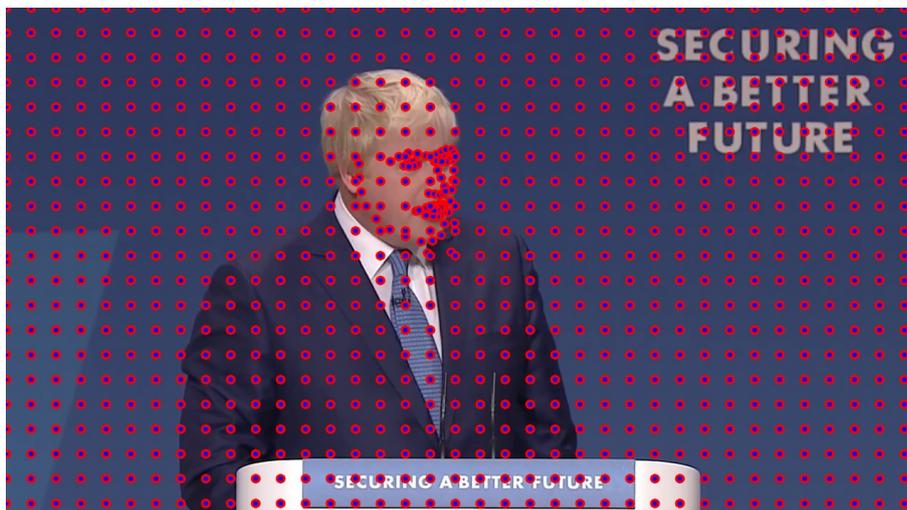
Após obter os novos pontos fiduciais da face a ser deformada, é realizada a deformação espacial dos *pixels* (aparência) da imagem a partir da diferença entre os pontos. Entretanto, após alguns testes realizados, percebeu-se que esta estratégia de deformação acarretava também deformações não desejadas ao redor da face. Para solucionar este problema, foram utilizados “pontos âncoras” por toda a imagem analisada, com o objetivo de concentrar as deformações apenas na região facial escolhida. Na Figura 4.7, apresenta-se um exemplo de um quadro com a utilização dos pontos âncoras para realizar a deformação espacial entre os *pixels*. Na Figura 4.7(a), é apresentado o quadro antes da deformação facial e na Figura 4.7(b) é apresentado o quadro após a deformação facial. É possível observar a constância dos pontos no *background* da imagem e as mudanças das deformações nos pontos próximos da face.

## 4.5 Detalhes de Projeto e Implementação

Para a validação da abordagem proposta apresentada nas seções anteriores, foi necessário desenvolver um sistema para a geração automática de caricaturas no contexto de vídeos, seguindo as etapas e o fluxo de execução apresentado e discutido nas seções anteriores. Nesta seção, são apresentados detalhes de projeto e implementação do software desenvolvido para a validação da abordagem descrita.

Figura 4.7 Exemplo do uso de pontos âncoras para realizar a deformação espacial dos *pixels*.

(a)



(b)

Fonte: Autoria própria

A implementação foi realizada utilizando a linguagem C++, a partir do ambiente de desenvolvimento integrado (*Integrated Development Environment - IDE*) *QT Creator* 5<sup>1</sup>. A linguagem C++ foi escolhida, dentre vários aspectos, por: (i) apresentar um desempenho superior para a tarefa, em relação a outras linguagens; (ii) ser muito utilizada na área de Visão Computacional; (iii) dispor de bibliotecas tais como a *OpenCV*<sup>2</sup>, dotada de facilidades destinadas ao suporte no desenvolvimento do trabalho. O ambiente de desenvolvimento *QT*

<sup>1</sup><https://www.qt.io/product/development-tools>

<sup>2</sup><https://opencv.org/>

*Creator* foi escolhido pela disponibilidade de utilização do projeto tanto em plataformas *Windows* como em plataformas baseadas no Linux. O projeto utilizado neste estudo pode ser encontrado no Github<sup>3</sup>

## 4.6 Metodologia Adotada para a Avaliação com Usuários

O processo de avaliação da abordagem proposta para a caricaturização de faces em vídeos consistiu de um experimento envolvendo participantes humanos. Para a realização do experimento, fez-se necessário seguir as seguintes etapas: (i) Seleção dos vídeos a serem utilizados no experimento; (ii) Definição das questões a serem abordadas; (iii) Definição do perfil do usuário por meio do QDPU (Questionário de Delineamento do Perfil do Usuário) e (iv) Recrutamento Online dos usuários, via e-mail.

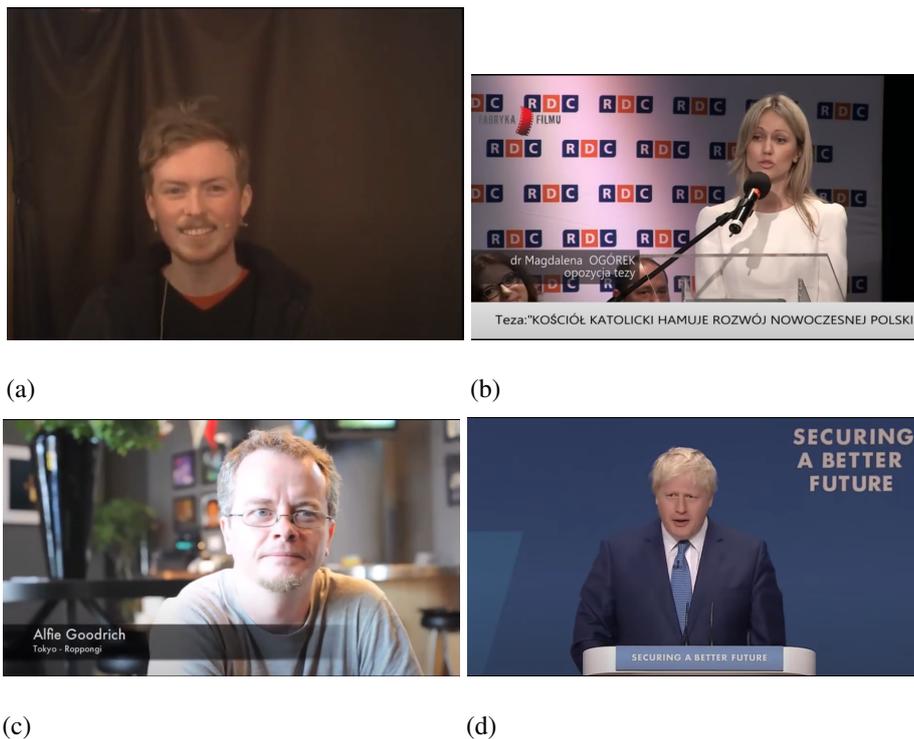
Para a escolha dos vídeos a serem utilizados no experimento, foram separados quatro vídeos da base de dados 300-vw (SAGONAS; TZIMIROPOULOS et al., 2013). Os vídeos foram selecionados por apresentarem diferentes *backgrounds*, presença de movimentações bruscas da cabeça e diferentes níveis de iluminação. Nas Figuras 4.8(a) a (d) são apresentados quadros extraídos dos vídeos selecionados. No Apêndice D, são apresentadas as sequências dos quadros utilizados no experimento e um quadro de vídeo correspondente, apresentando a deformação facial resultante.

Para a avaliação da abordagem proposta, a princípio, foi idealizado realizar a comparação da abordagem com diferentes estudos presentes na literatura. Entretanto, os estudos que apresentavam formas de replicação de resultados (seja via disponibilização de código fonte, seja via arquivo executável) não contemplavam os requisitos necessários para uma comparação justa. Por vezes, a deformação realizada em vídeo não tinha como intenção uma geração de caricaturas, como no estudo apresentado por Feng e Ge (2012). Outras vezes, as deformações realizadas eram apenas para imagens estáticas, como no estudo de Fujiwara et al. (2001). Por fim, não eram disponibilizados meios para separar as imagens deformadas das imagens estilizadas como, por exemplo, o estudo de Shi, Deb e Jain (2019).

---

<sup>3</sup><https://github.com/ViniFernandes/Uma-Abordagem-para-distor-es-de-faces-em-v-deos-digitais>

Figura 4.8 Quadros ilustrativos dos vídeos utilizados no experimento



Fonte: Autoria Própria

Por estas questões, optou-se por realizar uma avaliação interna da abordagem proposta. Para esta avaliação, foram gerados diferentes vídeos alterando-se o nível de deformação das características faciais com a intenção de definir o nível ideal de deformação dos vídeos analisados. Além disso, também foram gerados vídeos com e sem a utilização do algoritmo adaptativo apresentado na Seção 4.1, para assim, verificar o impacto final do algoritmo adaptativo nos vídeos gerados.

Para a definição das questões propostas aos participantes do experimento, foram formuladas três afirmações principais que objetivaram sintetizar um conjunto de características ideais que os vídeos resultantes da caricaturização automática proposta deveriam apresentar, a saber:

- Destaque das principais características faciais do alvo observado;
- Preservação da coerência das deformações;
- Preservação da coerência temporal;
- Qualidade visual do vídeo.

Após a apresentação do vídeo gerado por meio da abordagem proposta neste estudo, foram apresentadas ao usuário as afirmações presentes no Quadro 4.1. A cada questão associou-se uma escala de Likert de 5 pontos contendo as seguintes âncoras de semântica diferencial: 1 - Discordo totalmente; 2 - Discordo parcialmente; 3 - Não concordo nem discordo; 4 - Concordo Parcialmente; 5 - Concordo Totalmente.

Quadro 4.1 Afirmitivas apresentadas ao usuário para a avaliação dos vídeos gerados

---

Questão 1	Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).
Questão 2	O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, Cintilações).
Questão 3	O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

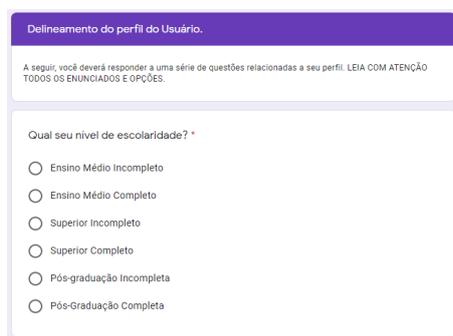
---

Fonte: Autoria própria.

Para o questionário de delineamento do perfil do usuário, foram formuladas questões acerca das características individuais do usuário e também sobre o seu nível de conhecimento sobre caricaturas. A última questão do questionário (“Como você julga sua facilidade de identificar uma caricatura”) foi elaborada com propósito eliminatório, ou seja, os usuários que por ventura assinalaram a alternativa "Muito Difícil" foram descartados da etapa de avaliação do modelo proposto. Uma parte do questionário pode ser observada na Figura 4.9. O questionário completo de delineamento do usuário pode ser visto com mais detalhes no Apêndice A.

Para a avaliação da abordagem proposta, foram gerados quatro vídeos diferentemente processados, conforme a descrição a seguir: Processamento 1 - Sem a utilização do algoritmo adaptativo de detecção dos pontos fiduciais e nível 3 de deformação da imagem; Processamento 2 - Com a utilização do algoritmo adaptativo e nível 3 de deformação; Processamento 3 - Com a utilização do algoritmo adaptativo e nível 2 de deformação; Processamento 4 - Com a utilização do algoritmo adaptativo e nível 1 de deformação.

Figura 4.9 Exemplo de parte do questionário de delineamento de perfil do usuário



Delineamento do perfil do Usuário.

A seguir, você deverá responder a uma série de questões relacionadas a seu perfil. LEIA COM ATENÇÃO TODOS OS ENUNCIADOS E OPÇÕES.

Qual seu nível de escolaridade? \*

- Ensino Médio Incompleto
- Ensino Médio Completo
- Superior Incompleto
- Superior Completo
- Pós-graduação Incompleta
- Pós-Graduação Completa

Fonte: Autoria própria

O nível de deformação consiste na distância na qual os novos pontos das características serão posicionados. No nível 1, o deslocamento dos novos pontos ocorreu de acordo com o resultado da proporção calculada, conforme apresentado na Seção 4.2. No nível 2, os novos pontos foram deslocados o dobro do nível da proporção calculada. Finalmente, no nível 3 o deslocamento dos novos pontos foi o triplo do nível da proporção calculada. Nas Figuras 4.10 (a), (b) e (c), são apresentados exemplos de quadros com os diferentes níveis de deformação utilizados.

Figura 4.10 Exemplos de quadros com os diferentes níveis de deformação utilizados. (a) Nível de deformação 1; (b) Nível de deformação 2 e (c) Nível de deformação 3



(a)

(b)

(c)

Fonte: Autoria própria

Conforme mencionado anteriormente, foram utilizados quatro vídeos para o experimento e, para cada vídeo, foram gerados quatro novos, alterando-se o tipo de processamento entre eles, ou seja, para cada resposta dos usuários foram obtidas 16 respostas para cada questão aplicada. A escolha dos vídeos se deu buscando cobrir diferentes cenários a serem observados,

como por exemplo, a ausência ou não de um *background* mais homogêneo e a presença ou não de movimentos rápidos da região facial. Foram definidos apenas quatro vídeos para que o processo de resposta dos voluntários não se transformasse em um processo cansativo.

Devido às circunstâncias impostas pela pandemia da COVID-19, o recrutamento dos participantes ocorreu de forma remota, assim como a sondagem de sua opinião, conduzida a partir dos questionários supradescritos. O recrutamento foi realizado por meio de um e-mail (ver Apêndice C).

## 4.7 Considerações Finais

Neste capítulo, foram descritos os diferentes aspectos que caracterizam a abordagem proposta neste estudo. Foram detalhados os procedimentos para a detecção de pontos fiduciais e foram definidos os critérios de escolha das características a serem deformadas. Também foram explicitadas as estratégias adotadas para realizar tanto a deformação facial, a partir dos pontos fiduciais encontrados, quanto à deformação espacial dos *pixels* nos quadros do vídeo-alvo. No próximo capítulo, são apresentados os experimentos realizados neste estudo e seus respectivos resultados.

# Capítulo 5

## Resultados e Discussões

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir dos experimentos realizados neste estudo. Na Seção 5.1, são apresentados e discutidos os resultados provenientes da avaliação dos algoritmos de detecção de pontos fiduciais. Na Seção 5.2, são apresentados e discutidos os resultados da comparação do algoritmo adaptativo com as demais abordagens concorrentes selecionadas.. Por fim, na Seção 5.3, são apresentados e discutidos os resultados provenientes da avaliação da abordagem proposta a partir da sondagem da opinião de usuários.

### 5.1 Avaliação dos Algoritmos de Detecção de Pontos Fiduciais

Para a avaliação dos algoritmos de detecção de pontos fiduciais apresentados no capítulo anterior fez-se necessária a pré-definição de uma série de etapas que estão detalhadas nesta seção.

Primeiramente, para que fosse possível realizar uma avaliação justa dos algoritmos de detecção de pontos fiduciais pré-definidos em comparação com o algoritmo proposto, foi necessário definir um banco de imagens único, com a finalidade de realizar o treinamento e a avaliação de cada um dos algoritmos. A pré-definição ocorreu conforme discutido na Subseção 5.1.1.

Após a definição do banco de imagens a ser utilizado, fez-se necessário estabelecer a métrica a ser adotada no processo de avaliação da acurácia dos algoritmos analisados. Esta

atividade encontra-se descrita na Subseção 5.1.2.

Posteriormente, iniciou-se a etapa de adaptação dos códigos selecionados, objetivando-se a manipulação de todos os dados do banco de imagens pré-definido por todos os algoritmos considerados, a fim de evitar erros no processamento dos dados. Esta etapa foi executada tanto para os algoritmos de treinamento, quanto para aqueles de detecção de pontos fiduciais. Por fim, foi realizada a validação experimental, detalhada na Subseção 5.1.3.

### 5.1.1 Banco de Imagens

Nesta etapa da pesquisa, foram pesquisados diversos bancos de imagens faciais que contivessem os pontos fiduciais demarcados, com diversas características específicas, e.g., o número de pontos fiduciais destacados, a apresentação de um fundo de imagem homogêneo ou heterogêneo, a quantidade de imagens disponíveis, a disponibilidade de imagens coloridas ou com níveis de cinza. Em função das características analisadas, verificou-se a adequação de alguns bancos de imagens ao propósito desejado. Um deles foi o *BioID Face Database* (JESORSKY; KIRCHBERG; FRISCHHOLZ, 2001), um banco que contém 1521 imagens monocromáticas, com resoluções de  $384 \times 286$  pixels, obtidas, em um ambiente de escritório comum e contemplando 20 pontos fiduciais distribuídos pela face. O *Milborrow/University of Cape Town (MUCT) Database* (MILBORROW; MORTEL; NICOLLS, 2010) contém 3755 imagens coloridas de faces frontais com 76 pontos fiduciais demarcados, resolução de  $640 \times 480$  pixels e variações de iluminação, idade e etnia.

O banco de dados *Pozna University of Technology (PUT) Face Database* (KASINSKI; FLOREK; SCHMIDT, 2008) contém 9971 imagens coloridas, com resoluções de  $2048 \times 1536$  pixels, fotografadas em um fundo único e com poses naturais, contando com 30 pontos, fiduciais demarcados em todas as imagens, além de conter um subconjunto de 2193 imagens de faces frontais com 194 pontos fiduciais demarcados.

O *IMM Face Database* (NORDSTRØM et al., 2004) é um banco de imagens composto por 240 imagens coloridas de 40 indivíduos diferentes, distribuídos entre 7 mulheres e 33 homens, com imagens de  $640 \times 480$  pixels de resolução, fotografadas em um fundo padrão para todas as imagens e cinco posições diferentes para cada indivíduo fotografado, e contém 58 pontos fiduciais demarcados na face.

O banco de dados *Helen Dataset* (LE et al., 2012) contém 2330 imagens coloridas, com

resoluções superiores a 500 pixels de largura, criado a partir de imagens da plataforma *Flickr*<sup>1</sup> e, portanto, dotadas de variações de ambientação, pose, iluminação e expressões faciais, todas contendo anotações de 194 pontos fiduciais.

O banco de dados 300-W (SAGONAS; ANTONAKOS et al., 2016) contém 600 imagens coloridas, contendo 68 pontos fiduciais demarcados na face, resoluções superiores a 800 *pixels* de altura e fundo homogêneo. Na verdade, trata-se de um banco de imagens que integra o AFW Dataset, o IBUG, LFPW e Helen Dataset.

Ao final do processo de investigação dos bancos de imagens, selecionou-se o 300-W, pelas razões listadas a seguir: (i) apresenta uma seleção de imagens mais propícias para a avaliação, com grupos de imagens separadas por fundo controlado e fundo ao ar livre; (ii) contém imagens com resoluções compatíveis, no geral, às obtidas por meio de webcams ou câmeras fotográficas (iii) reúne imagens de outros bancos de imagens utilizados nos estudos relacionados; (iv) tem sido empregado como base para a avaliação de algoritmos de localização de pontos fiduciais na competição *300 Faces In-the-Wild Challenge*<sup>2</sup>.

### 5.1.2 Definição da Métrica

Para a definição da métrica a ser utilizada, fez-se necessária a condução de uma pesquisa, a partir da qual, se constatou a adequação e pertinência do emprego da raiz do erro quadrático médio (*Root-Mean-Square Error* - RMSE) ao processo de avaliação desejado. Conforme explicitada na Equação 5.1, RMSE é uma métrica de precisão, mais sensível a erros. RMSE sempre retorna valores positivos ou nulos, sendo este último um indicativo de detecção perfeita. Valores altos de RMSE indicam menor precisão do algoritmo de detecção de pontos fiduciais (maiores diferenças entre a previsão  $y'_t$  e observação  $y_t$ , divididos pela quantidade de elementos observados  $n$ ). RMSE tem sido utilizado em diversos estudos de avaliação de algoritmos de detecção de pontos fiduciais como, por exemplo, os estudos apresentados nos artigos de Baltru, Robinson, Morency et al. (2016), Zhang et al. (2014) e Kazemi e Sullivan (2014).

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y'_t - y_t)^2}{n}}. \quad (5.1)$$

<sup>1</sup><https://www.flickr.com/>

<sup>2</sup><https://ibug.doc.ic.ac.uk/resources/300-VW/>

### 5.1.3 Validação Experimental

Para garantir que a comparação realizada entre os algoritmos seja justa, fez-se necessário, na etapa de treinamento, realizar um nivelamento de todos os modelos treinados, com a finalidade de que todos os modelos resultantes apresentassem uma taxa de erro média estatisticamente igual. Foi necessário, portanto, realizar alterações na forma de leitura dos dados nos códigos originais de cada algoritmo analisado para que, assim, fosse possível utilizar a mesma base no conjunto de treinamento para todos os algoritmos.

Para tanto, foi realizada uma validação cruzada de cinco dobras, que consiste na divisão do banco de imagens em cinco novos grupos, distintos entre si, segundo a qual foi extraído de cada grupo um subgrupo composto por setenta e cinco por cento das imagens, com a finalidade de treinamento e as restantes vinte e cinco por cento para a validação. Para cada dobra foram alterados os parâmetros de treinamento de cada algoritmo e avaliado o modelo resultante, com base em seu próprio banco de treinamento para, assim, garantir que todos os modelos treinados apresentassem estatisticamente níveis de acurácia iguais. De posse dos novos modelos re-treinados para cada algoritmo, foram analisados seus níveis de erro e escolhido um modelo resultante de cada técnica que apresentasse o menor nível de variância possível entre eles.

Após a definição dos modelos de treinamento, fez-se necessário, por fim, avaliar os algoritmos de detecção de pontos fiduciais escolhidos. Para esta avaliação, foi novamente conduzida uma validação cruzada de cinco dobras, feita empregando-se no processo de avaliação as imagens selecionadas que não foram utilizadas na fase de treinamento.

A partir da condução das etapas supradescritas, foi possível conferir imparcialidade ao processo de avaliação dos algoritmos, visto que tal praxe confere aos algoritmos analisados possuírem chances iguais de atingirem seus melhores desempenhos na etapa de detecção dos pontos fiduciais, uma vez que todos os algoritmos obtiveram a mesma entrada de imagens para a etapa de treinamento.

### 5.1.4 Resultados Obtidos

Baseando-se na métrica definida para a validação dos algoritmos, foi obtida, após várias iterações para a normalização dos dados, uma taxa de erro médio estatisticamente igual para

todos os modelos treinados. Na Tabela 5.1, são apresentados os dados obtidos de cada modelo treinado e seu valor respectivo coeficiente de variância, em que é possível observar que o modelo 03 re-treinado obteve menor coeficiente de variância, quando comparado com os demais modelos, tendo sido escolhido para ser utilizado no processo de validação.

Tabela 5.1 Erro médio na fase de de treinamento para os 5 modelos analisados de cada algoritmo

Implementação	Erro	no	Erro	no	Erro	no	Erro	no	Erro	no
	Modelo 01	Modelo 02	Modelo 03	Modelo 04	Modelo 05					
ASM	3,12	3,05	<b>3,19</b>	2,90	3,09					
AAM	2,89	2,95	<b>2,97</b>	2,80	2,86					
STASM	3,15	3,21	<b>3,11</b>	3,14	3,15					
CLM-Framework	3,01	3,19	<b>3,04</b>	3,20	3,18					
OpenFace	2,95	2,89	<b>3,03</b>	2,90	2,99					
DLib	3,07	3,20	<b>3,10</b>	3,21	3,11					
Clandmark	2,87	2,94	<b>2,99</b>	3,18	3,20					
Face Analysis	3,19	3,20	<b>2,99</b>	3,03	3,11					
Coefficiente de Variação	0,01	0,01	<b>0,01</b>	0,02	0,01					

Fonte: Autoria própria

Na Tabela 5.2, são apresentados os valores de média, desvio padrão e coeficiente de variação para cada modelo treinado. Observa-se que todos os algoritmos re-treinados obtiveram valores de média e desvio padrão relativamente similares.

Utilizando-se os modelos re-treinados definidos, foram avaliados os algoritmos de detecção de pontos fiduciais, adotando-se a estratégia da validação cruzada com  $K$  igual 5. Foram obtidos os valores contidos na Tabela 5.3, na qual é possível observar que o algoritmo *Face Analysis SDK* obteve o menor valor de taxa de erro dentre todos os algoritmos analisados.

Na Figura 5.1, é apresentado um *boxplot* no qual se observa que o algoritmo *Face Analysis SDK* obteve melhor desempenho, seguido pelos algoritmos *OpenFace* e *CLM-Framework*.

Tabela 5.2 Média, desvio padrão e coeficiente de variação na fase de treinamento

Implementação	Média	Desvio padrão	Coeficiente de Variação
ASM	3,07	0,10	0,01
AAM	2,89	0,06	0,01
STASM	3,15	0,03	0,01
CLM-Framework	3,12	0,09	0,01
OpenFace	2,95	0,06	0,01
DLib	3,14	0,06	0,01
Clandmark	3,03	0,14	0,02
Face Analysis SDK	3,10	0,09	0,01

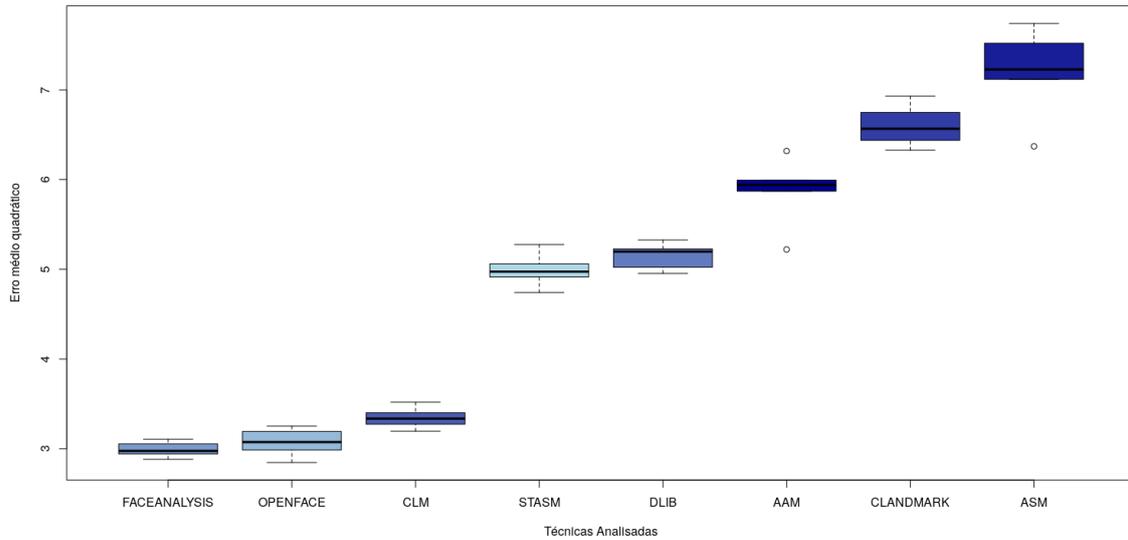
Fonte: Autoria própria

Tabela 5.3 Erro associado à detecção dos pontos fiduciais na fase de validação

Implementação	Erro no				
	Grupo 01	Grupo 02	Grupo 03	Grupo 04	Grupo 05
ASM	7,74	7,52	7,11	6,37	7,22
AAM	5,22	5,87	5,94	6,31	5,99
STASM	4,97	4,74	4,91	5,27	5,06
CLM-Framework	3,40	3,33	3,51	3,27	3,19
OpenFace	3,19	3,25	3,07	2,98	2,84
DLib	4,95	5,02	5,32	5,22	5,19
Clandmark	6,56	6,93	6,32	6,74	6,43
Face Analysis SDK	2,94	3,05	3,10	2,88	2,97

Fonte: Autoria própria

Figura 5.1 Boxplot dos erros na fase de validação



Fonte: Autoria própria

Entretanto, para que se pudesse garantir a veracidade de tal afirmação com maior confiança, foi necessário realizar uma análise estatística dos dados obtidos. Primeiramente, verificou-se a normalidade dos dados. Utilizando-se *p-value* de 0,01, foi possível constatar que os dados obtidos seguiam uma distribuição normal. Assim sendo, foi realizado o teste *ANOVA* para verificar a possível divergência estatística entre os algoritmos analisados. Foi obtido um *p-value* menor que  $2^{-16}$ , o que indica a existência de diferenças entre os modelos analisados considerando-se um valor de *Alpha* menor que 0,001.

A fim de verificar quais modelos apresentavam diferenças estatísticas, utilizou-se o teste de *Tukey*, considerando-se um grau de confiança de 99,9%, para o qual se obtiveram *p-value* contidos na Tabela 5.4. Na Tabela 5.5, apresentam-se os grupos de similaridade advindos do teste de *Tukey*. É possível observar que os algoritmos Face Analysis, CLM e OpenFace apresentam, estatisticamente, taxas de erro equivalentes. Com base nesses resultados, optou-se por utilizar o Openface, por ser um código mais recente, assim como por apresentar suporte à detecção de pontos fiduciais em imagens tridimensionais.

Tabela 5.4 Comparação entre as técnicas utilizando o teste de Tukey

Contraste	Diferença	Limite Inferior	Limite Superior	P-Value	Dif. Significativa para Alpha = 0,001
ASM Vs AAM	1,32	0,48	2,16	0,00	SIM
CLM Vs AAM	-2,52	-3,36	-1,68	0,00	SIM
DLIB Vs AAM	-0,72	-1,56	0,11	0,00	SIM
Face Analysis SDK Vs AAM	-2,87	-3,71	-2,03	0,00	SIM
OpenFace Vs AAM	-2,79	-3,63	-1,96	0,00	SIM
STASM Vs AAM	-0,87	-1,71	-0,03	0,00	SIM
CLM Vs ASM	-3,85	-4,68	-3,01	0,00	SIM
DLIB Vs ASM	-2,04	-2,88	-1,21	0,00	SIM
Face Analysis SDK Vs ASM	-4,20	-5,04	-3,36	0,00	SIM
OpenFace Vs ASM	-4,12	-4,96	-3,28	0,00	SIM
STASM Vs ASM	-2,20	-3,04	-1,36	0,00	SIM
CLM Vs CLANDMARK	-3,25	-4,09	-2,41	0,00	SIM
DLIB Vs CLANDMARK	-1,45	-2,29	-0,61	0,00	SIM
Face Analysis SDK Vs CLANDMARK	-3,61	-4,44	-2,77	0,00	SIM
OpenFace Vs CLANDMARK	-3,53	-4,37	-2,69	0,00	SIM
STASM Vs CLANDMARK	-1,60	-2,44	-0,77	0,00	SIM
DLIB Vs CLM	1,80	0,96	2,63	0,00	SIM
STASM Vs CLM	1,64	0,80	2,48	0,00	SIM
Face Analysis SDK Vs DLIB	-2,15	-2,99	-1,31	0,00	SIM
OpenFace Vs DLIB	-2,07	-2,91	-1,23	0,00	SIM
STASM Vs Face Analysis SDK	2,00	1,16	2,84	0,00	SIM
STASM Vs OpenFace	1,92	1,08	2,76	0,00	SIM
CLANDMARK Vs ASM	-0,59	-1,43	0,24	0,03	NÃO
CLANDMARK Vs AAM	0,73	-0,10	1,57	0,01	NÃO
Face Analysis SDK Vs CLM	-0,35	-1,19	0,48	0,49	NÃO
OpenFace Vs CLM	-0,27	-1,11	0,56	0,76	NÃO
STASM Vs DLIB	-0,15	-0,99	0,68	0,98	NÃO
OpenFace Vs Face Analysis SDK	0,07	-0,76	0,91	0,99	NÃO

Fonte: Autoria própria

Tabela 5.5 Grupos de similaridade obtidos por meio do teste de Tukey-Kramer

Técnicas	Média Estimada	Desvio Padrão	Grupos
Face Analysis	2,99	0,08	A
CLM	3,34	0,12	A
OpenFace	3,07	0,16	A
STASM	4,99	0,19	B
DLib	5,14	0,15	B
ASM	7,19	0,52	C
AAM	5,86	0,40	C
ClandMark	6,60	0,24	C

Fonte: Autoria própria

## 5.2 Avaliação do Algoritmo Adaptativo

A fim de validar a abordagem proposta do modelo de detecção de pontos fiduciais a partir do algoritmo adaptativo implementado nesta pesquisa, uma série de avaliações experimentais internas foi conduzida, considerando-se as diferentes abordagens pré-definidas. Tais experimentos foram essenciais para a consolidação da abordagem proposta, visto tratar-se de uma combinação das abordagens analisadas. Adicionalmente, foi possível destacar a eficácia das decisões adaptativas do algoritmo proposto quando comparado à eficácia de cada algoritmo, individualmente.

Nesta primeira etapa da avaliação, denominada Experimento 1, foram considerados os seguintes algoritmos: (i) MED (Uso da Mediana); (ii) MFO (Uso da Mediana do Fluxo Óptico); (iii) PFO (Predição por Fluxo Óptico); (iv) PDM (Predição por Fluxo Óptico combinado com o uso da Mediana do detector de ponto fiducial); e (v) AAD (Algoritmo Adaptativo Proposto).

A abordagem MED (Uso da Mediana) calcula a mediana dos pontos fiduciais detectados anteriormente, empregando-a como um auxílio à detecção de novos pontos fiduciais. A abordagem MFO (Uso da Mediana do Fluxo Óptico) emprega a mediana do resultado do fluxo óptico em uma série de quadros anteriores. A abordagem PFO (Predição por Fluxo Óptico) emprega as informações fornecidas pelo algoritmo de fluxo óptico apenas no quadro

atual. A abordagem PDM (Predição por Fluxo Óptico combinada com o uso da Mediana do detector de ponto fiducial) determina a média dos valores retornados pelo algoritmo de fluxo óptico e os valores dos pontos fiduciais previamente detectados. Finalmente, a abordagem AAD (Algoritmo Adaptativo Proposto) é a abordagem da Figura 4.2, conforme descrito anteriormente na Seção 4.2.

Após realizar a avaliação interna e obter a melhor configuração de parâmetros para o algoritmo proposto, foi realizada uma avaliação experimental externa com abordagens do estado da arte. O objetivo era validar o algoritmo proposto. Esta segunda etapa de avaliação foi denominada Experimento 2 e considerou as seguintes abordagens, além daquela proposta: DLib (DLB); (ii) OpenFace (OPF); (iii) Clandmark (CLM); e (iv) Alinhamento Facial (FAL). Elas são descritas com mais detalhes na Subseção 5.2.3.

### 5.2.1 Base de Dados

Com o objetivo da avaliação das abordagens selecionadas e propostas, uma base de dados de vídeos foi escolhida para analisar os erros na detecção dos pontos fiduciais. A presença de pontos fiduciais anotados das faces encontradas nos vídeos é essencial para permitir a avaliação dos erros de predição. O banco de dados 300-vw apresentado no estudo de Sagonas, Tzimiropoulos et al. (2013) atendeu a este requisito. Além disso, este banco de dados é amplamente utilizado na literatura relacionada, como nas pesquisas de Baltrusaitis et al. (2018), Shizhan et al. (2015), Dong et al. (2018) e Meilu et al. (2019).

Como o nome sugere, 300-vw tem 300 vídeos com duração aproximada de um minuto (entre 25 e 30 quadros por segundo), com todas as faces dos quadros anotadas com 68 pontos fiduciais. Cada vídeo é sobre um único indivíduo e, portanto, há apenas um rosto por quadro de vídeo. Para os experimentos de avaliação, foram escolhidos 73 vídeos que apresentavam diferentes variações de movimento, brilho e foco no rosto analisado. A Tabela 5.6 contém uma lista dos vídeos escolhidos.

### 5.2.2 Métrica de Erro

Assim como na avaliação anterior, a métrica de erro utilizada foi RMSE. Ver a Subseção 5.1.2 para mais detalhes.

Tabela 5.6 Lista de vídeos da base 300-vw selecionados para o processo de avaliação

Lista de vídeos														
001	002	004	009	013	019	022	028	029	033	046	047	112	114	119
120	123	124	125	126	138	143	144	150	158	160	203	205	208	211
212	213	214	218	224	401	402	403	404	405	406	407	408	410	411
412	505	506	507	508	511	515	516	518	522	524	525	526	528	530
531	533	537	538	540	541	546	548	550	551	553	557	558		

Fonte: Autoria própria

### 5.2.3 Abordagens Avaliadas

Nesta subseção, são descritas as abordagens consideradas nas avaliações experimentais comparativas, que se concentram na precisão da detecção de pontos fiduciais de faces em vídeo. Todas as abordagens escolhidas, nomeadamente CLandmark<sup>3</sup>, OpenFace<sup>4</sup>, DLib<sup>5</sup> e Face Alignment<sup>6</sup>, têm implementações Open Source C++ fornecidas pelos respectivos autores, o que permitiu uma comparação justa.

Clandmark (UŘIČÁŘ et al., 2016) fundamenta-se em um modelo de distribuição de pontos (PDM). Para modelar a aparência, a face inteira não é usada, mas apenas as localizações ao redor dos pontos fiduciais de interesse (semelhante à representação usada na técnica do Active Shape Model, (COOTES; TAYLOR et al., 1995)). Isso permite que a abordagem seja mais generalista, pois não envolve toda a complexidade de aprender a aparência de todo o rosto. O modelo utilizado pelo Clandmark pode ser descrito pelos parâmetros  $p = [s; R; q; t]$ , que podem ser variados para obter várias instâncias do modelo. Vale ressaltar que esses parâmetros podem ser variados quando o objetivo é obter diferentes instâncias do modelo. O PDM utilizado é formalizado na Equação 5.2.

$$x_i = sR(\bar{x}_i + \Phi_{iq}) + t, \quad (5.2)$$

<sup>3</sup><https://github.com/uricamic/clangmark>

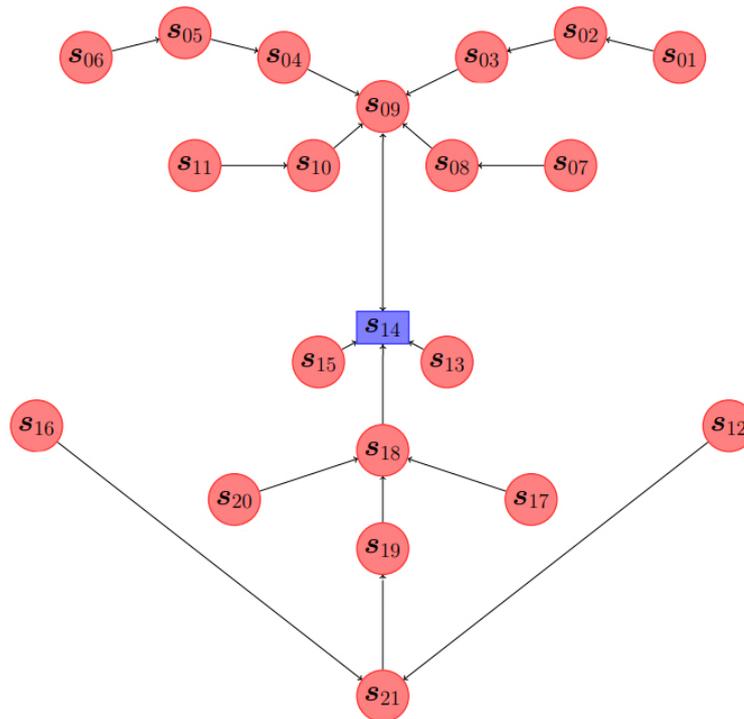
<sup>4</sup><https://github.com/TadasBaltrusaitis/OpenFace>

<sup>5</sup><http://dlib.net/>

<sup>6</sup><https://github.com/ladrianb/face-alignment>

em em que o parâmetro  $s$  é definido como o fator de escala,  $R$  é a rotação do objeto,  $t$  é a translação 2D e  $q$  é um vetor com a variação da forma. Na Figura 5.2 é exemplificado a utilização da Equação 5.2 para a detecção dos pontos da face. Cada ponto apresentado na Figura 5.2 representa uma localização específica dos pontos da face.

Figura 5.2 Modelo do PDM de 21 pontos fiduciais detectados de uma face frontal



Fonte: Adaptado de Uříčář et al. (2016)

OpenFace (BALTRUSAITIS et al., 2018) depende de Campos Neurais Locais Condicionais (CLNF) para a detecção e o rastreamento de pontos fiduciais de face. Os dois componentes principais do CLNF são: (i) o PDM que captura variações na forma dos pontos fiduciais e (ii) o conceito de *patchexperts*, para capturar as variações locais de aparência para cada ponto fiducial.

DLib <sup>1</sup> é uma biblioteca escrita em C++ que oferece um conjunto de ferramentas de aprendizado de máquina, entre as quais está a ferramenta projetada para detectar pontos fiduciais de faces. Para detecção de rosto, esta biblioteca usa uma abordagem baseada em histogramas de gradientes orientados (HOG) combinados com um classificador linear, uma pirâmide de imagem e uma janela deslizante de detecção. Além disso, o processo de detecção

de pontos fiduciais emprega o estimador de pose proposto por (KAZEMI; SULLIVAN, 2014).

A abordagem apresentada por (BULAT; TZIMIROPOULOS, 2017), denominada Face Alignment, consiste na localização dos pontos fiduciais faciais por uma rede neural de bloqueio residual profundo, originalmente, treinada para o alinhamento facial em modelos 2D e 3D. Além disso, os autores também propuseram uma arquitetura para a conversão de pontos 2D em 3D.

### 5.2.4 Apresentação e Discussão dos Resultados Experimentais

Nesta seção, são apresentados e discutidos os resultados experimentais desta pesquisa. Inicialmente, são avaliadas e comparadas as estratégias adotadas no Experimento 1. O RMSE das localizações dos pontos fiduciais foi calculado. Para identificar possíveis diferenças entre as abordagens comparadas, foi realizada uma análise estatística desses dados. Os resultados RMSE estão sumariados na Tabela 5.7.

Tabela 5.7 Estatísticas RMSE do Experimento 1

Abordagem	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão
<b>MED</b>	<b>1,00</b>	5,23	6,98	6,37	8,06	<b>9,97</b>	2,44
<b>MFO</b>	5,11	5,80	6,14	6,12	6,43	7,56	<b>0,51</b>
<b>PFO</b>	3,54	4,44	5,00	4,95	5,48	6,57	0,79
<b>PDM</b>	2,25	3,30	3,67	3,76	4,29	5,07	0,62
<b>AAD</b>	1,03	1,74	<b>2,02</b>	<b>2,17</b>	2,41	9,40	0,99

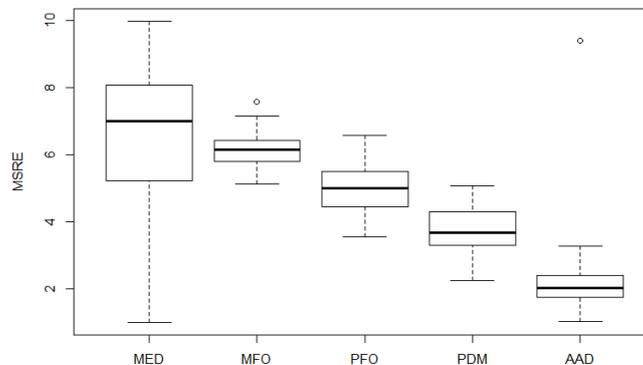
Fonte: Autoria própria

Dentre as abordagens investigadas, a abordagem MED produziu a maior variação RMSE. Por sua vez, o AAD obteve a menor mediana e a menor média. Por fim, o MFO obteve o menor desvio padrão (ver valores em negrito na Tabela 5.7). A abordagem AAD produziu a mediana mais baixa e os resultados médios mais baixos. Além disso, a abordagem MFO apresentou o menor desvio padrão, de acordo com os valores destacados em negrito na Tabela 5.7. Para melhor visualização dos valores RMSE obtidos, gráficos de *boxplot* foram

<sup>1</sup><http://dlib.net>

desenhados para as abordagens investigadas (ver Figura 5.3).

Figura 5.3 Boxplots do RMSE das abordagens avaliadas no Experimento 1



Fonte: Autoria própria

Após a sumarização do RMSE para as abordagens consideradas, foi verificada a normalidade dos dados. Todos os testes estatísticos contidos nos parágrafos seguintes têm uma significância de 0,05 ( $\alpha = 0,05$ ). Para isso, foi administrado o teste de Shapiro-Wilk (GHASEMI; ZAHEDIASL, 2012), cujos resultados são apresentados na Tabela 5.8.

Tabela 5.8 Teste de normalidade Shapiro-Wilk para distribuições RMSE de abordagens avaliadas no Experimento 1

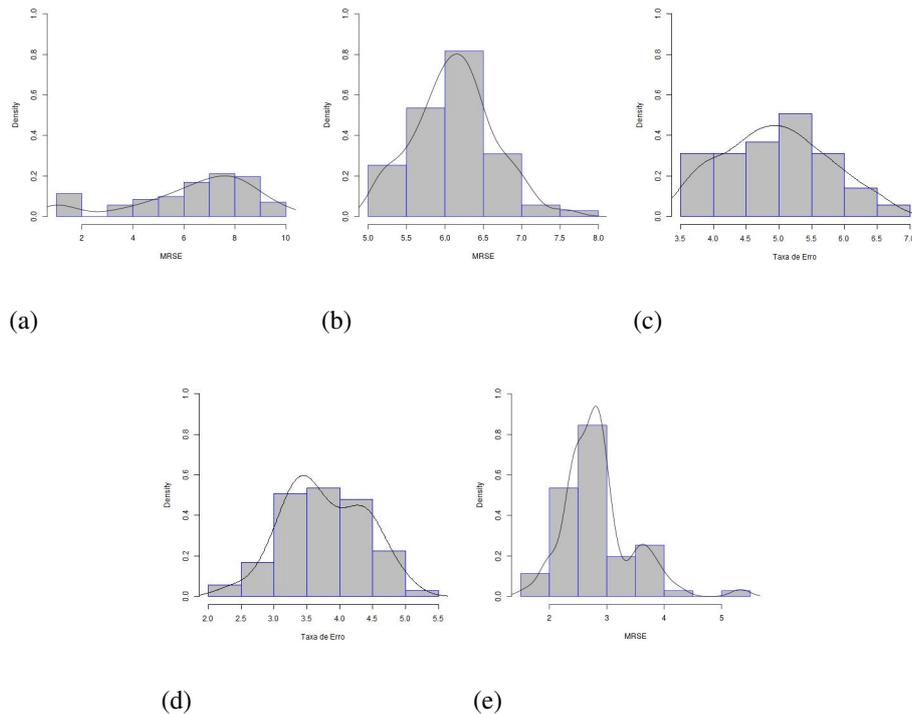
	<b>W</b>	<b>P-value</b>
<b>MED</b>	0,89	<b>2,24e-05</b>
<b>MFO</b>	0,98	0,72
<b>PFO</b>	0,97	0,22
<b>PDM</b>	0,98	0,48
<b>AAD</b>	0,53	<b>1,11e-13</b>

Fonte - Autoria Própria

De acordo com o teste de Shapiro-Wilk, as abordagens MED e AAD não apresentaram distribuições normais, uma vez que os p-valor obtidos eram superiores ao nível de significância considerado. Para facilitar a identificação das abordagens, cujos dados não apresentavam distribuições normais, histogramas relacionados aos resultados das abordagens MED, MFO, PFO, PDM e AAD foram plotados (ver Figuras 5.7(a) à 5.7(e), respectivamente). Os valores

RMSE para cada abordagem estão no eixo x, enquanto a frequência relativa de tais valores de erro está, em uma escala normalizada de zero a um, no eixo y.

Figura 5.4 Histogramas de valores RMSE para o Experimento 1. (a) MED, (b) MFO, (c) PFO, (d) PDM, (e) AAD



Fonte: Autoria própria

Uma vez que pelo menos uma das abordagens não seguiu uma distribuição normal, o teste de Kruskal-Wallis (OSTERTAGOVÁ; OSTERTAG; KOVÁČ, 2014) foi administrado para verificar as diferenças estatísticas entre as abordagens analisadas. Conforme apresentado na Tabela 5.9, infere-se que há diferenças estatísticas entre os resultados das abordagens analisadas ( $p\text{-valor} < \alpha$ ).

Tabela 5.9 Teste de Kruskal-Wallis para avaliar diferenças estatísticas entre as abordagens no Experimento 1

	Chi-squared	Df	P-value
<b>Kruskal-Wallis</b>	228,36	4,00	<b>2,20e-16</b>

Fonte - Autoria Própria

O teste de Kruskal-Wallis permitiu verificar a existência de diferenças significativas entre

as cinco abordagens avaliadas. No entanto, para verificar quais pares eram significativos, o teste *post-hoc* de Dunn foi administrado (HAHS-VAUGHN; LOMAX, 2013). Pode-se observar nos valores em negrito da Tabela 5.10 que todas as abordagens apresentaram diferenças estatísticas, apresentando  $p\text{-valor} > \alpha$ , com exceção das abordagens MED e MFO. Os números nas células da tabela representam a comparação entre o par de abordagens correspondente (valor superior) e o respectivo valor  $p$  (valor inferior).

A partir dos resultados estatísticos obtidos, é possível afirmar que o AAD apresentou melhor acurácia em relação às demais abordagens, pois apresentou a menor taxa de erro e nenhuma abordagem concorrente apresentou erro estatisticamente equivalente.

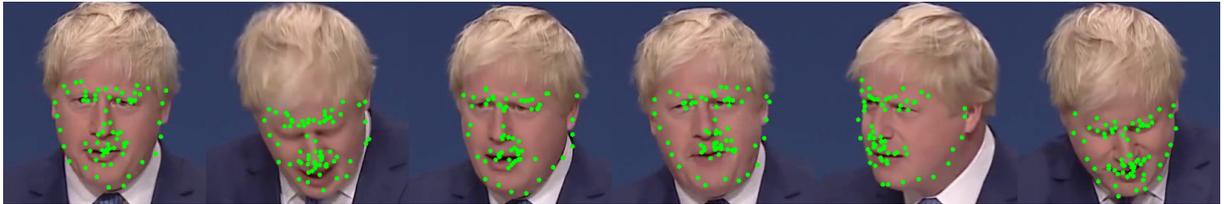
Tabela 5.10 Resultados do teste *post-hoc* de Dunn para identificar pares de métodos no Experimento 1 que têm diferenças estatisticamente significativas.

	<b>MED</b>	<b>MFO</b>	<b>PFO</b>	<b>PDM</b>
<b>MFO</b>	-0,41 0,33	-	-	-
<b>PFO</b>	<b>3,76</b> <b>0,00</b>	<b>4,18</b> <b>0,00</b>	-	-
<b>PDM</b>	<b>7,77</b> <b>0,00</b>	<b>8,19</b> <b>0,00</b>	<b>4,00</b> <b>0,00</b>	-
<b>AAD</b>	<b>12,15</b> <b>0,00</b>	<b>12,57</b> <b>0,00</b>	<b>8,38</b> <b>0,00</b>	<b>4,37</b> <b>0,00</b>

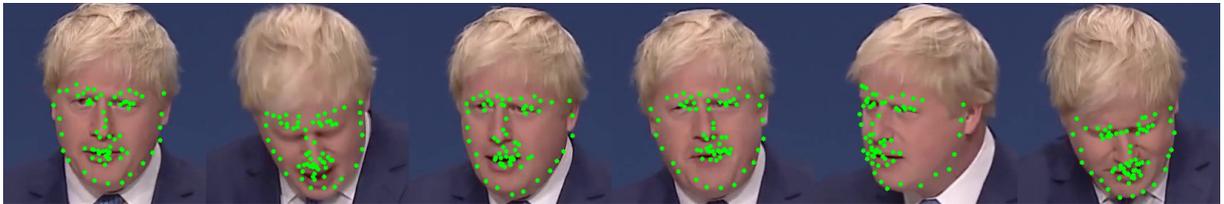
Fonte: Autoria própria

Também é possível observar nas Figuras 5.5(a) a 5.5(e) (MED, MFO, PFO, PDM, AAD, respectivamente) a diferença entre as abordagens analisadas no Experimento 1, em que os resultados dos pontos fiduciais obtidos a partir de uma sequência de quadros de um vídeo são apresentados. Nesta sequência de imagens, é possível verificar que a abordagem AAD (Figura 5.5(e)) é mais precisa na detecção de pontos fiduciais dos olhos e no contorno da face, principalmente nas imagens em que são apresentadas maiores rotações faciais.

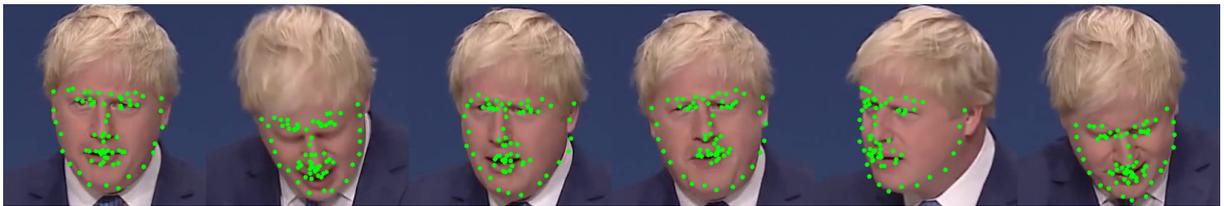
Figura 5.5 Resultados visuais das abordagens analisadas no Experimento 1, em uma sequência de quadros de um vídeo da base de dados 300-vw. (a) MED, (b) MFO, (c) PFO, (d) PDM, (e) AAD



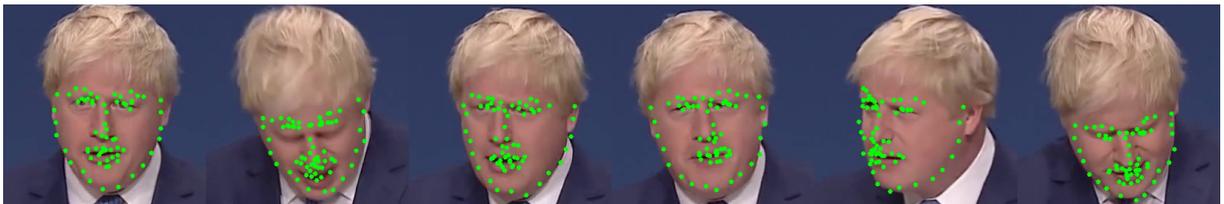
(a)



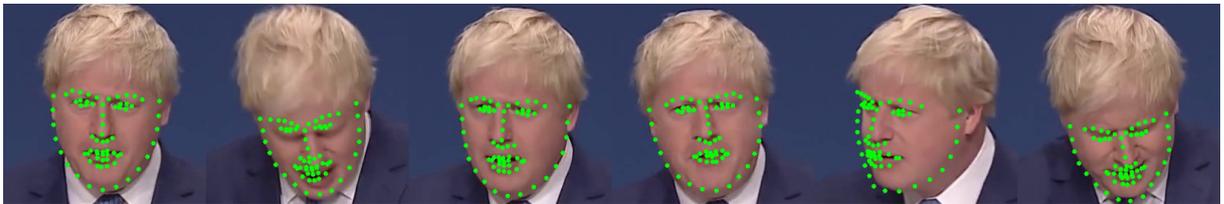
(b)



(c)



(d)



(e)

Fonte: Autoria própria

Após a comparação dos resultados entre as abordagens do Experimento 1, realizou-se o Experimento 2, no qual o melhor resultado (AAD) é comparado às abordagens concorrentes apresentadas na Seção 5.2.3.

Os resultados do RMSE para o Experimento 2 estão contidos na Tabela 5.11. A abordagem AAD produziu a maior variação RMSE entre os algoritmos avaliados, bem como os menores valores de mediana e média. Além disso, a abordagem FAL obteve o menor desvio padrão. Os valores acima estão destacados em negrito na Tabela 5.11. Para visualização das estatísticas RMSE comparadas, *boxplots* são apresentados na Figura 5.6.

Tabela 5.11 Estatísticas RMSE obtidas no Experimento 2

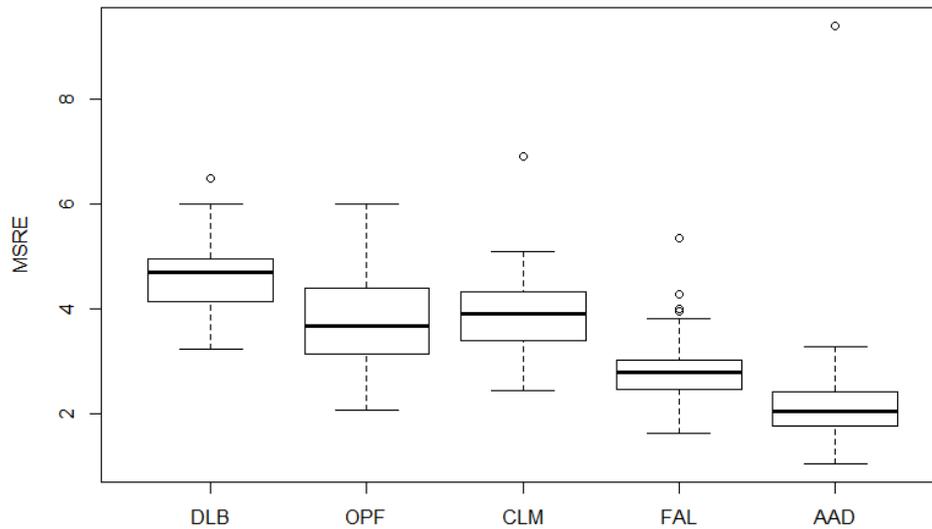
	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão
<b>DLB</b>	3,21	4,11	4,67	4,65	4,95	6,48	0,64
<b>OPF</b>	2,04	3,11	3,67	3,75	4,38	5,98	0,81
<b>CLM</b>	2,43	3,33	3,89	3,89	4,33	6,89	0,74
<b>FAL</b>	1,60	2,44	2,78	2,85	3,04	5,33	<b>0,60</b>
<b>AAD</b>	<b>1,03</b>	1,74	<b>2,02</b>	<b>2,17</b>	2,41	<b>9,40</b>	0,99

Fonte: Autoria própria

Similarmente ao primeiro processo de comparação, descrito anteriormente, após a sumariização do RMSE para as abordagens consideradas, foi testada a normalidade dos dados. A fim de facilitar a identificação das abordagens cujos dados não tinham distribuições normais, histogramas relacionados aos resultados das abordagens CLM, DLB, FAL, OPF e AAD foram plotados (ver Figuras 5.7(a) a 5.7(e), respectivamente). Os valores RMSE para cada algoritmo estão no eixo x, enquanto a frequência de tais valores de erro está, em uma escala normalizada de zero a um, no eixo y. Todos os testes estatísticos contidos nos parágrafos a seguir consideraram uma significância de 0,05 ( $\alpha = 0,05$ ). Para isso, foi administrado o teste de Shapiro-Wilk (GHASEMI; ZAHEDIASL, 2012), cujos resultados são sumariados na Tabela 5.12.

De acordo com esse teste, as abordagens CLM, FAL e AAD não apresentaram distribuições normais, pois apresentaram resultados de p-valor superiores ao nível de significância considerado.

Figura 5.6 Boxplots de estatísticas RMSE para o Experimento 2



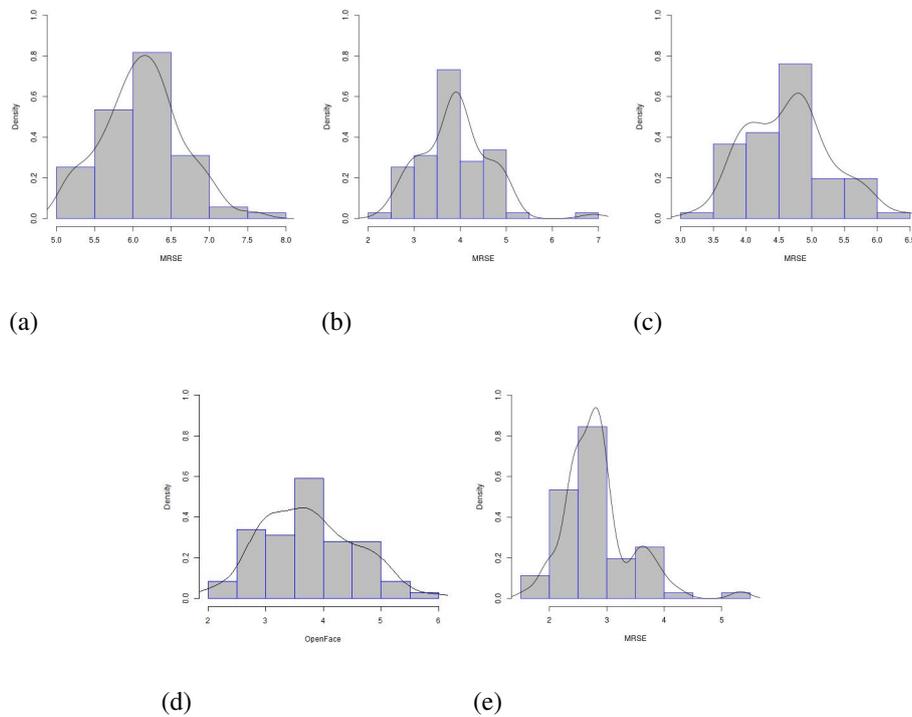
Fonte: Autoria própria

Tabela 5.12 Resultados do teste de Shapiro Wilk para distribuições RMSE de abordagens avaliadas no Experimento 2

	<b>W</b>	<b>P-value</b>
<b>DLB</b>	0,97	0,25
<b>OPF</b>	0,97	0,27
<b>CLM</b>	0,94	<b>0,00</b>
<b>FAL</b>	0,91	<b>0,00</b>
<b>AAD</b>	0,53	<b>1,11e-13</b>

Fonte: Autoria própria

Figura 5.7 Histogramas de valores RMSE para o Experimento 2. (a) CLM, (b) DLB, (c) FAL, (d) OPF e (e) AAD



Fonte: Autoria própria

Mais uma vez, como pelo menos uma das abordagens não seguia uma distribuição normal, o teste de Kruskal-Wallis foi administrado para verificar as diferenças estatísticas entre as abordagens analisadas (OSTERTAGOVÁ; OSTERTAG; KOVÁČ, 2014). Conforme apresentado na Tabela 5.13, pode-se inferir que houve diferenças estatísticas entre os resultados das abordagens analisadas ( $p\text{-valor} < \alpha$ ).

Tabela 5.13 Resultado do teste de Kruskal para experimento 2

	<b>Chi-squared</b>	<b>Df</b>	<b>P-value</b>
<b>Kruskal-Wallis</b>	229,85	4,00	<b>2,20e-16</b>

Fonte - Autoria Própria

Finalmente, para verificar em quais pares havia diferenças estatisticamente significativas, o teste *post-hoc* de Dunn foi administrado (HAHS-VAUGHN; LOMAX, 2013). Na Tabela 5.14, observa-se que todas as abordagens apresentaram diferenças estatísticas, adotando-se o nível de significância apresentado anteriormente, com exceção das abordagens OPF e CLM, que não apresentaram diferenças ( ver Tabela 5.14). Nos pares de valores apresentados, o valor superior corresponde à comparação entre os pares e o inferior, ao p-value. Por fim, pode-se inferir que a abordagem AAD obteve o melhor resultado quando comparada com outros algoritmos de código aberto que possuem a mesma finalidade.

Tabela 5.14 Teste *post-hoc* de Dunn para comparação entre as abordagens do Experimento 2.

	<b>DLB</b>	<b>OPF</b>	<b>CLM</b>	<b>FAL</b>
<b>OPF</b>	<b>-9,05</b> <b>0,00</b>	-	-	-
<b>CLM</b>	<b>-9,87</b> <b>0,00</b>	-0,82 0,20	-	-
<b>FAL</b>	<b>-13,62</b> <b>0,00</b>	<b>-4,56</b> <b>0,00</b>	<b>-3,74</b> <b>0,00</b>	-
<b>AAD</b>	<b>-3,84</b> <b>0,00</b>	<b>5,20</b> <b>0,00</b>	<b>6,02</b> <b>0,00</b>	<b>9,77</b> <b>0,00</b>

Fonte: Autoria própria

Também é possível observar nas Figuras 5.8(a) a 5.8(e) (AAD, CLM, DLB, OPF, FAL, respectivamente), a diferença entre as abordagens analisadas no Experimento 2, em que os resultados de pontos fiduciais foram obtidos a partir de uma sequência de quadros de um vídeo. Realizada uma comparação com algoritmos de código aberto, pode-se observar que todos os algoritmos, exceto o proposto, não apresentam bons resultados quando há mudanças na pose da face.

A partir dos resultados apresentados nesta seção, pode-se concluir que a abordagem proposta obteve o menor erro RMSE, seguido pelos algoritmos FAL e OPF. Isso pode ser explicado pelo fato de que a abordagem proposta utiliza informações temporais contidas em uma sequência de quadros de vídeo. Assim, a posição dos pontos fiduciais nos próximos quadros pode ser prevista, aumentando sua precisão, principalmente naqueles quadros em que as outras abordagens comparadas não conseguem detectar os pontos fiduciais.

### **5.3 Avaliação da Abordagem Proposta Com Usuários (QDPU)**

O objetivo desta pesquisa foi desenvolver uma metodologia para a geração de caricaturas em vídeos, que possa converter o processo de observação e as habilidades artísticas humanas em um sistema computacional. Neste contexto, definiram-se como características essenciais:

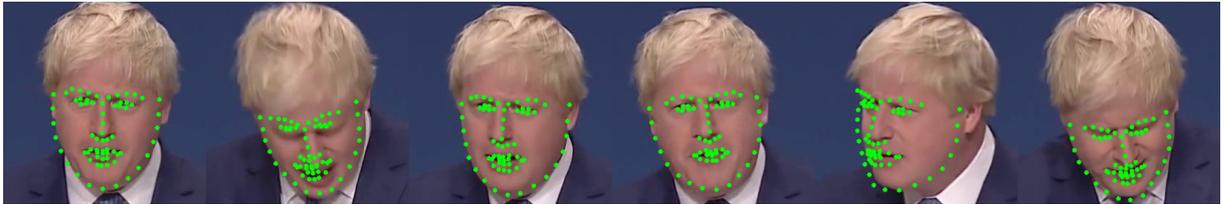
1. A preservação da coerência temporal
2. A precisão da detecção de características faciais mais evidentes
3. Aplicação voltada para o entretenimento

Os critérios definidos para a avaliação dos vídeos caricaturados automaticamente são:

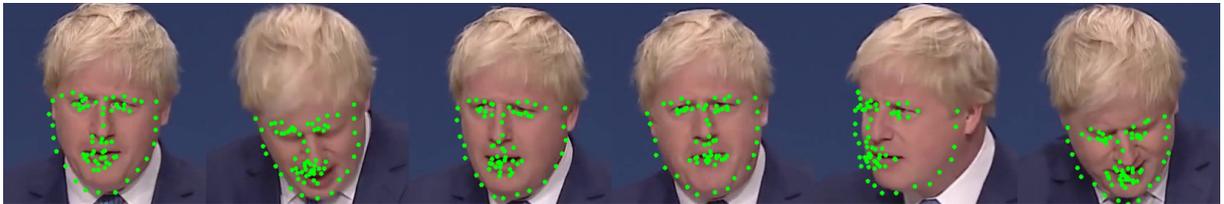
- Coerência das regras de deformação: o vídeo caricaturado deve apresentar coerência na deformação realizada sobre as características faciais de interesse. Isto é, as características faciais definidas como mais evidentes devem se manter deformadas por todo o vídeo;

Figura 5.8 Resultados visuais das abordagens analisadas no Experimento 2, em uma sequência de quadros de um vídeo da base de dados 300-vw. (a) AAD, (b) CLM, (c) DLB, (d) OPF, (e)

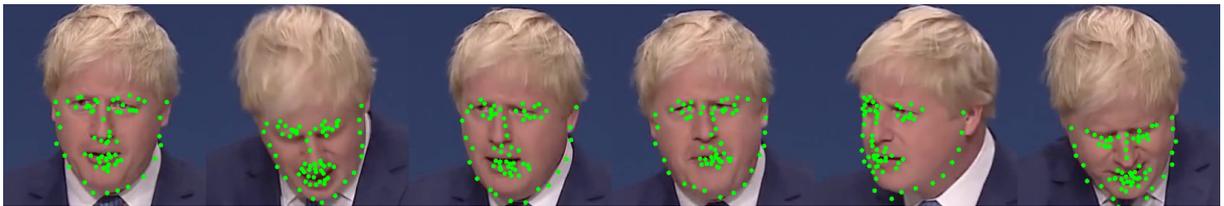
FAL



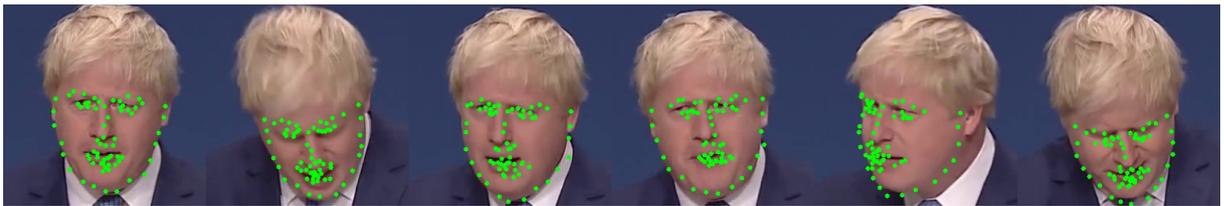
(a)



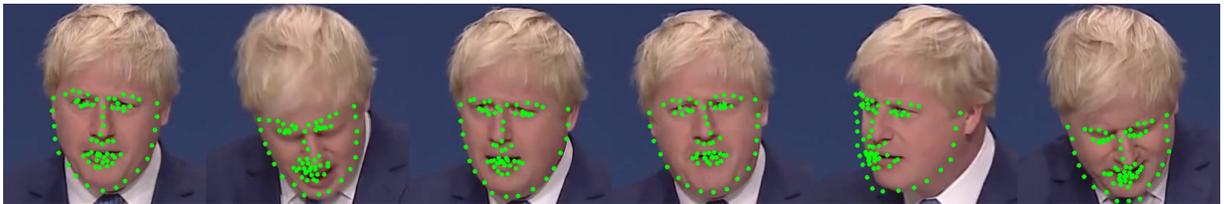
(b)



(c)



(d)



(e)

Fonte: Autoria própria

- Semelhança ao conteúdo original: a face caricaturada deve permitir ao observador o reconhecimento da face original. Portanto, o observador deve reconhecer na face caricaturada características essenciais para identificar a face original;
- Coerência temporal: o vídeo caricaturado não deve apresentar cintilações nem inconsistências no processo de deformação devido a variações de pose e expressões faciais;
- Preferência visual: o vídeo caricaturado deve apresentar um aspecto visual final agradável;
- Finalidade da Aplicação: os resultados devem evidenciar a finalidade de entretenimento do vídeo caricaturado.

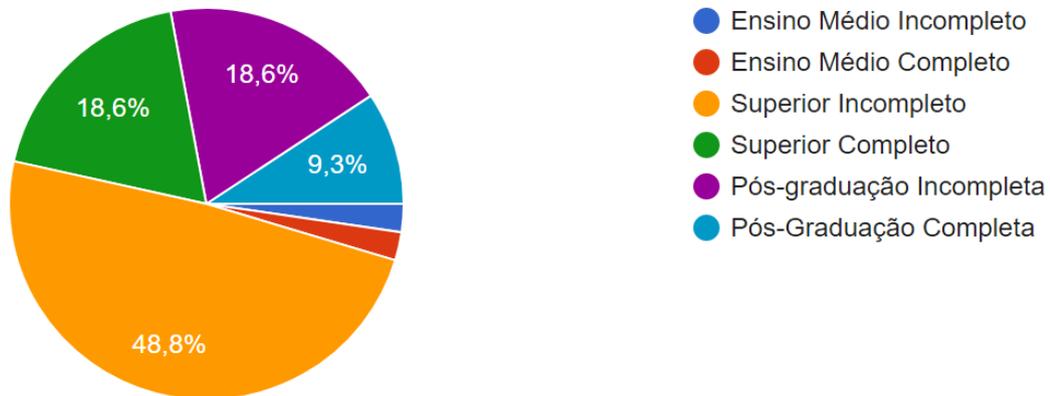
Nas próximas subseções serão apresentados os resultados obtidos a partir da sondagem subjetiva da opinião do usuário. Ao todo, foram coletados, por meio de formulários do *Google Forms*, as respostas de 43 usuários convocados via e-mail. Na Subseção 5.3.1, são apresentados os resultados da análise do perfil de usuário. Na Subseção 5.3.2, são apresentados os resultados da análise da avaliação da abordagem proposta.

### **5.3.1 Resultados da Análise do Questionário de Delineamento de Perfil do Usuário**

A partir dos dados coletados oriundos do QDPU, foi possível delinear o perfil dos participantes no processo de avaliação da abordagem proposta. As Figuras 5.9 a 5.14, apresentam as distribuições obtidas a partir do questionário de delineamento de perfil do usuário.

Na Figura 5.9, é possível observar que, para os usuários utilizados neste estudo, 48, 80% possuem ensino superior incompleto, 18, 60% possuem ensino superior completo, 18, 60% possuem pós-graduação incompleta, 9, 30% possuem pós-graduação completa, 2, 30% possuem ensino médio incompleto e 2, 40% possuem ensino médio completo.

Figura 5.9 Distribuição da Escolaridade dos Participantes



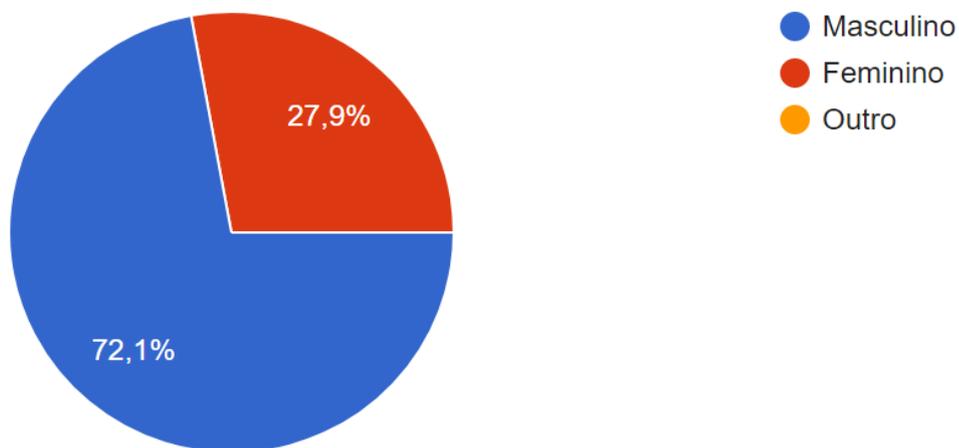
Fonte: Autoria própria

Na Figura 5.10, é possível observar que 72,20% dos usuários observados pertencem ao gênero masculino, enquanto 27,90% dos usuários declararam pertencer ao gênero feminino. Na Figura 5.11, é possível observar que, 65,10% dos usuários declararam utilizar algum tipo de corretivo ocular, enquanto 34,90% declararam que não utilizam nenhum tipo de corretivo ocular. Na Figura 5.12, 74,40% dos usuários declararam observar caricaturas nos meios de comunicação utilizados, como por exemplo, jornais, revistas e sites, enquanto 25,60% declararam não observar.

Na Figura 5.13, 32,20% dos usuários afirmaram observar caricaturas nos meios de comunicação em uma frequência alta, 30,20% nem baixa nem alta, 16,30% baixa, 14,00% muito alta e 7,00% afirmaram observar caricaturas em uma frequência muito baixa. Na Figura 5.14, 53,50% dos usuários afirmaram ter alta facilidade em identificar uma caricatura, 25,60% nem baixa nem alta, 16,30% muito alta, 4,70% baixa e nenhum usuário afirmou ter muito baixa a facilidade de identificação de uma caricatura.

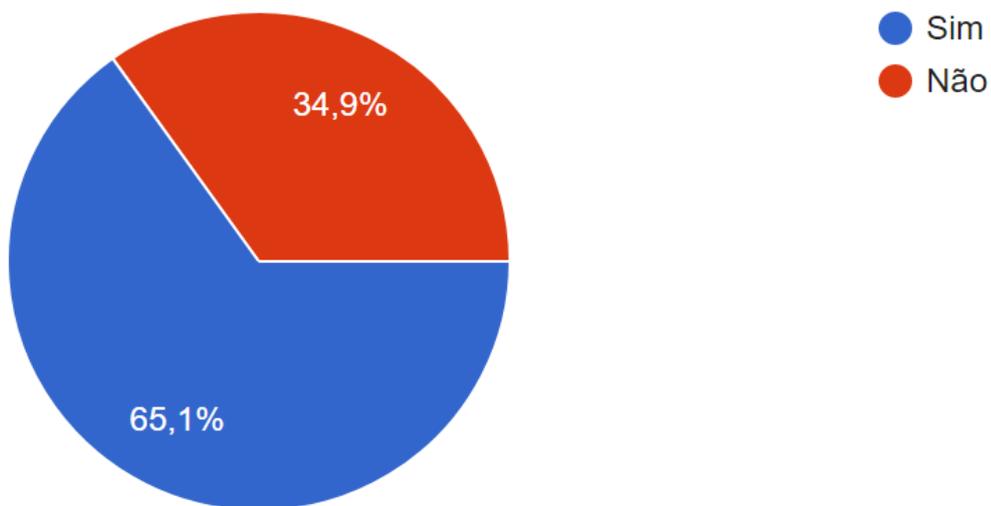
Visto que a questão referente à Figura 5.14 é de suma importância para a etapa de avaliação dos vídeos caricaturados gerados, dois usuários que afirmaram ter baixa facilidade em identificar uma caricatura não participaram do processo de avaliação da abordagem proposta.

Figura 5.10 Distribuição do Gênero dos Participantes



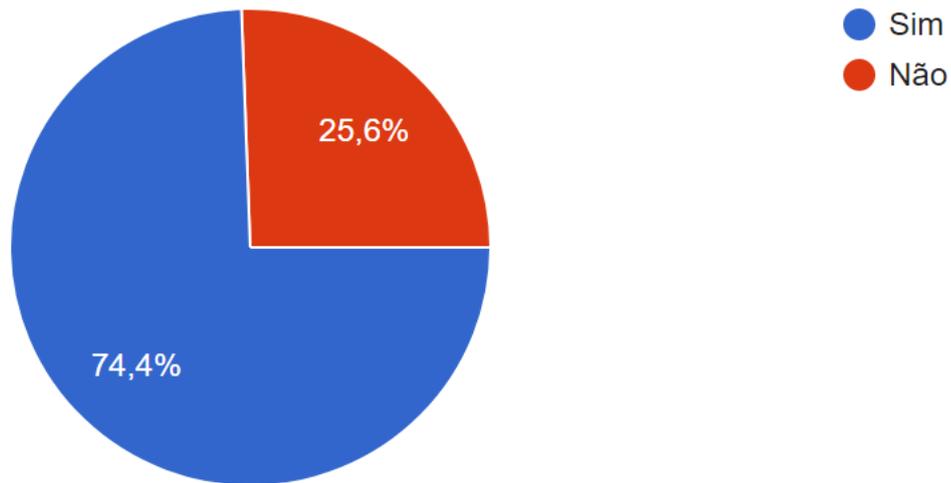
Fonte: Autoria própria

Figura 5.11 Distribuição do Uso de Corretivos Oculares dos Participantes



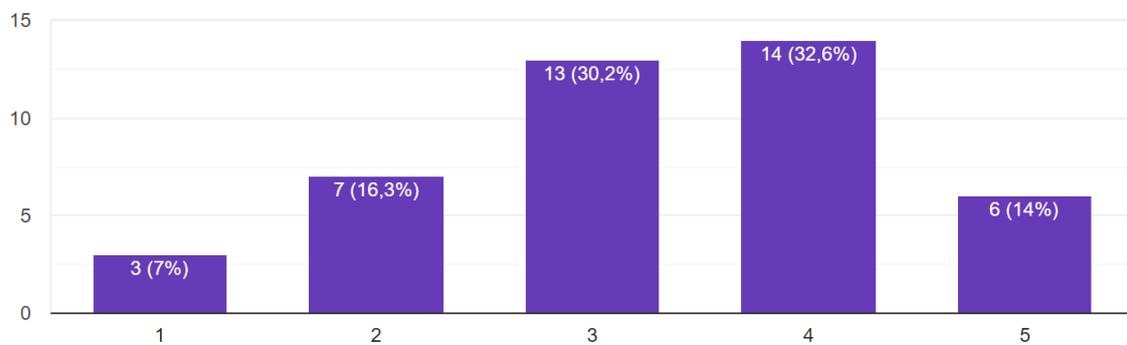
Fonte: Autoria própria

Figura 5.12 Distribuição do Costume dos Usuários de Observar Caricaturas nos Meios de Comunicação



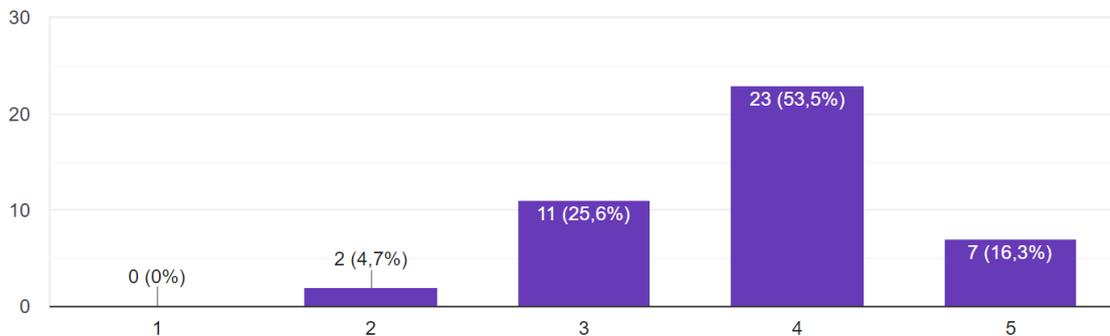
Fonte: Autoria própria

Figura 5.13 Distribuição da Frequência de Observações de Caricaturas nos Meios de Comunicação dos Participantes no Dia a Dia. Seguindo as seguintes âncoras semânticas: 1 - Muito Baixa; 2 - Baixa; 3 - Nem Baixa nem Alta; 4 - Alta; 5 - Muito Alta



Fonte: Autoria própria

Figura 5.14 Distribuição da Facilidade de Identificação de uma Caricatura. Seguindo as seguintes âncoras semânticas: 1 - Muito Baixa; 2 - Baixa; 3 - Nem Baixa nem Alta; 4 - Alta; 5 - Muito Alta



Fonte: Autoria própria

### 5.3.2 Resultados da Análise do Questionário para a Sondagem da Opinião do Usuário (QSOU)

Após a análise do perfil do usuário, foram realizadas as análises dos resultados obtidos a partir do questionário para a avaliação da abordagem proposta. Conforme informado na Seção 4.6, foi realizada uma avaliação interna da abordagem proposta. Para esta avaliação, foram apresentadas 3 afirmações para o usuário informar o seu grau de concordância, conforme apresentado na Tabela 4.1 que está presente na Seção 4.6. Também foram realizados quatro tipos de processamentos diferentes com o intuito de encontrar a melhor configuração de utilização da abordagem proposta. Para melhor entendimento, os valores das respostas foram normalizados para que, os valores mais próximos de 1 representem respostas desfavoráveis e valores mais próximos de 5 correspondam à respostas favoráveis.

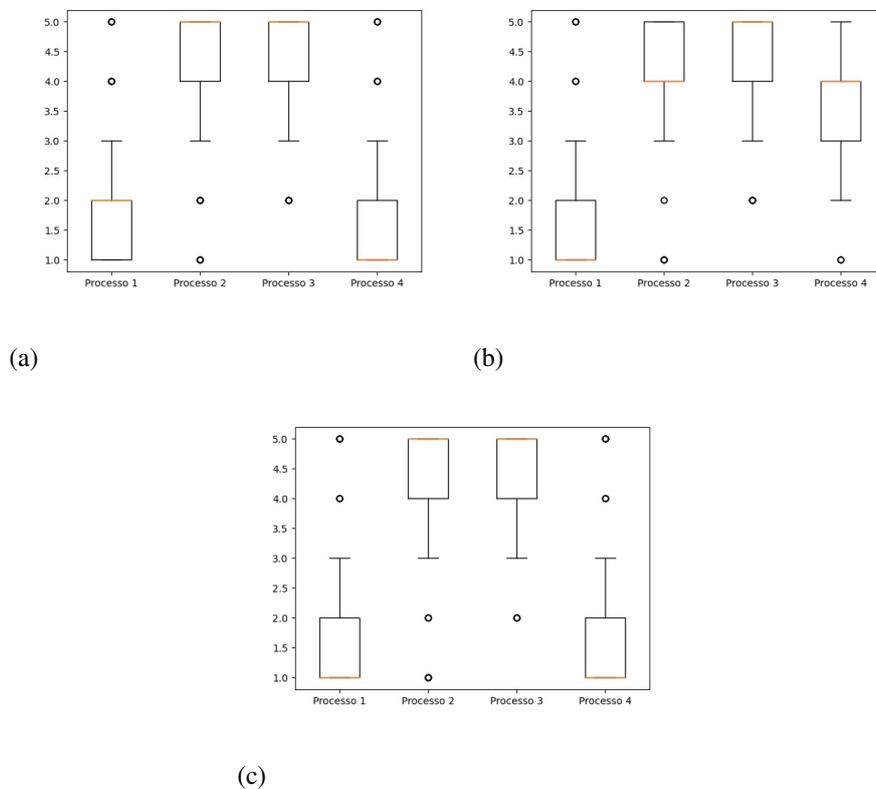
Após a obtenção dos dados, foram calculadas as médias das respostas dos usuários para cada questão observada. Estes valores podem ser observados na Tabela 5.15, em que é possível observar que os processamentos 2 e 3 obtiveram as maiores médias, e para o processamento 4 a Questão 2 obteve um valor médio de 3.80 de avaliação dos usuários. Nas Figuras 5.15(a) a (c) é possível observar a distribuição das avaliações obtidas por meio de um *boxplot*.

Para realizar a avaliação estatística dos dados, fez-se necessária primeiramente a verificação da normalidade dos dados. Para isso, foi aplicado o teste Shapiro-Wilk (GHASEMI; ZAHEDIASL, 2012), cujos resultados são apresentados nas Tabelas 5.16, 5.18 e 5.17, que

Tabela 5.15 Valores médios obtidos a partir da avaliação dos usuários

	<b>Média de Avaliação dos Usuários</b>		
	<b>Questão 1</b>	<b>Questão 2</b>	<b>Questão 3</b>
<b>Processamento 1</b>	1,87	1,72	1,85
<b>Processamento 2</b>	<b>4,19</b>	<b>4,19</b>	<b>4,20</b>
<b>Processamento 3</b>	<b>4,55</b>	<b>4,58</b>	<b>4,58</b>
<b>Processamento 4</b>	1,76	<b>3,80</b>	1,75

Fonte: Autoria própria

Figura 5.15 *Boxplots* das avaliações dos usuários para cada questão. (a) Questão 1; (b) Questão 2 e (c) Questão 3

Fonte: Autoria própria

apresentam os valores para as Questões 1, 2 e 3 respectivamente.

Tabela 5.16 Resultados do teste de Shapiro Wilk para a Questão 1

	<b>W</b>	<b>p-value</b>
<b>Processamento 1</b>	0,75	$\approx 0$
<b>Processamento 2</b>	0,71	$\approx 0$
<b>Processamento 3</b>	0,58	$\approx 0$
<b>Processamento 4</b>	0,70	$\approx 0$

Fonte: Autoria própria

Tabela 5.17 Resultados do teste de Shapiro Wilk para a Questão 2

	<b>W</b>	<b>p-value</b>
<b>Processamento 1</b>	0,70	$\approx 0$
<b>Processamento 2</b>	0,74	$\approx 0$
<b>Processamento 3</b>	0,56	$\approx 0$
<b>Processamento 4</b>	0,82	$\approx 0$

Fonte: Autoria própria

Tabela 5.18 Resultados do teste de Shapiro Wilk para a Questão 3

	<b>W</b>	<b>p-value</b>
<b>Processamento 1</b>	0,73	$\approx 0$
<b>Processamento 2</b>	0,71	$\approx 0$
<b>Processamento 3</b>	0,58	$\approx 0$
<b>Processamento 4</b>	0,70	$\approx 0$

Fonte: Autoria própria

Para uma maior segurança, também foram gerados gráficos de histograma para verificar visualmente se os dados seguiam uma distribuição normal. Nas Figuras 5.16, 5.17, 5.18 é possível observar que nenhum gráfico apresentado segue uma distribuição normal. Com base nos gráficos e nos testes de Shapiro Wilk é possível concluir que os dados utilizados não são normais.

Uma vez que pelo menos um dos valores de avaliação não seguiu uma distribuição normal, o teste de Kruskal-Wallis foi administrado para verificar as diferenças estatísticas entre os processamentos avaliados (OSTERTAGOVÁ; OSTERTAG; KOVÁČ, 2014). Na Tabela 5.19, são apresentados os valores obtidos, sendo possível observar a existência de diferenças estatísticas entre as variações de processamentos associadas a esta sondagem.

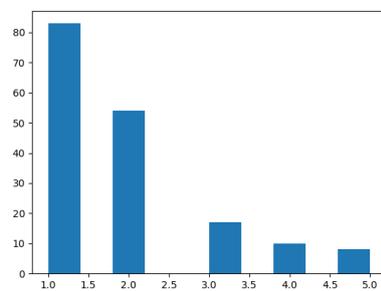
Tabela 5.19 Teste de Kruskal-Wallis para avaliar as diferenças estatísticas entre os processamentos abordados

	<b>Chi-squared</b>	<b>p-value</b>
<b>Questão 1</b>	395,65	$\approx 0$
<b>Questão 2</b>	358,38	$\approx 0$
<b>Questão 3</b>	399,02	$\approx 0$

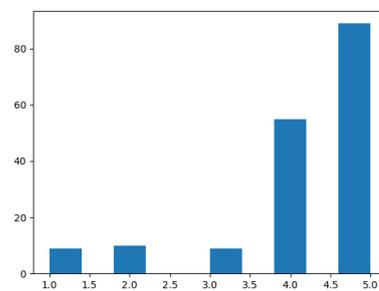
Fonte: Autoria própria

Foram quatro processamentos utilizados e o teste de Kruskal-Wallis mostrou que houve diferença significativa nos processamentos para todas as questões analisadas. No entanto, para verificar quais pares eram significativos, o teste *post-hoc* de Dunn foi administrado (HAHS-VAUGHN; LOMAX, 2013). Na Tabela 5.20, pode-se observar que, para a Questão

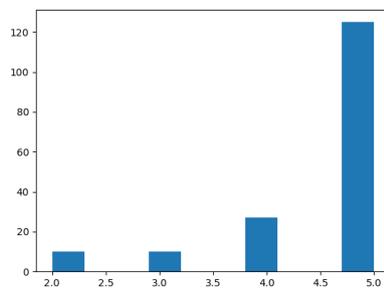
Figura 5.16 Histogramas das avaliações dos usuários para cada processamento da Questão 1. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4



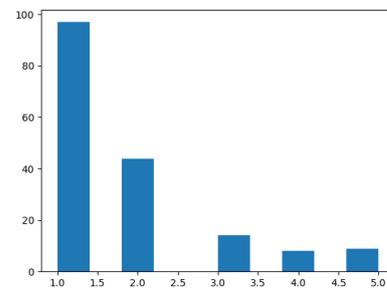
(a)



(b)



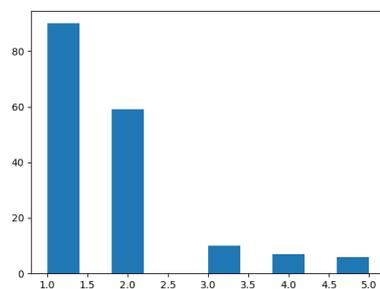
(c)



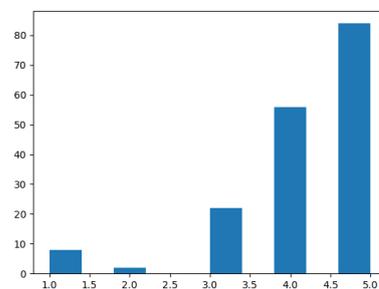
(d)

Fonte: Autoria própria

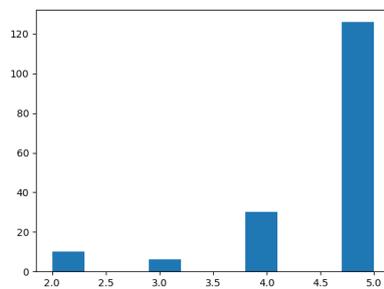
Figura 5.17 Histogramas das avaliações dos usuários para cada processamento da Questão 2. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4



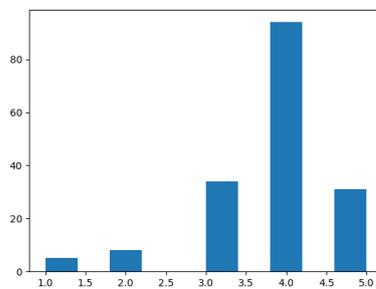
(a)



(b)



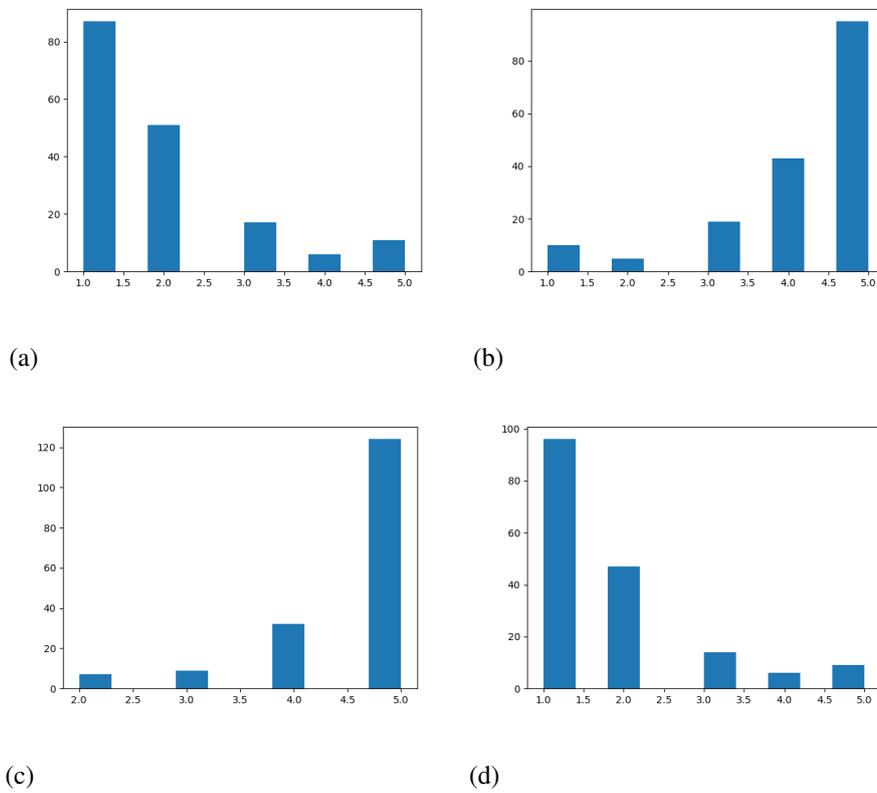
(c)



(d)

Fonte: Autoria própria

Figura 5.18 Histogramas das avaliações dos usuários para cada processamento da Questão 3. (a) Processamento 1; (b) Processamento 2; (c) Processamento 3 e (d) Processamento 4



Fonte: Autoria própria

1, para todos os processamentos ha diferenças estatísticas, com a exceção dos pares de processamentos 2-3 e 1-4.

Tabela 5.20 Resultados do teste *post-hoc* de Dunn para identificar pares de processamentos na Questão 1 que têm diferenças estatisticamente significativas

	<b>Proc 1</b>	<b>Proc 2</b>	<b>Proc 3</b>	<b>Proc 4</b>
<b>Proc 1</b>	1,00			
<b>Proc 2</b>	1,70e-34	1,00		
<b>Proc 3</b>	5,18e-49	<b>8,75e-02</b>	1,00	
<b>Proc 4</b>	<b>1,00</b>	3,10e-38	1,83e-53	1,00

Fonte: Autoria própria

Na Tabela 5.21, é possível observar que, para a Questão 2, todos os processamentos apresentaram diferenças estatísticas. Por fim, na Tabela 5.22 é possível observar que, para a Questão 3, para todos os processamentos existem diferenças estatísticas, com a exceção do par de processamentos 2-3.

Tabela 5.21 Resultados do teste *post-hoc* de Dunn para identificar pares de processamentos na Questão 2 que têm diferenças estatisticamente significativas

	<b>Proc 1</b>	<b>Proc 2</b>	<b>Proc 3</b>	<b>Proc 4</b>
<b>Proc 1</b>	1,00			
<b>Proc 2</b>	1,35e-46	1,00		
<b>Proc 3</b>	3,37e-70	4,75e-03	1,00	
<b>Proc 4</b>	7,27e-26	9,66e-04	6,06e-12	1,00

Fonte: Autoria própria

## 5.4 Considerações Finais

Neste capítulo, apresentaram-se e se discutiram os resultados obtidos a partir dos experimentos realizados o processo de validação da abordagem proposta. Inicialmente, foram apresentados os resultados obtidos no processo de definição dos algoritmos de detecção de pontos fiduciais a serem considerados. A partir da análise dos resultados obtidos, foi possível concluir que os

Tabela 5.22 Resultados do teste *post-hoc* de Dunn para identificar pares de processamentos na Questão 3 que têm diferenças estatisticamente significativas

	<b>Proc 1</b>	<b>Proc 2</b>	<b>Proc 3</b>	<b>Proc 4</b>
<b>Proc 1</b>	1,00			
<b>Proc 2</b>	2,49e-35	1,00		
<b>Proc 3</b>	3,44e-49	<b>1,23e-01</b>	1,00	
<b>Proc 4</b>	<b>1,00</b>	6,70e-39	2,13e-53	1,00

Fonte: Autoria própria

algoritmos Face Analysis, CLM e OpenFace, apresentavam, estatisticamente, taxas de erros equivalentes. Assim, optou-se pelo emprego do Openface, por ser um algoritmo mais recente, assim como por apresentar suporte à detecção de pontos fiduciais em imagens tridimensionais.

Também foram apresentados e discutidos os resultados obtidos a partir da análise do algoritmo adaptativo apresentado na Seção 4.2, na qual observou-se que a utilização do algoritmo adaptativo apresenta uma melhora da taxa de erro da detecção dos pontos fiduciais em vídeos.

Por fim, foram apresentados e discutidos os resultados obtidos a partir da avaliação da abordagem proposta com os voluntários recrutados. Foi possível concluir que os processamentos 2 (utilizando o algoritmo adaptativo e nível 3 de deformação) e 3 (utilizando o algoritmo adaptativo e nível 2 de deformação) obtiveram as melhores avaliações entre os usuários. Ver a Seção 4.6 para maiores detalhes das implementações utilizadas.

No próximo capítulo, serão apresentadas as considerações finais desta pesquisa, referentes às questões de pesquisa introduzidas no Capítulo 1, assim como as principais contribuições advindas desta pesquisa e recomendações de pesquisas futuras.

# Capítulo 6

## Considerações Finais

Neste capítulo, são formuladas as considerações finais desta pesquisa. Na Seção 6.1, apresenta-se uma síntese da pesquisa, bem como são feitas algumas considerações finais. Na Seção 6.2, apresentam-se as principais contribuições advindas da condução da pesquisa para a área de geração automática de caricaturas em vídeos. Por fim, na Seção 6.3, apresentam-se recomendações de pesquisas futuras nesta linha de pesquisa.

### 6.1 Síntese da Dissertação

No presente estudo, enfatiza-se que, para a geração de uma caricatura, o artista utiliza duas etapas básicas: a observação e o exagero. Com base nestes dois pontos principais, foi desenvolvida uma abordagem que visa a adaptação, a observação e o exagero realizado por humanos a partir do auxílio de pontos fiduciais e da análise das características faciais, além de algoritmos para conduzir o processo exagero das características de interesse. Além disso, a abordagem proposta focaliza a geração de vídeos caricaturados. Para tanto, fez-se necessária a utilização de estratégias que possibilitassem a uniformidade da deformação ao longo dos quadros do vídeo processado.

Conforme apresentado na Seção 1.4, o presente estudo objetivou responder as seguintes questões de pesquisa:

- **QP1:** É possível conceber uma abordagem que produza caricaturas de faces humanas em vídeos, as quais apresentem coerência temporal e sejam bem avaliadas por observadores humanos?

- **QP2:** Qual nível de deformação é mais adequado para um conjunto de vídeos caricaturizado a partir da abordagem proposta?

É importante observar que as conclusões obtidas neste estudo estão associadas à base de dados utilizada e ao conjunto de usuários que avaliaram os resultados da metodologia aplicada. Apesar dos usuários avaliarem positivamente a metodologia utilizada e conseqüentemente os resultados visuais, existe a possibilidade de que outros vídeos com outras características visuais possam produzir resultados de avaliação diferentes.

Considerando-se os resultados apresentados na Capítulo 5, formularam-se as seguintes conclusões:

**Argumento 1:** De acordo com a Questão 1 apresentada na Tabela 4.1 que diz: "Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção)". Fundamentando-se nos resultados apresentados na Tabela 5.15, é possível observar que os processamentos 2 e 3 obtiveram uma nota média de avaliação de 4,19 e 4,55 respectivamente. É importante lembrar que o processamento 2 consiste na utilização do algoritmo adaptativo e nível 3 de deformação, por sua vez, o processamento 3 consiste na utilização do algoritmo adaptativo e nível 2 de deformação, conforme apresentada na Seção 4.6. Na Tabela 5.20, é possível observar que, para estes dois processamentos não apresentou-se diferenças estatísticas nos resultados. Por fim, de acordo com a escala de Likert utilizada, a votação média dos usuários concordaram com a afirmação da Questão 1.

**Conclusão 1:** Para a amostra considerada, a geração dos vídeos distorcidos sem assemelham a uma caricatura utilizando os processamentos 2 e 3.

**Argumento 2:** Conforme apresentada na Seção 4.6 o processamento 1 não utiliza o algoritmo adaptativo, e por sua vez, o processamento 2 utiliza o algoritmo adaptativo. Observando os resultados obtidos a partir da Tabela 5.15 é possível observar que o processamento 2 foi melhor avaliado do que o processamento 1 em todas as questões apresentadas. Também é possível observar nas Tabelas 5.20, 5.21 e 5.22 que existem diferenças estatísticas entre os dois processamentos.

**Conclusão 2:** Para a amostra considerada, o uso do algoritmo adaptativo auxilia as distorções geradas.

**Argumento 3:** De acordo com a Questão 3 apresentada na Tabela 4.1 que diz: "O vídeo a

seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)". Fundamentando-se nos resultados apresentados na Tabela 5.15, é possível observar que os processamentos 2 e 3 obtiveram uma nota média de avaliação de 4.20 e 4.58 respectivamente. Na Tabela 5.22, é possível observar que, para estes dois processamentos, não apresentaram-se diferenças estatísticas nos resultados.

**Conclusão 3:** Para a amostra considerada, o modelo de deformação das características se comporta estável entre os quadros gerados para os processamentos 2 e 3.

**Argumento 4:** Conforme apresentada na Seção 4.6, a diferença entre os processamentos 2, 3 e 4 se dá apenas no nível de deformação utilizada. De acordo com a Tabela 5.15, os processamentos 2 e 3 obtiveram os maiores resultados de avaliação dos usuários. Também é possível observar nas Tabelas 5.21 e 5.22 que não existem diferenças estatísticas entre o processamento 2 e o processamento 3.

**Conclusão 4:** Para a amostra considerada, os níveis de deformação mais adequados para os vídeos analisados são aqueles dos processamentos 2 e 3.

**Argumento 5:** De acordo com os resultados apresentados na Tabela 5.11, é possível observar que o algoritmo adaptativo apresenta taxa de erro menor do que seu algoritmo original (Openface). Na Tabela 5.14, é possível observar que existem diferenças estatísticas entre as duas abordagens.

**Conclusão 5:** Para a amostra de vídeos utilizadas, o modelo do algoritmo adaptativo promove maior precisão na detecção dos pontos fiduciais.

**Argumento 6:** De acordo com os resultados apresentados na Tabela 5.11, é possível observar que o algoritmo adaptativo apresentou a menor taxa de erro quando comparado com algoritmos semelhantes. Na Tabela 5.14 é possível observar que, estatisticamente, a taxa de erro apresentada pelo algoritmo adaptativo não pertence a nenhum outro grupo dos algoritmos analisados.

**Conclusão 6:** Para a amostra de vídeos utilizados e para os algoritmos analisados, o modelo do algoritmo adaptativo apresenta os menores erros.

## 6.2 Contribuições do Estudo Realizado

A partir dos resultados obtidos com a avaliação do algoritmo adaptativo, considera-se tal algoritmo como uma contribuição relevante desta pesquisa. A utilização da estratégia abordada neste estudo pode vir a servir para diversas outras áreas que necessitam de uma detecção mais precisa de pontos fiduciais em vídeos.

Outra contribuição importante diz respeito à metodologia desenvolvida para a geração de caricaturas no contexto de vídeos digitais. É uma metodologia que pode facilmente ser aplicada a outros contextos que não sejam necessariamente a geração de caricaturas, como por exemplo, a criação de filtros para “embelezamento” ou envelhecimento da face no contexto de vídeos.

Por fim, a própria avaliação estatística dos resultados por meio de um questionário de sondagem da opinião de usuários pode ser considerada como uma contribuição deste estudo. Mesmo que, a amostra possa ser considerada pequena se comparada com toda a população de potenciais usuários de uma aplicação fundamentada na abordagem proposta nesta dissertação, a amostra representa os resultados estatísticos de um recorte desta população.

## 6.3 Estudos Futuros

Uma parte importante da caracterização das caricaturas geradas por artistas é sua estilização. Neste contexto, se torna importante realizar uma pesquisa para a geração de caricaturas estilizadas em vídeo. Focando-se na deformação da face (realizada no presente estudo) e na sua estilização (trabalho futuro) entre os quadros dos vídeos, mantendo sempre a coerência temporal. Entende-se por estilização, quaisquer modificações na aparência do vídeo em termos da distribuição de suas cores, do estilo dos contornos dos objetos, entre outros.

Outro ponto importante da geração das caricaturas criadas por artistas é alguma mensagem que é passada para o público, neste contexto, se torna relevante também uma pesquisa que objetive uma análise de todos os elementos do vídeo com a finalidade de extrair alguma informação relevante que sirva de apoio na geração das caricaturas.

Por fim, é de fundamental relevância realizar uma análise estatística dos resultados a partir de usuários especialistas, ou seja, de caricaturistas. Dessa forma seria possível comparar os

resultados obtidos a partir de uma população em geral com a opinião dos especialistas na arte de desenhar caricaturas.

# Referências

- AKLEMAN, E. Automatic Creation of Expressive Caricatures: A Grand Challenge for Computer Graphics. In: COMPUTATIONAL Aesthetics in Graphics, Visualization and Imaging. [S.l.: s.n.], 2005. p. 1–2.
- BALTRU, Tadas; ROBINSON, Peter; MORENCY, Louis-Philippe et al. OpenFace: an open source facial behavior analysis toolkit. In: IEEE. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). [S.l.: s.n.], 2016. p. 1–10.
- BALTRUŠAITIS, Tadas; ROBINSON, Peter; MORENCY, Louis-Philippe. 3D constrained local model for rigid and non-rigid facial tracking. In: IEEE. COMPUTER Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. [S.l.: s.n.], 2012. p. 2610–2617.
- BALTRUSAITIS, Tadas et al. Openface 2.0: Facial behavior analysis toolkit. In: IEEE. 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). [S.l.: s.n.], 2018. p. 59–66.
- BEIER, Thaddeus; NEELY, Shawn. Feature-based image metamorphosis. **ACM SIGGRAPH computer graphics**, ACM New York, NY, USA, v. 26, n. 2, p. 35–42, 1992.
- BLANZ, Volker; VETTER, Thomas. Face recognition based on fitting a 3D morphable model. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 25, n. 9, p. 1063–1074, 2003.
- BRENNAN, Susan Elise. **Caricature generator**. 1982. Tese (Doutorado) – Massachusetts Institute of Technology.
- BROWN, Matthew; LOWE, David G. Automatic panoramic image stitching using invariant features. **International journal of computer vision**, Springer, v. 74, n. 1, p. 59–73, 2007.

BULAT, Adrian; TZIMIROPOULOS, Georgios. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In: INTERNATIONAL Conference on Computer Vision. [S.l.: s.n.], 2017.

CAO, Kaidi; LIAO, Jing; YUAN, Lu. Carigans: Unpaired photo-to-caricature translation. **arXiv preprint arXiv:1811.00222**, 2018.

CLARKE, Lyndsey; CHEN, Min; MORA, Benjamin. Automatic generation of 3D caricatures based on artistic deformation styles. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 17, n. 6, p. 808–821, 2011.

COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. J. Computer Vision - 5th European Conference on Computer Vision Freiburg, Germany, June 2–6, 1998 Proceedings, Volume II. In: BURKHARDT, Hans; NEUMANN, Bernd (Ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. Active appearance models, p. 484–498. ISBN 978-3-540-69235-5. DOI: 10.1007/BFb0054760. Disponível em: <<http://dx.doi.org/10.1007/BFb0054760>>.

COOTES, Tim; BALDOCK, ER; GRAHAM, J. An introduction to active shape models. **Image processing and analysis**, Oxford University Press, p. 223–248, 2000.

COOTES, Timothy F; EDWARDS, Gareth J; TAYLOR, Christopher J et al. Comparing active shape models with active appearance models. In: CITESEER. BMVC. [S.l.: s.n.], 1999. p. 173–182.

COOTES, Timothy F; TAYLOR, Christopher J et al. Active shape models-their training and application. **Computer vision and image understanding**, Elsevier, v. 61, n. 1, p. 38–59, 1995.

DAL, Upasna; DAL, D; ABRAHAM, Siby. A facial caricature generation system using adaptive thresholding. In: IEEE. INFORMATION and Communication Technologies (WICT), 2011 World Congress on. [S.l.: s.n.], 2011. p. 682–687.

DAVIS, Malcolm H et al. A physics-based coordinate transformation for 3-D image matching. **IEEE transactions on medical imaging**, IEEE, v. 16, n. 3, p. 317–328, 1997.

- DONG, Xuanyi et al. Style aggregated network for facial landmark detection. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2018. p. 379–388.
- DU, Chunhua et al. SVM based ASM for facial landmarks location. In: IEEE. COMPUTER and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on. [S.l.: s.n.], 2008. p. 321–326.
- DU, Xiaorong et al. A Method of Human Facial Portrait Generation Based on Features Exaggeration. In: TRANSACTIONS on Edutainment XI. [S.l.]: Springer, 2015. p. 90–102.
- FENG, Xiaofei; GE, Yujia. Video Deformation Based on ASM. In: COMPUTER, Informatics, Cybernetics and Applications. [S.l.]: Springer, 2012. p. 1119–1127.
- FROWD, Charlie et al. An application of caricature: how to improve the recognition of facial composites. **Visual Cognition**, Taylor & Francis, v. 15, n. 8, p. 954–984, 2007.
- FUJIWARA, Takayuki et al. 3D modeling system of human face and full 3D facial caricaturing. In: IEEE. 3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on. [S.l.: s.n.], 2001. p. 385–392.
- GHASEMI, Asghar; ZAHEDIASL, Saleh. Normality tests for statistical analysis: a guide for non-statisticians. **International journal of endocrinology and metabolism**, Kowsar Medical Institute, v. 10, n. 2, p. 486, 2012.
- HAHS-VAUGHN, Debbie L; LOMAX, Richard G. **An introduction to statistical concepts**. [S.l.]: Routledge, 2013.
- HAN, Myoung-hun; SONG, Han-na; YOON, Kyung-hyun. Cartooning using reference images. In: IEEE. HUMAN-CENTRIC Computing (HumanCom), 2010 3rd International Conference on. [S.l.: s.n.], 2010. p. 1–5.
- HAN, Xiaoguang; HOU, Kangcheng et al. Caricatureshop: Personalized and photorealistic caricature sketching. **IEEE transactions on visualization and computer graphics**, IEEE, v. 26, n. 7, p. 2349–2361, 2018.
- HILL, Matthew Q et al. Deep convolutional neural networks in the face of caricature. **Nature Machine Intelligence**, Nature Publishing Group, v. 1, n. 11, p. 522–529, 2019.

- HUO, Jing et al. Webcaricature: a benchmark for caricature recognition. **arXiv preprint arXiv:1703.03230**, 2017.
- JESORSKY, Oliver; KIRCHBERG, Klaus J; FRISCHHOLZ, Robert W. Robust face detection using the hausdorff distance. In: SPRINGER. INTERNATIONAL Conference on Audio-and Video-Based Biometric Person Authentication. [S.l.: s.n.], 2001. p. 90–95.
- KASINSKI, Andrzej; FLOREK, Andrzej; SCHMIDT, Adam. The PUT face database. **Image Processing and Communications**, v. 13, n. 3–4, p. 59–64, 2008.
- KAZEMI, Vahid; SULLIVAN, Josephine. One millisecond face alignment with an ensemble of regression trees. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2014. p. 1867–1874.
- KING, Davis E. Dlib-ml: A machine learning toolkit. **The Journal of Machine Learning Research**, JMLR. org, v. 10, p. 1755–1758, 2009.
- LE, Vuong et al. Interactive facial feature localization. In: SPRINGER. EUROPEAN Conference on Computer Vision. [S.l.: s.n.], 2012. p. 679–692.
- LI, Cheng; MIAO, Zhenjiang. A Method for Exaggerative Caricature Generation from Real Face Image. In: IEEE. PATTERN Recognition (ACPR), 2013 2nd IAPR Asian Conference on. [S.l.: s.n.], 2013. p. 828–832.
- LIANG, Lin et al. Example-based caricature generation with exaggeration. In: IEEE. COMPUTER Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on. [S.l.: s.n.], 2002. p. 386–393.
- LIU, Junfa et al. Creative cartoon face synthesis system for mobile entertainment. In: SPRINGER. PACIFIC-RIM Conference on Multimedia. [S.l.: s.n.], 2005. p. 1027–1038.
- MACEDO, Marcio; APOLINÁRIO, Antônio. Improved anti-aliasing for Euclidean distance transform shadow mapping. **Computers & Graphics**, Elsevier, v. 71, p. 166–179, 2018.
- MEILU et al. Robust facial landmark detection via occlusion-adaptive deep networks. In: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. p. 3486–3496.

- MILBORROW, S.; MORTEL, J.; NICOLLS, F. The MUCT Landmarked Face Database. **Pattern Recognition Association of South Africa**, 2010.  
<http://www.milbo.org/muct>.
- MILBORROW, S.; NICOLLS, F. Active Shape Models with SIFT Descriptors and MARS. **VISAPP**, 2014.
- MONDAL, Sharmistha; BISWAS, Arindam. Sharpening of 3D Digital Objects using Triangulation. In: IEEE. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). [S.l.: s.n.], 2020. p. 1–7.
- MONZO, David et al. A Comparative Study of facial landmark localization methods for Face Recognition using HOG descriptors. In: IEEE. PATTERN Recognition (ICPR), 2010 20th International Conference on. [S.l.: s.n.], 2010. p. 1330–1333.
- NIGAM, Swati; SINGH, Rajiv; MISRA, AK. Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. **Multimedia tools and applications**, Springer, v. 77, n. 21, p. 28725–28747, 2018.
- NORDSTRØM, Michael M et al. **The IMM face database-an annotated dataset of 240 face images**. [S.l.], 2004.
- OSTERTAGOVÁ, Eva; OSTERTAG, Oskar; KOVÁČ, Jozef. Methodology and application of the kruskal-wallis test. In: TRANS TECH PUBL. APPLIED Mechanics and Materials. [S.l.: s.n.], 2014. v. 611, p. 115–120.
- RANFTL, Andreas; ALONSO-FERNANDEZ, Fernando; KARLSSON, Stefan. Face tracking using optical flow. In: IEEE. 2015 International Conference of the Biometrics Special Interest Group (BIOSIG). [S.l.: s.n.], 2015. p. 1–5.
- RIBEIRO, Felipe M.L. et al. Detecção de pontos fiduciais sobre a face em tempo real. **XXX SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES**, v. 200, p. 200, 2012.
- RICHMOND, Tom. **The mad art of caricature**. Burnsville - United States: Deadline Demon Publishing, 2011.
- RIEGLER, Gernot et al. Anatomical landmark detection in medical applications driven by synthetic data. In: PROCEEDINGS of the IEEE International Conference on Computer Vision Workshops. [S.l.: s.n.], 2015. p. 12–16.

- SADIMON SURIATI, Bte et al. Computer generated caricature: A survey. In: IEEE. CYBERWORLDS (CW), 2010 International Conference on. [S.l.: s.n.], 2010. p. 383–390.
- SAGONAS, Christos; ANTONAKOS, Epameinondas et al. 300 faces in-the-wild challenge: Database and results. **Image and Vision Computing**, Elsevier, v. 47, p. 3–18, 2016.
- SAGONAS, Christos; TZIMIROPOULOS, Georgios et al. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In: PROCEEDINGS of the IEEE International Conference on Computer Vision Workshops. [S.l.: s.n.], 2013. p. 397–403.
- SHI, Yichun; DEB, Debayan; JAIN, Anil K. Warpgan: Automatic caricature generation. In: PROCEEDINGS of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. p. 10762–10771.
- SHIZHAN et al. Face alignment by coarse-to-fine shape searching. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015. p. 4998–5006.
- SOHAIL, Abu Sayeed Md.; BHATTACHARYA, Prabir. Signal Processing for Image Enhancement and Multimedia Processing. In: DAMIANI, Ernesto et al. (Ed.). Boston, MA: Springer US, 2008. Detection of Facial Feature Points Using Anthropometric Face Model, p. 189–200. ISBN 978-0-387-72500-0. DOI: 10.1007/978-0-387-72500-0\_17. Disponível em: <[http://dx.doi.org/10.1007/978-0-387-72500-0\\_17](http://dx.doi.org/10.1007/978-0-387-72500-0_17)>.
- TULYAKOV, Sergey et al. Mocogan: Decomposing motion and content for video generation. In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2018. p. 1526–1535.
- UŘIČÁŘ, Michal et al. Multi-view facial landmark detector learned by the Structured Output SVM. **Image and Vision Computing**, Elsevier, v. 47, p. 45–59, mar. 2016. ISSN 0262-8856.
- VALSTAR, Michel et al. Facial point detection using boosted regression and graph models. In: IEEE. COMPUTER Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. [S.l.: s.n.], 2010. p. 2729–2736.
- WANG, Ching-Wei; HUANG, Cheng-Ta et al. Evaluation and comparison of anatomical landmark detection methods for cephalometric X-ray images: a grand challenge. **IEEE transactions on medical imaging**, IEEE, v. 34, n. 9, p. 1890–1900, 2015.

- WANG, Nannan; LI, Jie et al. Heterogeneous image transformation. **Pattern Recognition Letters**, Elsevier, v. 34, n. 1, p. 77–84, 2013.
- YANG, Fei et al. Facial expression editing in video using a temporally-smooth factorization. In: IEEE. 2012 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2012. p. 861–868.
- ZHANG, Zhanpeng et al. Facial landmark detection by deep multi-task learning. In: SPRINGER. EUROPEAN Conference on Computer Vision. [S.l.: s.n.], 2014. p. 94–108.
- ZHAO, Long et al. Cartoonish sketch-based face editing in videos using identity deformation transfer. **Computers & Graphics**, Elsevier, v. 79, p. 58–68, 2019.
- ZOPE, Bhushan; ZOPE, Soniya B. A Survey of Morphing Techniques. **International Journal of Advanced Engineering, Management and Science**, Infogain Publication, v. 3, n. 2, p. 239773, 2017.

## **Apêndice A**

# **Questionário de Delineamento do Perfil do Usuário**

## Questionário para avaliação de uma abordagem para geração automática de caricaturas

\*Obrigatório

### 1. E-mail \*

Uma Abordagem para a  
Geração Automática de  
Caricaturas em Vídeos Digitais

Pesquisador responsável: Vinicius Fernandes de Sousa  
Instituição/Departamento: UFCG/DSC/CEEI  
Telefone e endereço postal completo: (83) 3341 - 6810 - Rua Aprígio Veloso, 882, Bairro  
Universitário, Campina Grande - PB - CEP: 58429-900  
Email para contato: [viniciusfernandes@copin.ufcg.edu.br](mailto:viniciusfernandes@copin.ufcg.edu.br)

Eu, Vinicius Fernandes de Sousa, responsável pela pesquisa Uma Abordagem para a Geração Automática de Caricaturas em Vídeos Digitais, o convidamos a participar como voluntário deste nosso estudo.

Durante todo o período da pesquisa você terá a possibilidade de não aceitar participar ou de retirar sua permissão a qualquer momento, sem nenhum tipo de prejuízo pela sua decisão.

Em caso de algum problema relacionado com a pesquisa, você terá direito à assistência gratuita que será prestada pelo pesquisador mediante a explicação detalhada dos passos a serem realizados durante a pesquisa.

As informações desta pesquisa poderão divulgadas em eventos ou publicações.

2. Após a leitura deste documento, estou suficientemente informado, ficando claro que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Diante do exposto e de espontânea vontade, expresso minha concordância em participar deste estudo. \*

*Marque todas que se aplicam.*

Autorizo a utilização dos dados

Delineamento do perfil  
do Usuário.

A seguir, você deverá responder a uma série de questões relacionadas a seu perfil. LEIA COM ATENÇÃO TODOS OS ENUNCIADOS E OPÇÕES.

### 3. Qual seu nível de escolaridade? \*

*Marcar apenas uma oval.*

- Ensino Médio Incompleto  
 Ensino Médio Completo  
 Superior Incompleto  
 Superior Completo  
 Pós-graduação Incompleta  
 Pós-Graduação Completa

4. Qual seu gênero? \*

Marcar apenas uma oval.

- Masculino  
 Feminino  
 Outro

5. Você utiliza corretivos visuais (e.g., óculos ou lente de contato)? \*

Marcar apenas uma oval.

- Sim  
 Não

6. Você costuma observar caricaturas nos meios de comunicação que você acessa? (ex.: Jornais, Sites, Aplicativos) \*

Marcar apenas uma oval.

- Sim  
 Não

7. Com que frequência você observa essas caricaturas ao consumir os meios de comunicações? Para a escala de 5 pontos a seguir, considere as seguintes âncoras semânticas: 1 - Muito Baixa; 2 - Baixa; 3 - Nem Baixa nem Alta; 4 - Alta; 5 - Muito Alta \*

Marcar apenas uma oval.

	1	2	3	4	5	
Muito Baixa	<input type="radio"/>	Muito Alta				

8. Como você julga sua facilidade de identificar uma caricatura. Para a escala de 5 pontos a seguir, considere as seguintes âncoras semânticas: 1 - Muito difícil; 2 - Difícil; 3 - Nem Difícil nem Fácil; 4 - Fácil; 5 - Muito Fácil \*

Marcar apenas uma oval.

	1	2	3	4	5	
Muito difícil	<input type="radio"/>	Muito Fácil				

## **Apêndice B**

# **Questionário de Sondagem da Opinião do Usuário sobre os resultados de uma Abordagem para Geração Automática de Caricaturas**

## Questionário para avaliação de uma abordagem para geração automática de caricaturas

\*Obrigatório

Uma Abordagem para a Geração Automática de Caricaturas em Vídeos Digitais

Pesquisador responsável: Vinicius Fernandes de Sousa  
 Instituição/Departamento: UFCG/DSC/CEEI  
 Telefone e endereço postal completo: (83) 3341 - 6810 - Rua Aprígio Veloso, 882, Bairro Universitário, Campina Grande - PB - CEP: 58429-900  
 Email para contato: [viniciusfernandes@copin.ufcg.edu.br](mailto:viniciusfernandes@copin.ufcg.edu.br)

Eu, Vinicius Fernandes de Sousa, responsável pela pesquisa Uma Abordagem para a Geração Automática de Caricaturas em Vídeos Digitais, o convidamos a participar como voluntário deste nosso estudo.

Durante todo o período da pesquisa você terá a possibilidade de não aceitar participar ou de retirar sua permissão a qualquer momento, sem nenhum tipo de prejuízo pela sua decisão.

Em caso de algum problema relacionado com a pesquisa, você terá direito à assistência gratuita que será prestada pelo pesquisador mediante a explicação detalhada dos passos a serem realizados durante a pesquisa.

As informações desta pesquisa poderão divulgadas em eventos ou publicações.

1. Após a leitura deste documento, estou suficientemente informado, ficando claro que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Diante do exposto e de espontânea vontade, expresso minha concordância em participar deste estudo. \*

Marque todas que se aplicam.

Autorizo a utilização dos dados

Questionário de Sondagem da Opinião

Os pares de vídeo a seguir obedecem sempre a mesma ordem: o vídeo da esquerda é o vídeo original, enquanto o vídeo da direita é o vídeo que contém um grau de distorção facial. Cada questão contém uma escala de cinco pontos. Selecione na escala a opção que melhor reflete sua opinião, após visualizar cada par de vídeos em tela cheia (clique no título do vídeo e, em seguida, dê um duplo clique no vídeo apresentado). LEIA COM ATENÇÃO TODOS OS ENUNCIADOS E OPÇÕES.

ATENTE TAMBÉM PARA AS ÂNCORAS SEMÂNTICAS DAS ESCALAS DE 5 PONTOS

Para a escala DISCORDO TOTALMENTE a CONCORDO TOTALMENTE as âncoras são:

- 1 - Discordo totalmente
- 2 - Discordo parcialmente
- 3 - Não concordo nem discordo
- 4 - Concordo Parcialmente
- 5 - Concordo Totalmente

Para a escala MUITO DIFÍCIL a MUITO FÁCIL as âncoras são:

- 1 - Muito difícil
- 2 - Difícil
- 3 - Nem difícil nem fácil
- 4 - Fácil
- 5 - Muito fácil

Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=ugA3OoOWrCQ>

2. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

3. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

4. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=Wl5QTKwEhow>





13. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=nFSL0Pt0u-g>

14. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

15. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

16. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



[http://youtube.com/watch?v=\\_mlKrezWvT4](http://youtube.com/watch?v=_mlKrezWvT4)

17. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

18. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

19. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=o409p-7Snlk>





28. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=TXa5iDLm6yE>

29. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

30. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

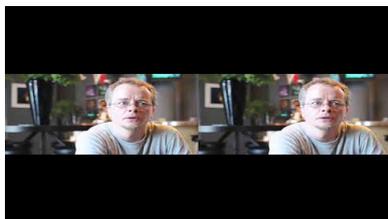
31. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

*Marcar apenas uma oval.*

	1	2	3	4	5	
Discordo Totalmente	<input type="radio"/>	Concordo Totalmente				

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=cLcWxLAU2Q>

32. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

*Marcar apenas uma oval.*

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

33. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

*Marcar apenas uma oval.*

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

34. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

*Marcar apenas uma oval.*

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=cLcWxLAU2Q>





43. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=JT8IQEnaKlg>

44. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

Marcar apenas uma oval.

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

45. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

Marcar apenas uma oval.

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

46. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

1    2    3    4    5

---

Discordo Totalmente      Concordo Totalmente

#### Questionário de Sondagem da Opinião

Vídeo a ser analisado



<http://youtube.com/watch?v=JT8lQEnaKlg>

47. Com relação ao vídeo original, a face se assemelha à uma face caricaturada (e.g., contém exageros nas características faciais que mais chamam a atenção).

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

48. O vídeo a seguir apresenta artefatos visuais indesejáveis (e.g., deformações fora da região facial, deformações incômodas visualmente).

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

49. O vídeo a seguir apresenta uma deformação que se mantém estável ao longo dos quadros dos vídeos (e.g., deformação facial se mantém coerente durante todo o vídeo)

Marcar apenas uma oval.

1 2 3 4 5

Discordo Totalmente      Concordo Totalmente

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

# Apêndice C

## E-mail de Recrutamento dos Usuários

Bom dia, caros colegas!

Sou mestrando em Ciência da Computação, sob a orientação dos professores Eustáquio Rangel e Herman Gomes. Minha pesquisa envolve a geração automática de deformações de faces em vídeos (ao estilo de caricaturas). Gostaria de poder contar com sua participação em um divertido experimento on-line, cujo propósito é avaliar os resultados de diferentes algoritmos.

No experimento, você será solicitado a assistir e julgar sequências de vídeos no tocante a um conjunto de critérios. A duração do experimento será de aproximadamente 40 minutos.

Caso tenha interesse em participar do experimento, peço-lhes a gentileza de preencher o questionário acessível a partir deste link.

Oportunamente, entrarei em contato com cada um dos interessados, a fim de agendar um horário a sua conveniência para a condução do experimento. Ao final do experimento, os participantes que desejarem, poderão ter um vídeo pessoal caricaturizado pelo algoritmo desenvolvido na pesquisa.

Agradeço-lhes antecipadamente por toda a colaboração!

## **Apêndice D**

### **Exemplos dos Quadros Deformados Utilizados no Experimento com a Avaliação com Usuários**

Figura D.1 Exemplos de Quadros Utilizando o Processamento 1 - Processamento sem a utilização do algoritmo de adaptativo de detecção dos pontos fiduciais e nível 3 de deformação da imagem. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autorial Própria

Figura D.1 Exemplos de Quadros Utilizando o Processamento 1 - Processamento sem a utilização do algoritmo de adaptativo de detecção dos pontos fiduciais e nível 3 de deformação da imagem. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.2 Exemplos de Quadros Utilizando o Processamento 2 - Processamento com a utilização do algoritmo adaptativo e nível 3 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.2 Exemplos de Quadros Utilizando o Processamento 2 - Processamento com a utilização do algoritmo adaptativo e nível 3 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.3 Exemplos de Quadros Utilizando o Processamento 3 - Processamento com a utilização do algoritmo adaptativo e nível 2 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.3 Exemplos de Quadros Utilizando o Processamento 3 - Processamento com a utilização do algoritmo adaptativo e nível 2 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.4 Exemplos de Quadros Utilizando o Processamento 4 - Processamento com a utilização do algoritmo adaptativo e nível 1 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria

Figura D.4 Exemplos de Quadros Utilizando o Processamento 4 - Processamento com a utilização do algoritmo adaptativo e nível 1 de deformação. A esquerda são apresentados os quadros originais e a direita o resultado do processamento.



Fonte: Autoria Própria