LISETH URPY SEGUNDO CARPIO

**Geração de caricaturas para representação de emoções usando
processamento de imagens faciais e grafos And-Or**

Versão Corrigida

São Paulo
2016

LISETH URPY SEGUNDO CARPIO

**Geração de caricaturas para representação de emoções usando processamento de
imagens faciais e grafos And-Or**

Versão corrigida

Dissertação apresentada à Escola de Artes,
Ciências e Humanidades da Universidade de
São Paulo para obtenção do título de Mestre em
Ciências pelo Programa de Pós-graduação em
Sistemas de Informação.

Área de Concentração:
Sistemas de Informação

Orientadora:
Prof. Dra. Arianne Machado Lima

Coorientadora:
Prof. Dra. Fátima de Lourdes dos Santos Nunes

São Paulo

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)

Carpio, Liseth Urpy Segundo

Geração de caricaturas para representação de emoções usando processamento de imagens faciais e grafos And-Or / Liseth Urpy Segundo Carpio ; orientadora, Ariane Machado Lima ; coorientadora, Fátima de Lourdes dos Santos Nunes Marques. – São Paulo, 2016
128 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Sistemas de Informação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, em 2015
Versão corrigida

1. Reconhecimento de padrões. 2. Processamento de imagens. 3. Reconhecimento de imagem. 4. Emoções. 5. Percepção de face. I. Machado-Lima, Ariane, orient. II. Marques, Fátima de Lourdes dos Santos Nunes, coorient. III. Título

CDD 22.ed. – 006.42

SEGUNDO CARPIO, Liseth Urpy

Geração de caricaturas para representação de emoções usando processamento de imagens faciais e grafos And-Or

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências do Programa de Pós-Graduação em Sistemas de Informação.

Área de Concentração:

Sistemas de Informação

Aprovado em: 17/ 12/ 2015

Banca Examinadora

Prof. Dr. Hélio Pedrini

Instituição: IC - UNICAMP

Julgamento: Aprovação

Prof. Dr. Luciano A. Pedrini

Instituição: EACH - USP

Julgamento: Aprovação

Prof. Dr. Ronaldo F. Hashimoto

Instituição: IME - USP

Julgamento: Aprovação

Dedico este trabalho a toda minha família.

Agradecimentos

Em primeiro lugar, gostaria de agradecer a Deus, por me amparar nos momentos difíceis, me dar força interior para superar as dificuldades, mostrar o caminho nas horas incertas e me suprir em todas minhas necessidades.

Um agradecimento especial a minha orientadora e amiga, Profa. Dra. Arianne Machado Lima, pela confiança, pela oportunidade de trabalhar ao seu lado, pela disposição em discutir dúvidas e certezas, pelo tempo gasto nas incansáveis correções de textos e principalmente pelo apoio nos momentos mais difíceis no mestrado.

À Profa. Dra. Fátima de Lourdes dos Santos Nunes Marques, minha coorientadora, pela infinita disponibilidade, por todos os ensinamentos e pelo incentivo na superação de meus limites.

Agradecimentos aos meus pais, dona Laura e don Roberto (meu anjo, que falta você me faz!), por me ter dado educação, valores, amor e apoio incondicional. A vocês que, muitas vezes, renunciaram a seus sonhos para que eu pudesse realizar o meu.

Aos irmãos que Deus colocou em minha vida, Vanessa e Sihuar. Amor incondicional, sempre. A distância não nos separa. Seus corações estão comigo e o meu com vocês. Amo vocês!

Aos amigos que fizeram parte desses momentos sempre me ajudando e incentivando.

Aos alunos Rafael Testa, Daniel Carvalho e Arthur Moraes pelo trabalho em conjunto.

Aos professores, funcionários e colegas do Programa de Pós-Graduação em Sistemas de Informação da EACH-USP, pelas aulas que estarão sempre em minha memória e pela solicitude e solidariedade perante minhas dificuldades.

*“Sua vida pode ser dividida em dois períodos: antes de agora e a partir de
agora.”*

(Prof. Obvious Stating)

RESUMO

SEGUNDO CARPIO, L. U. **Geração de caricaturas para representação de emoções usando processamento de imagens faciais e grafos And-Or**. 2016. 121 p. Dissertação (Mestrado em Sistemas de Informação) - Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2015. Versão corrigida.

Emoções faciais desempenham um papel fundamental nas relações interpessoais humanas. Certas doenças psiquiátricas podem se caracterizar por déficits no reconhecimento das emoções. Nesse contexto, o presente projeto faz parte de um projeto maior que visa a desenvolver um jogo sério para auxílio ao diagnóstico e treinamento no reconhecimento de expressões faciais utilizando grafos And-Or para representação de emoções. Este projeto de mestrado define, implementa e avalia uma abordagem que gera caricaturas de várias emoções faciais (satisfação, tristeza, surpresa, medo, aversão, raiva e neutra) a partir de imagens de treinamento. Foi desenvolvida uma metodologia de processamento de imagens que localiza pontos de controle faciais em imagens reais de pessoas expressando as diferentes emoções, e aprende as variações de todos os pontos de controle, da face neutra para uma determinada emoção, para caracterização das expressões faciais. Para poder comparar a fidelidade da representação da emoção desejada nas imagens geradas, com respeito a uma abordagem anterior (que utiliza antropometria e FACS - Facial Action Coding System), um experimento com voluntários foi conduzido. Os resultados mostraram que a taxa de reconhecimento das imagens geradas com a metodologia proposta neste projeto, em relação às imagens geradas com a metodologia anterior, foi inferior em cinco emoções (medo, raiva, satisfação, surpresa e tristeza), igual para uma emoção (neutra) e superior para uma emoção (aversão). Para a emoção de aversão, cuja taxa de reconhecimento foi superior na abordagem proposta neste projeto, conseguiu-se aprender variações significativas de pontos de controle que não estão presentes nas unidades de ação que o FACS prevê para essa emoção. Esses dados sugerem que a proposta de aprender as variações de todos os pontos de controle da face pode ajudar em representar melhor as emoções faciais. Foram identificadas as possíveis causas para as taxas de reconhecimento inferiores das cinco emoções citadas, tanto na seleção da amostra de treinamento quanto nas etapas de processamento de imagens, as quais podem ser aperfeiçoadas a fim de melhorar a geração das caricaturas. Portanto, os resultados indicam que a abordagem apresentada é promissora, uma vez que considera todos os pontos de controle identificados em imagens reais e que o

aperfeiçoamento de algumas fases do processo pode gerar caricaturas mais próximas das expressões reais, considerando diferentes intensidades das emoções.

Palavras-chave: Processamento de imagens. Emoções Faciais. Grafos And-Or.

ABSTRACT

SEGUNDO CARPIO, L. U. **Generation of cartoons for representing emotions using facial images processing and And-Or graph.** 2016. 121 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2015. Corrected version.

Facial emotions play a key role in human interpersonal relationships. Some psychiatric disorders may be characterized by deficits in recognizing emotions. In this context, this project is part of a major project that aims to develop a serious game to aid diagnosis and training in the recognition of facial expressions using And-Or graphs to emotions representing. This master's project defines, implements and evaluates an approach for generating caricatures of facial emotions (satisfaction, sadness, surprise, fear, disgust, anger, neutral) from training images. An image processing method that locates facial points in facial expression images was developed, and learns the variations of all facial points, from neutral face to a particular emotion for facial expression characterization. In order to compare the fidelity representation of the desired emotion in images generated with respect to anterior approach (using Anthropometry and Facial Action Coding System FACS), an experiment with volunteers was conducted. The rate recognition results showed that images generated with the proposed approach, compared to previous approach, were lower in five emotions (fear, anger, satisfaction, surprise and sadness), same to an emotion (neutral) and higher to an emotion (disgust). For disgust emotion, it was possible to learn significant variations of facial points that are not present in the action units that FACS provides for that emotion. This data suggests that the proposal for learning the variations of all facial points may help the facial emotion characterization. Possible causes for the lower rates were identified, both in the selection of training sample images and steps of image processing, which can be optimized in order to enhance the generation of caricatures. Therefore, the results indicate this approach is promising, since considers all facial points, and the improvement of some stages of the process can generate closest caricatures of real expressions, considering different intensities of emotions.

Keywords: Image processing. Facial emotions. And-Or graphs.

Lista de figuras

Figura 1.1 – Imagens estáticas geradas pela metodologia 1.	22
Figura 1.2 – A curva em preto corresponde à parte de um componente da face e em verde seus correspondentes pontos de controle. O retângulo azul interno define uma região que envolve a curva. O retângulo vermelho delimita um envoltório que é o dobro do tamanho dos lados do retângulo azul.	23
Figura 2.1 – Um exemplo de um grafo <i>And-Or</i> para a categoria do objeto relógio. Ele tem os dois grafos sintáticos apresentados na Figura 2.2, um dos quais está ilustrado com setas escuras.	29
Figura 2.2 – Dois exemplos de grafo sintático de imagens de relógios, gerados a partir do grafo <i>And-Or</i> da Figura 2.1.	29
Figura 2.3 – Exemplo de análise sintática.	30
Figura 2.4 – (a) Imagem original em tons de cinza. (b) Exemplo de histograma de uma imagem.	32
Figura 2.5 – Passos fundamentais para executar uma tarefa de processamento de imagens.	34
Figura 2.6 – Histograma de níveis de cinza que podem ser particionados por (a) um limiar único e (b) limiares múltiplos.	35
Figura 2.7 – (a) <i>Landmarks</i> antropométricos faciais. (b) Medições antropométricas faciais.	37
Figura 2.8 – Pontos de controle definidos na metodologia 1 para representar uma face inicial neutra.	38
Figura 2.9 – Exemplos de unidades de ação: (a) unidade de ação 1, (b) unidade de ação 2.	39
Figura 2.10 – Representação parcial do grafo <i>And-Or</i>	41
Figura 3.1 – Processo de seleção dos estudos primários.	43
Figura 3.2 – Quantidade de artigos publicados.	44
Figura 3.3 – Tendências de uso de grafos <i>And-Or</i>	44
Figura 3.4 – Quantidade de artigos publicados por grupo de pesquisa.	45

Figura 4.1 – Metodologia de processamento de imagens para localizar pontos de controle faciais e identificar variações desses pontos para representar emoções faciais a partir de imagens de treinamento.	54
Figura 4.2 – Processo de localização de pupilas: (a) localização da face em uma imagem original; (b) localização da região dos olhos; (c) região dos olhos; (d) imagem binária obtida após aplicação do algoritmo <i>Canny</i> ; (e) localização das pupilas.	57
Figura 4.3 – Representação das coordenadas das pupilas.	57
Figura 4.4 – Exemplos da aplicação de rotação nas imagens: (a) imagens originais; (b) imagens após rotação.	58
Figura 4.5 – Processo de ajuste de escala das imagens de uma mesma pessoa: (a) localização das faces; (b) cálculo das distâncias entre as pupilas; (c) ajuste de escala das imagens com respeito à distância mínima entre as pupilas.	59
Figura 4.6 – (a) Localização das áreas da face. (b) Recorte das áreas das faces. . . .	60
Figura 4.7 – (a) Localização dos Componentes Faciais. (b) Estimativa da área das sobrancelhas a partir da área dos olhos.	61
Figura 4.8 – Sequência de técnicas de processamento de imagens aplicada à região do olho direito: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.	63
Figura 4.9 – Sequência de técnicas de processamento de imagens aplicada à região da sobrancelha: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.	63

Figura 4.10—Sequência de técnicas de processamento de imagens aplicada à região do nariz: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.	64
Figura 4.11—Localização dos pontos de controle na imagem binária do nariz.	66
Figura 4.12—Exemplos de localização dos pontos de controle da metodologia 2.	66
Figura 4.13—(a) Envoltório da curva superior do olho direito, tendo essa curva sido estimada a partir dos pontos de controle que a compõem, a saber <i>ex</i> , <i>ps</i> e <i>en</i> . (b) Envoltório da curva inferior do olho direito, tendo essa curva sido estimada a partir dos pontos de controle que a compõem, a saber <i>ex</i> , <i>pi</i> e <i>en</i>	67
Figura 5.1 — Expressões faciais de satisfação geradas por ambas as metodologias para as três intensidades consideradas.	70
Figura 5.2 — Expressões faciais de tristeza geradas por ambas as metodologias para as três intensidades consideradas.	71
Figura 5.3 — Expressões faciais de medo geradas por ambas as metodologias para as três intensidades consideradas.	72
Figura 5.4 — Expressões faciais de surpresa geradas por ambas as metodologias para as três intensidades consideradas.	73
Figura 5.5 — Expressões faciais de raiva geradas por ambas as metodologias para as três intensidades consideradas.	74
Figura 5.6 — Expressões faciais de aversão geradas por ambas as metodologias para as três intensidades consideradas.	75
Figura 5.7 — Expressão facial neutra.	76
Figura 5.8 — Taxas de acerto das emoções em cada metodologia.	77
Figura 5.9 — Exemplos de faces sutis representando a emoção satisfação no banco de dados.	78
Figura 5.10—Taxas de acerto para cada emoção da metodologia 1 (linhas retas) e da metodologia 2 (linhas tracejadas).	80

Figura 5.11—(a) Estimação da área das sobrancelhas (retângulo verde) a partir da localização da área dos olhos (retângulo amarelo). (b) Imagem binária da sobrancelha direita. (c) Localização dos pontos de controle da face. .	82
Figura 5.12—(a) Imagem segmentada (binária) da sobrancelha direita. (b) Imagem segmentada (binária) da sobrancelha esquerda. (c) Localização dos pontos de controle da face.	82
Figura 5.13—Exemplos de caricaturas geradas individualmente pela metodologia 2. Na coluna esquerda encontra-se as imagens originais dos sujeitos para emoção satisfação. Na coluna direita encontra-se sua correspondente caricatura gerada pela metodologia 2.	86
Figura D.1—Estrutura do grafo <i>And-Or</i>	108

Lista de tabelas

Tabela 1 – Metodologia 1: matriz de confusão das respostas obtidas. Em negrito estão as taxas de acerto.	76
Tabela 2 – Metodologia 2: matriz de confusão das respostas obtidas. Em negrito estão as taxas de acerto.	77
Tabela 3 – Percentagem de indivíduos que atingiram determinada faixa de coerência referente às imagens duplicadas em ambas metodologias.	81
Tabela 4 – Taxa de acerto para cada gênero da metodologia 1.	84
Tabela 5 – Taxa de acerto para cada gênero da metodologia 2.	84
Tabela 6 – Taxa de acerto para cada faixa etária da metodologia 1.	85
Tabela 7 – Taxa de acerto para cada faixa etária da metodologia 2.	85
Tabela 8 – Descrições da aparência de cada UA.	96
Tabela 8 – Descrições da aparência de cada UA.	97
Tabela 8 – Descrições da aparência de cada UA.	98
Tabela 9 – Inferências feitas a partir da antropometria.	99
Tabela 9 – Inferências feitas a partir da antropometria.	100
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	101
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	102
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	103
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	104
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	105
Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.	106
Tabela 11 – Valores das variações dos pontos de controle obtidos neste projeto. . .	111

Lista de abreviaturas e siglas

ACM	<i>Association for Computing Machinery</i>
CIE	<i>Commission Internationale Eclairage</i>
FACS	<i>Facial Action Coding System</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
JPEG	<i>Joint Photographic Experts Group</i>
LApIS	Laboratório de Aplicações de Informática em Saúde
MCTS	<i>Monte Carlo Tree Search</i>
OpenCV	<i>Open Source Computer Vision Library</i>
RGB	Espaço de cores <i>Red Green Blue</i>
ST-AOG	Grafo <i>And-OR</i> espaço-temporal
UA	Unidade de Ação

Lista de símbolos

Σ	Alfabeto
θ	Ângulo de inclinação das pupilas
ε	Cadeia vazia
α	Deslocamento máximo de um ponto de controle
σ	Desvio padrão do filtro Gaussiano
I	Intensidade
L	Linguagem
$G_{\text{And-Or}}$	Grafo <i>And-Or</i>
G	Gramática
D_A	Tamanho do envoltório
$\oplus$	Transformação morfológica dilatação
$\ominus$	Transformação morfológica erosão
$\circ$	Transformação morfológica abertura
$\bullet$	Transformação morfológica fechamento

Sumário

1	INTRODUÇÃO	20
1.1	Versão inicial do projeto geral	21
1.2	Hipótese	23
1.3	Proposta deste projeto	24
1.4	Organização do documento	24
2	FUNDAMENTAÇÃO TEÓRICA	26
2.1	Gramáticas	26
2.2	Processamento de imagens	30
2.3	Geração de imagens	36
2.4	Uso do grafo <i>And-Or</i> para representação das faces	40
2.5	Trabalhos correlatos de localização de pontos de controle	41
3	REVISÃO SISTEMÁTICA	42
3.1	Planejamento e condução	42
3.2	Resultados	42
3.2.1	Análise global	42
3.2.2	Geração de imagens	45
3.2.3	Algoritmos de reconhecimento	47
3.2.4	Algoritmos de aprendizado de estruturas	47
3.2.5	Aplicações específicas	48
3.2.5.1	Representação de Objetos Articulados	48
3.2.5.2	Categorização	50
3.2.5.3	Temporalidade	51
3.3	Discussões	53
4	METODOLOGIA	54
4.1	Bancos de imagens	54
4.2	Pré processamento	56
4.3	Segmentação de imagens	58

4.3.1	Detecção da face	59
4.3.2	Localização dos componentes faciais	60
4.3.3	Segmentação dos componentes faciais	61
4.4	Cálculo dos pontos de controle e de suas variações entre diferentes emoções	62
4.4.1	Localização dos pontos de controle	62
4.4.2	Identificação das variações dos pontos de controle	67
4.5	Comparação entre as duas metodologias	68
5	RESULTADOS E DISCUSSÕES	69
5.1	Reconhecimento das caricaturas geradas	69
5.2	Coerência das respostas	80
5.3	Avaliação da metodologia de processamento de imagens	81
5.4	Considerações Finais	84
6	CONCLUSÕES	87
6.1	Principais resultados obtidos e contribuições	87
6.2	Trabalhos futuros	89
	Referências	91
	APÊNDICE A – DESCRIÇÕES DA APARÊNCIA DE CADA UNIDADE DE AÇÃO	96
	APÊNDICE B – DESCRIÇÃO DE CADA PONTO DE CONTROLE	99
	APÊNDICE C – POSICIONAMENTO DOS PONTOS DE CONTROLE NA ABORDAGEM INICIAL	101
	APÊNDICE D – ESTRUTURA DO GRAFO AND-OR	107
	APÊNDICE E – VALORES DAS VARIAÇÕES DOS PONTOS DE CONTROLE OBTIDOS NESTE PROJETO	109

APÊNDICE F – UTILIZAÇÃO DE GRAFOS <i>AND-OR</i> PARA RE- CONHECIMENTO DE PADRÕES EM IMAGENS E GERAÇÃO DE IMAGENS: PROTOCOLO DE REVISÃO	112
APÊNDICE G – AQUISIÇÃO DE IMAGENS	115
ANEXO A – APROVAÇÃO DO COMITÊ DE ÉTICA DO PRO- JETO DE FORMAÇÃO DE UM BANCO DE DADOS DE EXPRESSÕES FACIAIS	116
ANEXO B – TERMO DE CONSENTIMENTO LIVRE E ESCLA- RECIDO	118

1 Introdução

As emoções são respostas instintivas, associadas às diversas situações da vida. Elas acontecem em todos os relacionamentos que nos interessam: no trabalho, em nossas amizades, nas interações familiares e em relacionamentos de qualquer natureza (EKMAN; FRIESEN, 1971). A comunicação de uma emoção pode ser estabelecida por meio verbal ou não verbal. As manifestações não verbais podem revelar-se por meio de expressões faciais, gestos, posturas corporais, dilatação da pupila, etc. Pesquisas em psicologia social (BOYLE; ANDERSON; NEWLANDS, 1994; LEVENSON et al., 1991) sugerem que as expressões faciais são a maior modalidade na comunicação humana não verbal e desempenham um papel muito importante na interação interpessoal humana e na adaptação do homem na sociedade (CALDER et al., 2000).

Doenças psiquiátricas distintas podem se caracterizar por déficits diferentes em aspectos específicos do processamento emocional. Por exemplo, crianças autistas possuem menor capacidade de discriminar emoções representadas por meio das expressões faciais do que crianças sem patologia psiquiátrica (NEWSCHAFER; CURRAN, 2003).

Técnicas computacionais envolvendo processamento de imagens, computação gráfica, reconhecimento de padrões e realidade virtual têm se mostrado aliadas na construção de ferramentas para diagnóstico, avaliação e treinamento no reconhecimento de emoções faciais (LAHIRI et al., 2013; CHENG; LING, 2008; GRYNSZPAN; MARTIN; NADEL, 2008). Estudos que visam a gerar imagens virtuais que representem emoções faciais fazem uso de motores de jogo ou pacotes prontos de realidade virtual (BEKELE et al., 2014; DYCK et al., 2008; KU et al., 2005; CHENG; LING, 2008; LAHIRI et al., 2013). No entanto, com respeito à flexibilidade, a maioria desses recursos é limitada em gerar uma variedade adequada nas intensidades das emoções e características dos avatares¹ como gênero, etnia e idade. Seria desejável que a intensidade da emoção e as características do avatar fossem parâmetros do jogo, de forma a serem utilizados para graduar o nível de dificuldade do jogo de forma separada do roteiro do mesmo.

Nesse contexto, a presente proposta faz parte de um projeto de pesquisa mais abrangente que visa a desenvolver um jogo sério de reconhecimento de expressões faciais,

¹ No contexto de jogos, um avatar é a manifestação corporal de alguém no espaço cibernético.

com o propósito de auxiliar no diagnóstico e na terapia de transtornos psiquiátricos, permitindo tal parametrização do jogo, em tempo de execução, conferindo flexibilidade para que terapeutas atendam às necessidades individuais de cada paciente. A geração das imagens será a partir de gramáticas, mais especificamente grafos *And-Or*, que devem caracterizar as várias emoções faciais, suas diferentes intensidades e ainda futuramente grupo étnico, gênero e faixa etária. De fato, gramáticas em imagens têm se mostrado aliadas na geração de imagens sintéticas. Trabalhos como (CHEN et al., 2006), (XU et al., 2008) e (SUO et al., 2010) apresentam o uso de diferentes tipos de gramática para gerar imagens de roupas, representar e desenhar faces, e gerar imagens que simulam o envelhecimento da face, respectivamente.

A seguir são apresentados mais detalhes para entender o escopo abordado no presente documento.

1.1 Versão inicial do projeto geral

Em uma versão inicial do jogo, a qual chamaremos de metodologia 1, são geradas imagens estáticas, caricaturas de várias expressões faciais (satisfação, tristeza, raiva, aversão, medo, surpresa e medo) com diferentes graus de intensidade (Figura 1.1). A forma de geração dessas caricaturas foi proposta durante o trabalho de iniciação científica de Rafael Luiz Testa (TESTA et al., 2015), e é descrita brevemente nesta seção e com mais detalhes na seção 2.3.

As caricaturas são construídas considerando que cada componente da face (olho, boca, etc.) é composto por curvas *splines*. Uma curva *spline* é definida matematicamente por dois ou mais pontos de controle (FOLEY, 1996). A definição, modificação e manipulação dessas curvas são realizadas por meio de operações em seus pontos de controle.

Uma face inicial neutra é desenhada utilizando medidas antropométricas (FARKAS, 1994) para definir a dimensão de cada curva e o posicionamento correto dos seus pontos de controle.

Para contemplar a tarefa de reproduzir as demais expressões faciais, é necessário simular os movimentos provocados pelas emoções na face, e aplicar esse efeito nas curvas que compõem a face. No contexto deste projeto, esta tarefa usa o conceito de Unidades de Ação (UAs), definidas por um sistema de codificação facial denominado FACS (*Facial Action Coding System*) (EKMAN; FRIESEN, 1976). As UAs são compostas por um músculo

Figura 1.1 – Imagens estáticas geradas pela metodologia 1.

Fonte: (TESTA et al., 2015)

ou um conjunto de músculos que se movimentam quando uma emoção é expressa pela face humana. Uma emoção pode movimentar uma ou mais UAs. O sistema FACS define quais e como essas UAs se movimentam em função de cada emoção expressa. No projeto em questão, os movimentos das UAs são relacionados com os deslocamentos das curvas que formam as caricaturas.

Apesar da ampla descrição de cada unidade de ação, as ciências comportamentais não estão preocupadas em medir exatamente com qual intensidade se varia um músculo em uma dada unidade de ação. Há apenas uma descrição de qual músculo se move em qual direção e a implicação dessa ação nos componentes da face.

Para contornar esse problema, foi preciso encontrar uma maneira alternativa de realizar as variações nos pontos de controle das curvas *splines* sem gerar inconsistências. Para isso, definiu-se um envoltório como o dobro do tamanho dos lados do retângulo que envolve a curva, como ilustrado na Figura 1.2. Assim, o reposicionamento de um ponto consiste em variar um percentual da medida do envoltório correspondente. Nesta versão inicial do projeto geral, esse percentual foi definido empiricamente (Apêndice C).

Figura 1.2 – A curva em preto corresponde à parte de um componente da face e em verde seus correspondentes pontos de controle. O retângulo azul interno define uma região que envolve a curva. O retângulo vermelho delimita um envoltório que é o dobro do tamanho dos lados do retângulo azul.

Fonte: (TESTA et al., 2015)

A qualidade da representação das emoções nas imagens geradas foi avaliada por voluntários supostamente saudáveis (sem nenhum tipo de transtorno psiquiátrico conhecido). Verificou-se que, exceto para a face neutra e a emoção de aversão, a taxa de acerto aumentou à medida que aumentou a intensidade das emoções. As imagens da emoção de medo foram as que apresentaram os menores níveis de reconhecimento em todas as intensidades, seguidas pelas imagens das emoções aversão e raiva. Esses resultados provavelmente não significavam inabilidade dos avaliadores em reconhecerem expressões faciais, mas indícios de que as caricaturas precisavam ser aperfeiçoadas (TESTA et al., 2015).

Embora o FACS defina as principais alterações faciais envolvidas em cada emoção, a crítica, no entanto, é que pessoas reais não conseguem isolar os músculos faciais presentes nessas unidades de ação (DYCK et al., 2008). Este fato poderia ser um dos pontos a ser trabalhado no intuito de melhorar a representação das emoções nas caricaturas.

1.2 Hipótese

É possível gerar imagens que expressem melhor a emoção desejada, comparativamente às imagens geradas pela metodologia 1, se aprender os parâmetros do grafo And-Or a partir de imagens de pessoas reais e movimentando todos os pontos de controle.

1.3 Proposta deste projeto

A partir da motivação apresentada, o objetivo do presente projeto é definir, implementar e avaliar uma abordagem para melhorar as caricaturas anteriormente citadas. A proposta consiste em: i) variar potencialmente todos os pontos de controle de todos os componentes faciais em cada emoção, já que pessoas reais normalmente não conseguem isolar apenas alguns músculos faciais em uma expressão e ii) aprender as variações de todos esses pontos de controle a partir de imagens reais, ao invés de defini-los empiricamente. Esta metodologia proposta será denominada metodologia 2 para um melhor entendimento do texto.

Para isso, foram definidos os seguintes objetivos específicos:

- pesquisar e/ou criar uma base de imagens de faces contendo conjuntos de imagens de uma mesma pessoa expressando mais de uma emoção;
- desenvolver uma metodologia de processamento de imagens para localizar os pontos de controle definidos na metodologia 1 em imagens reais de pessoas expressando as diferentes emoções;
- aprender essas variações para a caracterização das expressões faciais referentes à face neutra e às emoções de satisfação, tristeza, raiva, aversão, medo e surpresa;
- avaliar a abordagem, comparando os resultados obtidos com a metodologia anterior por meio de condução de um novo experimento com voluntários.

1.4 Organização do documento

Além desta introdução, este documento é dividido em mais cinco capítulos. O capítulo 2 faz uma introdução sobre os conceitos mais importantes na área de linguagens formais, grafos *And-Or* e processamento de imagens. A revisão sistemática realizada é descrita no capítulo 3, que tem como objetivo apresentar uma visão geral sobre o uso de grafos *And-Or* para reconhecimento e geração de imagens. O capítulo 4 traz a metodologia utilizada neste projeto. No capítulo 5 são apresentados e discutidos os resultados obtidos. Finalmente, o capítulo 6 apresenta as conclusões e os trabalhos futuros.

No Apêndice A apresenta-se uma descrição da aparência de cada unidade de ação. A descrição detalhada de cada ponto controle definido na metodologia 1 é apresentado em B. Os posicionamentos dos pontos de controle na metodologia 1 são apresentados no

Apêndice C. O esquema do grafo *And-Or* utilizado é apresentado no Apêndice D. No Apêndice E são mostrados os valores das variações dos pontos de controle obtidos a partir da abordagem proposta neste projeto (metodologia 2). O protocolo utilizado na revisão sistemática é descrito no Apêndice F. No Apêndice G apresenta-se o protocolo de aquisição de imagens do banco de dados *LApIS Database*.

Finalmente, no Anexo A apresenta-se a aprovação do comitê de ética da Universidade de São Paulo do projeto de formação do banco de dados de expressões faciais *LApIS Database*. O termo de consentimento livre e esclarecido definido neste projeto é apresentado no Anexo B.

2 Fundamentação Teórica

Neste capítulo são apresentados os conceitos fundamentais relacionados ao presente trabalho. Na seção 2.1 são abordados os conceitos de gramáticas de *strings* e grafos *And-Or* para imagens. Os principais conceitos de processamento de imagens são apresentados na seção 2.2. Uma descrição detalhada da geração de caricaturas na metodologia 1 é apresentada em 2.3. A seção 2.4 descreve o uso do grafo *And-Or* para representação das faces no contexto do presente projeto. Por fim, alguns dos trabalhos correlatos sobre localização de pontos de controle são apresentados na seção 2.5.

2.1 Gramáticas

As definições aqui apresentadas foram extraídas de (AHO; ULLMAN, 1972; SALOMAA, 1973; HOPCROFT; ULLMAN, 1979; MENEZES, 2005).

Alguns conceitos são necessários para as definições apresentadas nesta seção. Um **alfabeto** Σ é um conjunto finito, não vazio, de símbolos (elementos). Sobre Σ pode-se definir uma **cadeia** w como uma sequência finita de símbolos. O tamanho de uma cadeia w , denotado por $|w|$, é definido pelo número de símbolos que a compõem. Uma **cadeia vazia** ε é uma cadeia constituída de nenhum símbolo, isto é, uma cadeia de tamanho 0 (zero). O **fechamento reflexivo** de um conjunto Σ , representando por Σ^* é dado pelo conjunto de todas as possíveis sequências que podem ser formadas a partir de Σ , inclusive a cadeia vazia.

Definição 1. Uma **linguagem** L sobre um alfabeto Σ é um subconjunto de Σ^* , isto é, $L \subseteq \Sigma^*$.

Uma **gramática** G define uma estrutura sobre um alfabeto Σ de forma a permitir que apenas determinadas combinações de símbolos, ou seja, que apenas determinadas cadeias, sejam válidas, definindo assim a linguagem que ela representa. Em outras palavras, a finalidade de uma gramática é definir um subconjunto de Σ^* que forma uma determinada linguagem.

Definição 2. Uma **gramática** G é uma quádrupla $G = (V_n, V_t, R, S)$ na qual:

- V_n é um conjunto finito de símbolos denominados *não terminais*¹;
- V_t é um conjunto finito de símbolos denominados *terminais*²;
- R é um conjunto finito de pares (α, β) denominados produções (ou regras gramaticais ou de sintaxe). Uma produção é dada por $\alpha \rightarrow \beta$, na qual $\alpha \in (V_n \cup V_t)^* V_n (V_n \cup V_t)^*$ e $\beta \in (V_n \cup V_t)^*$;
- $S \in V_n$ é o símbolo inicial da gramática.

Definição 3. Dadas duas cadeias c_1 e c_2 , c_2 é uma **derivação direta** de c_1 , denotada por $c_1 \Rightarrow c_2$ se existem quatro cadeias f_1, f_2, α e β (vazias ou não), tais que:

- $c_1 = f_1 \alpha f_2$
- $c_2 = f_1 \beta f_2$
- $\alpha \rightarrow \beta \in R$

Definição 4. Uma cadeia c_2 é uma **derivação de uma cadeia** c_1 , denotada por $c_2 \Rightarrow^* c_1$, se existem cadeias $c_{11}, c_{12}, \dots, c_{1n}$, onde $c_1 \Rightarrow c_{11} \Rightarrow c_{12} \Rightarrow \dots \Rightarrow c_{1n}$.

Definição 5. Uma cadeia c_1 é uma **sentença** de uma gramática G se $c_1 \in \{c \in V_t^* \mid S \Rightarrow^* c\}$. Ou seja, uma sentença é uma cadeia composta apenas de símbolos terminais que seja derivação do símbolo inicial.

Definição 6. Dada uma gramática G , a **linguagem gerada** por G , denotada por $L(G)$, o conjunto de todas as sentenças de G .

As gramáticas podem ser hierarquizadas em quatro tipos segundo a classificação de Chomsky ([MENEZES, 2005](#)):

- Gramática **tipo 0** ou gramática irrestrita:

$$R = \{ \alpha \rightarrow \beta \mid \alpha \in (V_n \cup V_t)^* V_n (V_n \cup V_t)^* \text{ e } \beta \in (V_n \cup V_t)^* \}$$

- Gramática **tipo 1** ou gramática sensível ao contexto:

$$R = \{ \alpha \rightarrow \beta \mid |\alpha| \leq |\beta|, \alpha \in (V_n \cup V_t)^* V_n (V_n \cup V_t)^* \text{ e } \beta \in (V_n \cup V_t)^* \}$$

¹ Os símbolos não terminais correspondem a símbolos auxiliares utilizados na descrição da linguagem gerada pela gramática.

² Os símbolos terminais correspondem aos símbolos que compõem as sentenças da linguagem gerada por G , ou seja V_t é o alfabeto da linguagem gerada por G .

- Gramática **tipo 2** ou gramática livre de contexto:

$$R = \{ A \rightarrow \beta \mid A \in V_n \text{ e } \beta \in (V_n \cup V_t)^* \}$$

- Gramática **tipo 3** ou gramática regular:

$$R = \{ A \rightarrow aX \mid A \in V_n, a \in V_t \text{ e } X \in \{V_n \cup \{\varepsilon\}\} \}$$

Um grafo *And-Or* é um tipo de gramática sensível ao contexto que pode ser aplicada a imagens. Diferente de uma *string* na qual os símbolos aparecem concatenados, em uma imagem os símbolos podem apresentar outras relações, como adjacência, concentricidade, etc. Logo, é também necessário definir os tipos das relações entre os símbolos (ZHU; MUMFORD, 2007). A seguir apresenta-se a definição de grafos *And-Or*, um exemplo dele é apresentado na Figura 2.1.

Definição 7. Um grafo *And-Or* é uma quádrupla $G_{And-Or} = (S, V_t, V_n, R)$, na qual:

- S é o símbolo inicial, graficamente representado por um nó raiz³;
- V_t é o conjunto de nós terminais ou primitivas⁴, graficamente representado por nós folhas;
- V_n é o conjunto de nós não-terminais, graficamente representado por nós internos. $V_n = V^{and} \cup V^{or}$, sendo V^{and} o conjunto de nós *And* e V^{or} o conjunto de nós *Or*. Um nó *And* representa a composição desse nó por seus nós filhos⁵. Um nó *Or* representa as formas alternativas definidas por seus filhos⁶;
- R é o conjunto de relações entre os nós em todos os níveis, graficamente representado por arestas entre os nós⁷.

Um **grafo sintático** é definido como uma interpretação hierárquica de uma imagem (subgrafo) no qual, para cada nó *Or*, apenas uma das arestas direcionada é percorrida (Figura 2.2). A tarefa de decomposição de uma imagem em suas partes visuais constituintes (Figura 2.3), é definida como **análise sintática** de imagens, cuja saída é representada por um grafo sintático como a interpretação mais provável de uma imagem de entrada.

³ Esse nó pode ser considerado como uma cena ou categoria de objeto.

⁴ Estruturas mínimas de uma imagem, como por exemplo linhas, pedaços de imagens, etc.

⁵ Um nó *And* equivale, na gramática de *strings*, ao lado esquerdo de uma produção se decompondo nos símbolos do lado direito da produção.

⁶ Um nó *Or* equivale, na gramática de strings, às produções alternativas para um mesmo lado esquerdo.

⁷ Essas relações podem ser entre nós de diferentes níveis, compondo os filhos de um nó pai (graficamente representadas por arestas verticais e direcionadas), ou entre nós de um mesmo nível caracterizando relações como concentricidade, adjacência, etc. (graficamente representadas por arestas horizontais e não-direcionadas).

Figura 2.1 – Um exemplo de um grafo *And-Or* para a categoria do objeto relógio. Ele tem os dois grafos sintáticos apresentados na Figura 2.2, um dos quais está ilustrado com setas escuras.

Fonte: (ZHU; MUMFORD, 2007)

Figura 2.2 – Dois exemplos de grafo sintático de imagens de relógios, gerados a partir do grafo *And-Or* da Figura 2.1.

Fonte: (TU et al., 2005)

Figura 2.3 – Exemplo de análise sintática.

Fonte: (TU et al., 2005)

2.2 Processamento de imagens

Esta subseção tem como objetivo apresentar os conceitos básicos sobre processamento de imagens, extraídos de (GONZALEZ; WOODS, 2011; NUNES, 2006; RUSS, 1995).

O termo **imagem digital** refere-se à função bidimensional de intensidade da luz $f(x,y)$, x e y denotam as coordenadas espaciais e o valor de f em qualquer ponto (x,y) é proporcional ao brilho da imagem naquele ponto. Como a luz é uma forma de energia, $f(x,y)$ deve ser positiva e finita, como apresentado na Equação 2.1.

$$0 < f(x,y) < \infty \quad (2.1)$$

Uma imagem digital no domínio espacial é uma função $f(x,y)$ discretizada tanto em coordenadas espaciais quanto em brilho, e pode ser considerada como sendo uma matriz cujos índices de linhas e de colunas identificam um ponto na imagem, e o correspondente valor do elemento da matriz identifica a cor naquele ponto. Quando se trata de uma imagem monocromática, esta cor geralmente se refere a um nível da escala de cinza, que varia do preto (valor zero) a um valor máximo que denota a cor branca. Os elementos dessa matriz digital são chamados de elementos da imagem ou “*pixels*”, este último abreviação de “*picture element*” (elemento da figura).

As imagens que as pessoas percebem em atividades visuais corriqueiras consistem de luz refletida dos objetos. A natureza básica de $f(x,y)$ pode ser caracterizada por dois

componentes: i) a quantidade de luz incidindo na cena sendo observada e ii) a quantidade de luz refletida pelos objetos na cena. Esses objetos são chamados de iluminação e reflectância, respectivamente, e são representados por $i(x,y)$ e $r(x,y)$. O produto das funções $i(x,y)$ e $r(x,y)$ resulta em $f(x,y)$ (Equações 2.2 a 2.4).

$$f(x,y) = i(x,y).r(x,y) \quad (2.2)$$

$$\text{onde } 0 < i(x,y) < \infty \quad (2.3)$$

$$\text{e } 0 \leq r(x,y) < 1. \quad (2.4)$$

A natureza de $i(x,y)$ é determinada pela fonte de luz, e $r(x,y)$ é determinada pelas características dos objetos na cena. Os valores dados nas equações 2.3 e 2.4 são limites teóricos. Ao longo deste texto, denomina-se l o nível de cinza de uma imagem monocromática f nas coordenadas (x,y) da imagem. Das equações 2.2 e 2.4 verifica-se que l fica restrito ao intervalo apresentado na Equação 2.5.

$$L_{min} \leq l \leq L_{max} \quad (2.5)$$

O intervalo $[L_{min}, L_{max}]$ é denominado escala de cinza. A prática comum é deslocar esse intervalo para $[0, L]$, onde $l=0$ refere-se à cor preta e $l=L$ refere-se à cor branca. Os valores intermediários são tons de cinza variando continuamente entre o branco e o preto.

O *histograma* de uma imagem digital com níveis de cinza no intervalo $[0, L-1]$ é uma função discreta $p(r_k) = n_k/n$, em que r_k é o k -ésimo nível de cinza, n_k é o número de pixels na imagem com esse nível de cinza, n é o número total de pixels na imagem e $k=[0...L-1]$. Um gráfico (Figura 2.4.b) dessa função para todos os valores de k fornece uma descrição global da aparência de uma imagem.

Quando o objetivo é representar imagens digitalmente, há diversos tipos de representação que podem ser empregados de acordo com a forma que se deseja trabalhar com as cores. O **espaço RGB** é formado pela nossa sensação da soma ponderada do *red* (R), *green* (G) e *blue* (B), os quais geram a maioria das cores visíveis. Seu espaço complementar CMY é formado pelo *cyan* ($C=255-R$), *magenta* ($M=255-G$) e *yellow* ($Y=255-B$).

Uma cor pode ser decomposta em três componentes independentes: intensidade (I), matiz (M) e saturação (S). A intensidade é responsável pela sensação de brilho, o

Figura 2.4 – (a) Imagem original em tons de cinza. (b) Exemplo de histograma de uma imagem.

Fonte: (NUNES, 2006)

matiz pela sensação de cor (comprimento de onda), e a saturação pelo grau de pureza da cor. O **espaço HSV** representa o matiz (H), a saturação (S) e o brilho (V). Como os componentes primários são de correlacionados, melhoramentos na imagem através de transformações radiométricas aplicadas à saturação e/ou brilho não afetarão a matiz.

Uma **transformação geométrica** 2D (bidimensional) é uma função que leva um ponto $P(x,y)$ em outro ponto $P(x',y')$ no espaço R^2 . A transformação de translação t_x, t_y e sua inversa são dadas em coordenadas homogêneas pelas equações 2.6 e 2.7, respectivamente.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.6)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (2.7)$$

A transformação de rotação por um ângulo θ no sentido horário aplicada a um ponto (x,y) , e sua inversa são dadas em coordenadas homogêneas pelas equações 2.8 e 2.9, respectivamente.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.8)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (2.9)$$

A transformação de escala (s_x, s_y) aplicada a um ponto (x,y) , e sua inversa são dadas em coordenadas homogêneas pelas equações 2.10 e 2.11, respectivamente.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.10)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_x} & 0 & 0 \\ 0 & \frac{1}{s_y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (2.11)$$

O processamento de imagens digitais abrange uma ampla escala de hardware, software e fundamentos teóricos (GONZALEZ; WOODS, 2011). Os passos fundamentais necessários para executar uma tarefa de processamento de imagens são apresentados na Figura 2.5.

As operações de processamento de imagens podem ser divididas em três grupos diferentes: baixo, médio e alto nível.

No **processamento de baixo nível** é feita a remoção de ruídos ou realce de dados de interesse em uma imagem.

Uma *máscara* (também denominada “molde”, janela ou filtro) é definida como uma matriz bidimensional (por exemplo de tamanho 3x3 pixels) cujos valores dos coeficientes determinam a natureza do processo, tal como o aguçamento de imagens. O uso de máscaras para processamento de imagens é usualmente chamado *filtragem* no domínio espacial.

Figura 2.5 – Passos fundamentais para executar uma tarefa de processamento de imagens.

Fonte: (GONZALEZ; WOODS, 2011)

Filtros de **suavização** são utilizados para diminuir efeitos espúrios resultantes dos processo de aquisição da imagem (por exemplo ruídos). Entre eles, o filtro **Gaussiano** utiliza a função Gaussiana unidimensional (Equação 2.12) para obter valores de uma máscara a ser definida digitalmente, onde σ é o desvio padrão.

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (2.12)$$

Considerando que a imagem é armazenada como uma coleção de pixels discretos, precisa-se produzir uma forma discreta da distribuição Gaussiana para obter uma máscara. A Equação 2.13 apresenta a forma inteira 5x5 da máscara da Gaussiana para $\sigma = 1.0$.

$$M_{Gaus} = \frac{1}{273} \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix} \quad (2.13)$$

A **segmentação** tem como objetivo identificar as formas significativas de uma imagem a fim de fornecer informações para sua interpretação. Segmentar uma imagem consiste em particioná-la em regiões de pixels relevantes para uma dada aplicação. A segmentação consiste em duas tarefas básicas: identificação e delineamento. A identificação indica a localização aproximada do objeto e o delineamento extrai sua extensão na imagem.

Limiarização (*Thresholding*) é uma das mais importantes metodologias para segmentação de imagens. Esta técnica consiste em definir valores de limiares e reconhecer

como partes de um mesmo objeto os pixels dentro de um determinado intervalo desses valores. Suponha que o histograma de níveis de cinza mostrado na Figura 2.6.a corresponde a uma imagem $f(x,y)$, composta por objetos iluminados sobre um fundo escuro, de maneira que os pixels do objeto e os do fundo tenham seus níveis de cinza agrupados em dois grupos dominantes. Uma maneira de extrair os objetos do fundo é através da seleção de um limiar T que separe os dois grupos. Então, cada ponto (x,y) tal que $f(x,y) > T$ é denominado um ponto do objeto, caso contrário, o ponto é denominado um ponto do fundo.

A Figura 2.6.b mostra um caso mais geral dessa abordagem, a *limiarização multiníveis*. Nesse caso, três grupos dominantes caracterizam o histograma dominante da imagem (por exemplo, dois tipos de objetos iluminados sobre um fundo escuro). A mesma abordagem básica classifica um ponto (x,y) como pertencendo à classe de um dos objetos se $T_1 < f(x,y) \leq T_2$, à classe do outro objeto se $f(x,y) > T_2$, e ao fundo se $f(x,y) \leq T_1$.

Figura 2.6 – Histograma de níveis de cinza que podem ser particionados por (a) um limiar único e (b) limiares múltiplos.

Fonte: (GONZALEZ; WOODS, 2011)

O operador **Canny** (CANNY, 1986) usa um algoritmo multi-passos para detectar bordas. O filtro de Canny é um filtro de convolução que usa a primeira derivada. Ele suaviza o ruído e localiza bordas, combinando um operador diferencial com um filtro Gaussiano.

A **morfologia matemática** (SERRA, 1982) é a parte do processamento de imagem não-linear que tem por objetivo extrair características da imagem associadas à geometria dos objetos. Uma transformação morfológica consiste essencialmente da comparação da imagem com outra menor, cuja geometria é conhecida, denominada **elemento estruturante** e

definido como um conjunto de coordenadas de pixel. Uma transformação morfológica requer uma operação não-linear entre a imagem e o elemento estruturante.

A transformação morfológica **dilatação** ($\oplus$) combina dois conjuntos utilizando adição vetorial. É definida pela Equação 2.14, onde A representa a imagem sendo processada e B é o elemento estruturante.

$$A \oplus B = \{c | c = a + b, a \in A \text{ e } b \in B\} \quad (2.14)$$

A transformação morfológica **erosão** ($\ominus$) combina dois conjuntos utilizando vetores de subtração (Equação 2.15). Ela é expressa como uma interseção de A e B . Assim é definido $A \ominus B = B \cap A$.

$$A \ominus B = \{x | x + b \in A \text{ para todo } b \in B\} \quad (2.15)$$

A transformação morfológica **abertura** ($\circ$) em geral suaviza o contorno de uma imagem e elimina proeminências delgadas. Ela aplica uma erosão imediatamente seguida de uma dilatação usando o mesmo elemento estruturante (Equação 2.16).

$$A \circ B = (A \ominus B) \oplus B \quad (2.16)$$

A transformação morfológica **fechamento** ($\bullet$) trabalha de um modo oposto ao método abertura. Se uma abertura cria pequenos espaços vazios na imagem, um fechamento irá preencher os espaços vazios (Equação 2.17).

$$A \bullet B = (A \oplus B) \ominus B \quad (2.17)$$

2.3 Geração de imagens

A seguir são apresentados os conceitos necessários para entender o escopo da abordagem inicial do jogo (metodologia 1). Como descrito no capítulo 1, as caricaturas desenhadas são construídas considerando que cada componente da face é composto por curvas *spline* definidas por seus pontos de controle.

Medidas antropométricas definidas em (FARKAS, 1994) foram utilizadas para definir e calcular o posicionamento dos pontos de controle em uma face neutra inicial. Para a realização das medidas, foi determinado um conjunto de pontos bem definidos sobre o corpo

humano, chamados de *landmarks*. Um exemplo dos pontos definidos na face é mostrado na Figura 2.7.a. A distância entre dois *landmarks* caracteriza uma medida antropométrica (Figura 2.7.b).

Figura 2.7 – (a) *Landmarks* antropométricos faciais. (b) Medições antropométricas faciais.

Fonte: (FARKAS, 1994)

As medidas antropométricas variam de acordo com o gênero, o grupo étnico e a idade dos indivíduos. Assumindo que essas medidas assumem uma distribuição normal, a Antropometria define, para cada gênero, grupo étnico e faixa etária, a média e o desvio padrão de cada medida. Para a constituição do modelo antropométrico utilizado nesta abordagem, foram considerados os dados para uma população composta de indivíduos caucasianos norte-americanos saudáveis, masculinos, na faixa etária de 19-25 anos.

As medidas antropométricas referem-se às distâncias entre os *landmarks* e não às posições dos mesmos. Portanto, a fim de desenhar uma face neutra inicial, foi necessário calcular o posicionamento de cada um desses pontos no espaço bidimensional da imagem com base nas relações de distância entre dois pontos, fornecidos pelas medidas antropométricas. Finalmente, os *landmarks* foram utilizados na definição dos pontos de controle que compõem as curvas que formam cada parte da face (Figura 2.8). Uma descrição mais detalhada para cada ponto é encontrada no Apêndice B.

Para contemplar a tarefa de reproduzir as expressões faciais, é necessário simular os movimentos provocados pelas emoções na face, e aplicar esse efeito nas curvas que compõem a face. A seguir é descrito como o FACS foi empregado para obter uma representação das expressões faciais, a saber; satisfação, tristeza, medo, surpresa, raiva e aversão.

A fim de ilustrar as unidades de ação, a Figura 2.9 mostra as unidades de ação 1 e 2. A unidade de ação 1 define uma elevação dos cantos internos da sobrancelha e um enrugamento horizontal na testa (Figura 2.9.a). Já a unidade de ação 2 define uma elevação da porção exterior das sobrancelhas, produção de uma forma arqueada para as sobrancelhas e a possibilidade de rugas laterais entre as sobrancelhas (Figura 2.9.b). No Apêndice A é apresentada uma descrição dos movimentos considerados por todas as unidades de ação.

Figura 2.9 – Exemplos de unidades de ação: (a) unidade de ação 1, (b) unidade de ação 2.

Fonte: (TESTA et al., 2015)

Apesar da ampla descrição de cada unidade de ação, as ciências comportamentais não estão preocupadas em medir exatamente com qual intensidade se varia um músculo em uma dada unidade de ação. Há apenas uma descrição de qual músculo (ou grupo de músculos) se move, e a implicação dessa ação nas partes da face.

Para contornar esse problema, foi preciso encontrar uma maneira alternativa de realizar as variações nos pontos de controle das curvas *splines* sem gerar inconsistências. Para isso definiu-se um envoltório definido como o dobro do tamanho do retângulo que envolve a curva, como ilustrado na Figura 1.2. Assim, o reposicionamento de um ponto consiste em variar um percentual da medida do envoltório correspondente.

A posição do ponto de controle em uma dada emoção é calculada com base nas coordenadas x e y da face neutra. Nas equações 2.18 e 2.19, o elemento α é o percentual necessário de variação para a unidade de ação de acordo com o envoltório para a emoção que está sendo representada, I regula o quão intensa será a emoção (valor no intervalo

de 0 a 1) e $PN1-PN0$ e $PN3-PN2$ são as medidas do envoltório horizontal e vertical respectivamente (Figura 1.2).

$$PE_x = PN_x + \alpha.I.|PN1 - PN0| \quad (2.18)$$

$$PE_y = PN_y + \alpha.I.|PN3 - PN2| \quad (2.19)$$

No Apêndice C é apresentado como este cálculo das posições se estabelece para cada ponto de controle em cada unidade de ação.

2.4 Uso do grafo *And-Or* para representação das faces

O grafo *And-Or* foi utilizado nesse projeto com o objetivo de flexibilizar a maneira e a estrutura para armazenar e gerenciar o desenho das caricaturas⁸. Ele representa somente a face neutra e é esquematizado no Apêndice D.

O grafo é composto apenas por nós *And* estruturados em três níveis, como pode ser visto parcialmente⁹ na Figura 2.10. No primeiro nível, há apenas um nó, que representa a face. No segundo nível, há todos os componentes que a formam, como: olho esquerdo *OE*, olho direito *OD*, boca *B*, nariz *N*, etc. No terceiro e último nível, há os nós das curvas que formam cada um dos componentes, por exemplo, *OE_C1* é uma curva de *OE*.

O grafo é lido a partir de um arquivo de entrada composto por três tipos de informações: (i) dados de criação de nós, cujos parâmetros são um identificador numérico, um rótulo ou identificador de nome, um vetor de pontos de controle (utilizado no caso de ser um nó folha) e seu tipo (*And* ou *Or*); (ii) dados das coordenadas de pontos de controle que serão armazenadas nos nós curvas (média e desvio padrão); (iii) dados das relações entre nós, nas quais são determinados os nós pais e os nós filhos.

Quando é solicitado que a caricatura da face neutra seja gerada, as rotinas acessam o grafo. É acessado o primeiro nível, no qual se inicia todo o percurso. O segundo nível é então acessado, no qual um componente da face será desenhado. Para isso, são utilizadas

⁸ A estrutura inicial do grafo e as principais rotinas de criação e manipulação do grafo foram desenvolvidas durante a Iniciação Científica do aluno Antônio Henrique Nunes Muniz. Uma reestruturação desse grafo e dessas rotinas assim como a integração com as rotinas de desenho das faces foram desenvolvidas no Trabalho de Conclusão de Curso do aluno Arthur Felipe Batista Moraes.

⁹ Para uma melhor visualização, essa Figura descreve apenas 3 componentes faciais com o intuito de exemplificar a estrutura. O grafo completo é apresentado no Apêndice D.

Figura 2.10 – Representação parcial do grafo *And-Or*.

Fonte: Autor Arthur Felipe Batista

as coordenadas que formam a curva e dessa forma o terceiro nível é requisitado. Como os nós do terceiro nível são nós folhas, as rotinas de desenho são invocadas e é passado como parâmetro o vetor de coordenadas (média e desvio padrão) contido dentro do nó. As coordenadas x e y exatas de cada ponto são sorteadas dentro da distribuição normal definida. Desta forma, cada solicitação de desenho de caricatura desenhará uma face neutra ligeiramente diferente, o que é uma característica desejável neste projeto.

2.5 Trabalhos correlatos de localização de pontos de controle

O problema de localização de pontos de controle faciais ainda é considerado desafiador em visão computacional. Trabalhos como (KIM; KIM; BANG, 2001), (PHIMOLTARES; LURSINSAP; CHAMNONGTHAI, 2002), (LEE et al., 2001) utilizam técnicas de aprendizado de máquina, como análise de componentes principais (PCA), redes neurais e algoritmos genéticos, respectivamente, para estimar as posições de pontos de controle faciais. Para treinar seus modelos, essas técnicas exigem um grande número de imagens e alto tempo computacional.

Por sua vez, em (SOHAIL; BHATTACHARYA, 2008) os autores propõem várias combinações de técnicas de processamentos de imagens com o intuito de localizar dezoito pontos de controle. Cada combinação de técnicas é destinada a um componente facial específico, como boca, olhos, etc.

No capítulo seguinte será apresentada a revisão sistemática que motivou as ideias empregadas neste projeto geral com relação ao uso do grafo *And-Or*.

3 Revisão Sistemática

Neste capítulo são apresentados e discutidos os resultados de uma revisão sistemática feita para identificar o estado da arte na área de geração e reconhecimento de imagens utilizando grafos *And-Or*. O planejamento, a condução e o levantamento dos trabalhos, já desenvolvidos nesta área, são descritos nas seções a seguir.

3.1 Planejamento e condução

A revisão sistemática foi realizada considerando as três fases sugeridas em (KIT-CHENHAM, 2004): planejamento, condução e extração e síntese dos dados. Inicialmente foi elaborado um protocolo de pesquisa de forma detalhada, no qual é descrito o processo e os métodos aplicados durante a revisão sistemática (Apêndice F).

Na sequência, com o intuito de responder a pergunta de pesquisa (Apêndice G) definida no protocolo, foram construídas duas *strings* de busca, as quais foram submetidas a diferentes máquinas de busca (*ACM, IEEE, Scopus e ISI*), obtendo como resultado um total de 207 artigos. Desses primeiros 207 estudos retornados, 20 foram incluídos após aplicação dos critérios de inclusão e exclusão, definidos para apoiar a seleção de estudos relevantes. Finalmente, esses 20 artigos foram lidos na íntegra para extrair informações relevantes. A sequência de passos do processo de seleção é apresentada na Figura 3.1.

3.2 Resultados

Uma análise global dos trabalhos é exposta na seção 3.2.1. Os artigos incluídos são descritos nas seções 3.2.2 a 3.2.5, divididos em categorias considerando o tipo de algoritmo utilizado e o propósito para o qual foi utilizado.

3.2.1 Análise global

A revisão sistemática realizada mostrou que, nos últimos anos, a quantidade de estudos utilizando grafos *And-Or* para tarefas de geração e reconhecimento de imagens vem crescendo, mas ainda é pequena. A evolução desta área de pesquisa com relação à

Figura 3.1 – Processo de seleção dos estudos primários.

Fonte: Autoria própria

quantidade de artigos publicados pode ser observada na Figura 3.2. A maioria dos artigos foi publicada nos últimos quatro anos, com picos nos anos de 2013 e 2014.

Figura 3.2 – Quantidade de artigos publicados.

Fonte: Autoria própria

Diferentes artigos focam diferentes aspectos da utilização de grafos *And-Or*. Por exemplo, em (LIN et al., 2015) é apresentado um algoritmo de aprendizado da estrutura do grafo, Wu, Xia e Zhu (2007) apresenta um algoritmo de análise sintática e estimação de parâmetros. Já em (DONG et al., 2014), (LIN et al., 2012) e (JIANG et al., 2012) apresentam-se aplicações específicas como representação de objetos articulados, categorização e temporalidade, respectivamente. Dessa forma, uma classificação pode ser feita, como a representada na Figura 3.3.

Figura 3.3 – Tendências de uso de grafos *And-Or*.

Fonte: Autoria própria

Na Figura 3.4 é possível observar que, do total de artigos, 12 são de um mesmo grupo de pesquisa, liderados pelo Prof. Dr. Song Chun Zhu, quem pela primeira vez propôs o uso de grafos *And-Or* em imagens (CHEN et al., 2006) no ano 2006.

Figura 3.4 – Quantidade de artigos publicados por grupo de pesquisa.

Fonte: Autoria própria

3.2.2 Geração de imagens

Os três artigos que utilizam grafos *And-Or* para gerar imagens partem, na verdade, de alguma imagem de entrada. Eles destinam-se a diferentes finalidades e utilizam diferentes tipos de imagem de entrada. Em (CHEN et al., 2006) e (XU et al., 2008), um grafo *And-Or* é utilizado para interpretar uma imagem de entrada e a reconstruí-la em forma de esboço. Por sua vez, Suo et al. (2010) têm como objetivo simular o envelhecimento de faces humanas.

Em (CHEN et al., 2006) são utilizados pela primeira vez grafos *And-Or* para reconhecimento e geração de imagens de vestuário. Neste artigo foram utilizadas imagens de estudantes sentados em uma cadeira, com boas condições de luz e fundo uniforme para reduzir a oclusão e controlar a iluminação. Um artista foi convidado para desenhar esboços o mais próximos possível das imagens originais. Tais esboços foram manualmente decompostos em duas camadas: (i) a primeira contendo as partes das pessoas e vestuário, como por exemplo cabelo, face, braços, calça, sapatos, etc., e (ii) a segunda camada contendo as estruturas remanescentes, correspondentes às texturas, como, por exemplo, o texto impresso na camiseta. As partes relacionadas à primeira camada foram utilizadas

para construir manualmente um grafo *And-Or*, cujos nós terminais representam as partes identificadas nos esboços. Este grafo é então utilizado para, a partir de uma imagem de pessoa, identificar as partes das roupas e gerar um esboço das mesmas. Para a identificação das partes, a partir de imagens, é utilizado o algoritmo da análise sintática proposto em (HAN; ZHU, 2005). Este algoritmo faz uso de um processo *bottom-up* para a detecção das partes que são mais discrimináveis, como, por exemplo, a face. Em seguida, o algoritmo ativa buscas *top-down* para prever as partes com base nas informações de contexto codificadas no grafo.

Na pesquisa de Xu et al. (2008) apresenta-se um grafo *And-Or* para gerar desenhos de faces humanas. Para representar as faces em diferentes escalas e resoluções, o grafo *And-Or* foi construído contendo três camadas: (i) a primeira camada trata cada face como um todo, (ii) na segunda camada representam-se os componentes faciais, como, por exemplo, olhos e boca, e (iii) em uma terceira camada são representados detalhes faciais mais finos como marcas e rugas. Uma imagem de entrada, representando uma face, é analisada por um processo de análise sintática que então reconstrói o desenho mais adequado, segundo a resolução da imagem de entrada, percorrendo o grafo a partir do nó raiz até os nós folhas. Os resultados mostraram que o modelo ajuda a reconstruir imagens de faces altamente detalhadas.

Simular o envelhecimento de faces humanas e estimar a idade de uma pessoa são duas tarefas com aplicações interessantes em diferentes áreas de pesquisa. Considerando o potencial dessas aplicações, Suo et al. (2010) apresentam um modelo dinâmico para representar o processo de envelhecimento de faces humanas utilizando grafos *And-Or*. Um grafo *And-Or* é desenhado manualmente para representar as diferentes faixas etárias das faces. Para simular o envelhecimento das faces é criado um conjunto de grafos sintáticos, cada um dos quais representa uma face mais velha. Para avaliar o desempenho do modelo foi feita uma comparação entre as imagens sintéticas geradas e fotografias reais das faces. Segundo os autores, os resultados mostraram que as imagens geradas por este modelo preservam a identidade das faces e não têm diferenças significativas.

3.2.3 Algoritmos de reconhecimento

Em (WU; XIA; ZHU, 2007) apresenta-se um modelo baseado em grafos *And-Or* para identificar determinadas primitivas em imagens. Essas primitivas, chamadas de *graphlets*, são imagens comuns de junções simples e compostas por linhas. A estrutura do grafo foi definida manualmente a partir de uma base de dados com 200 imagens de treinamento, as quais foram segmentadas manualmente e utilizadas para aprender as relações de dependência entre as primitivas. O aprendizado dos parâmetros do grafo *And-Or* foi realizado de forma recursiva, no qual cada um dos *graphlets* são codificados de forma independente com seus respectivos parâmetros tanto geométricos (posição central, escala, orientação e deformação plástica), como de aparência (contraste e intensidade).

Wu e Zhu (2011) definiram três processos (alfa, beta e gama) para reconhecimento de imagens com oclusões utilizando grafos *And-Or*. O processo alfa lida com situações nas quais as imagens têm resolução média e sem nenhum tipo de oclusão; nesse caso, a estratégia para identificar um nó do grafo é utilizar diretamente os dados da imagem. O processo beta lida com imagens de resolução alta e é mais provável ter algum tipo de oclusão; a estratégia é identificar um nó mediante a união de seus nós filhos. Por fim, o processo gama lida com imagens de resolução muito baixa, e um nó pode ser predito somente a partir da informação do contexto que seu nó pai fornece a ele. Os três processos foram treinados separadamente com o objetivo de encontrar a quantidade de informação que cada processo contribui. Para a tarefa de reconhecimento, os três processos são integrados em um algoritmo da análise sintática, o qual percorre cada nó *And* do grafo aplicando os três processos em cada um deles, com exceção do nó raiz (só executa o processo gama) e dos nós folhas (só executam o processo beta).

3.2.4 Algoritmos de aprendizado de estruturas

Objetos podem apresentar muitas variações estruturais. Como exemplo, um objeto da classe cadeira pode ter pernas compridas e outro objeto da mesma classe pode ter pernas curtas. Essas variações tornam complexa a tarefa de aprendizado de classes. Em (SI; ZHU, 2013) é apresentado um *framework* para aprender de forma não supervisionada classes de objetos na forma de um *And-Or Template* - AOT (grafo *And-Or* que não apresenta relações entre nós do mesmo nível). Para aprender as partes reusáveis dos objetos, o algoritmo de

aprendizado define uma matriz de dados, na qual cada linha corresponde a um exemplo de treinamento e cada coluna representa uma característica do objeto (cor, textura, etc.), na qual pode-se encontrar conjuntos de características compartilhadas (*templates*) por meio do agrupamento dos valores da matriz. Para aprender a estrutura do AOT foi adicionado um modelo *a priori* das configurações estruturais (vetores binários) que indicam quais características do objeto aparecem nas imagens de treinamento.

Em (SONG et al., 2013) é utilizado um aprendizado fracamente supervisionado – somente o objeto como um todo é anotado, mas não as suas partes – para aprender a estrutura de uma árvore *And-Or*. O processo de aprendizado é composto por três etapas: (i) são aprendidas as subcategorias (diferentes pontos de vista de um objeto) de forma não supervisionada utilizando um método de agrupamento; (ii) é criada uma árvore *And-Or* para cada subclasse; um algoritmo percorre cada árvore para encontrar a melhor configuração; (iii) a partir das subcategorias aprendidas e suas correspondentes configurações, foram treinadas Máquinas de Vetores de Suporte (*Support Vector Machine* - *SVM*) para aprender os parâmetros estruturais e de aparência dos objetos.

O estudo (SI et al., 2011) propõe aprender um grafo *And-Or* de eventos automaticamente, de forma não supervisionada, a partir de vídeos em ambientes controlados. Em (PEI et al., 2013) é apresentada uma continuidade do trabalho anterior, no qual esse modelo de aprendizado é utilizado para gerar classes de eventos mais escaláveis e reduzir, ao mesmo tempo, o processo manual de anotação de imagens. O modelo é criado em três etapas: (i) detecção de posição e poses de agentes para cada *frame* de vídeo, (ii) agrupamento destas posições em um alfabeto de ações atômicas transitórias, e (iii) a partir dessas ações atômicas, eventos curtos são agrupados em eventos mais longos, recursivamente.

3.2.5 Aplicações específicas

3.2.5.1 Representação de Objetos Articulados

No trabalho (LIN et al., 2015) é apresentado um algoritmo de otimização estrutural para aprender um grafo *And-Or* de forma fracamente supervisionada. O algoritmo de aprendizado mapeia, de forma explícita, os nós do grafo em um vetor de características, a partir de cada exemplo de treinamento. A partir desse vetor de características são extraídos subvetores de características correspondentes para cada nó. Em cada subvetor é aplicado

um procedimento de agrupamento e, com base nos resultados deste procedimento, o vetor de características é reorganizado. Consequentemente, a estrutura do grafo pode ser gerada e seus parâmetros podem ser atualizados. Por exemplo, se os subvetores correspondentes a uma determinada parte de um objeto são agrupados em três grupos, serão criados três nós folhas e incorporados na estrutura.

Devido à complexidade e ambiguidade da representação visual de objetos articulados, definir que tipo de representação é capaz de modelar a grande variação na aparência de objetos não é um problema trivial. Em (ZHU et al., 2011) é proposto um *framework* para modelar objetos articulados – especificamente imagens de cavalos e jogadores de *basebol* – e estimar poses. Cada objeto é representado em um grafo *And-Or*, cuja estrutura foi desenhada de forma manual. Os nós do grafo especificam posição, orientação e escala das sub-partes dos objetos. Tarefas de análise sintática, aprendizado e estimação de parâmetros são contempladas no *framework*. Para o aprendizado da estrutura é utilizado o método apresentado em (ZHU et al., 2008). Por sua vez, o processo de inferência refere-se à estimação das posições e características (posição, orientação e escala) de todas as partes do objeto, procurando quais partes estão presentes na imagem.

Em (ROTHROCK; PARK; ZHU, 2013) apresenta-se um modelo na forma de um grafo *And-Or* para estimar poses de humanos. O diferencial do modelo é a integração, dentro da gramática, da composição de partes e a geometria articulada (relações de posição e orientação entre as partes) dos objetos. Nesse artigo também é descrito o aprendizado do modelo a partir de uma perspectiva de minimização de riscos.

Um modelo unificado para analisar sintaticamente imagens humanas utilizando *parselets* é apresentado em (DONG et al., 2013). *Parselets* são grupos de primitivas, segmentos de imagens analisáveis, que podem ser obtidos utilizando algoritmos de segmentação de imagens. Os *parselets* são utilizados como elementos de construção do modelo da análise sintática, agrupados para formar as partes do corpo humano, como cabeça e corpo, utilizando dois tipos de modalidades: (i) coocorrência, que representa a relação de coexistência e junção entre *parselets* e (ii) exclusividade, que modela as relações de diferentes *parselets* que não podem coexistir logicamente (por exemplo saia e calça em um mesmo objeto). Para avaliar o modelo foi construído um protótipo de um sistema de

recuperação de imagens de pessoas visualmente similares. Os resultados mostraram que as imagens são recuperadas com sucesso independentemente da pose dos humanos.

Geralmente, estimação de pose e identificação das partes do corpo humano, são tratadas de duas formas diferentes, separadamente ou de forma sequencial. Segundo (DONG et al., 2014), de forma separada ignora-se a correlação entre ambas, e de forma sequencial inclui-se erros na estimação de parâmetros. Para superar essas limitações, os autores propõem um modelo de análise híbrido para unificar as duas tarefas. O modelo define dois elementos básicos unificadores: *parselets* e *templates* de imagens. Para representar de forma hierárquica as restrições geométricas e de compatibilidade desses elementos, foi utilizado um grafo *And-Or*, cuja diferença dos grafos *And-Or* tradicionais reside no fato de que o grafo proposto tem dois tipos de nós folhas (*P-Leaf* e *M-Leaf*), para representar *parselets* e *templates* de imagens, respectivamente.

3.2.5.2 Categorização

No contexto de reconhecimento de categorias de objetos existem duas necessidades opostas: i) ter um número grande de instâncias de treinamento para conseguir aprender as várias possíveis aparências de um objeto; ii) apresentar baixo tempo de execução da análise sintática de um novo objeto, mas esse tempo aumenta de forma diretamente proporcional à variabilidade de aparência aprendida pelo grafo. Para resolver estes dois problemas, o trabalho (LIN et al., 2012) apresenta um *framework* para a categorização de objetos utilizando grafos *And-Or*. Para tratar o primeiro problema, são criados novos exemplos de objetos que, embora semelhantes em aparência aos dados de treinamento originais, incrementam significativamente o conjunto de treinamento. Para resolver o segundo problema é proposta uma poda em cascata do grafo com a finalidade de obter um conjunto pequeno de objetos e categorias candidatas para uma imagem nova. De acordo com os pesquisadores, os resultados experimentais mostraram que o modelo proposto permite resolver os dois problemas descritos, já que permite uma melhor representação de objetos e executa de forma eficiente a classificação de imagens novas utilizando poucos exemplos de treinamento.

No contexto de classificação multiclasse de objetos, há dois problemas que muitas vezes são ignorados ou muito simplificados: (i) capturar a grande variação de aparência e posição de objetos, e (ii) o tratamento das diferentes classes de objetos como entidades sem

conexão. Para resolver estes problemas, o artigo (WANG et al., 2013) propõe um modelo hierárquico e reconfigurável capaz de reconhecer objetos multi-classe utilizando grafos *And-Or*. No grafo, os nós Or escolhem as composições alternativas e os nós folhas são compartilhados entre as diferentes classes para representar as categorias de objetos. Para treinar o modelo foi utilizado um algoritmo *EM-type*, baseado no método de Esperança-Maximização (EM), que determina as configurações estruturais e os parâmetros do grafo.

3.2.5.3 Temporalidade

Uma outra aplicação de grafos *And-Or* é análise de vídeos. Em (JIANG et al., 2012) e (PEI; DONG; ZHAO, 2013) são analisados vídeos de atividades humanas no cotidiano. O artigo (JIANG et al., 2012) apresenta uma abordagem para prever eventos em vídeos de futebol a partir de ações realizadas por jogadores. Para este fim, foi construído um grafo *And-Or* que descreve as diferentes configurações de eventos e as relações temporais entre os subeventos. A estrutura do grafo foi definida manualmente e tanto os parâmetros quanto as relações entre os nós foram aprendidas a partir de dados anotados. No grafo, os nós terminais representam ações humanas atômicas, em que cada ação é realizada por um determinado jogador em um intervalo de tempo contínuo. Um algoritmo de inferência de eventos foi desenvolvido com o objetivo de encontrar a melhor configuração do grafo, misturando um processo *bottom-up*, o qual propõe as diferentes configurações do grafo, e um processo *top-down* para aperfeiçoar essas configurações propostas.

Por sua vez, o artigo (PEI; DONG; ZHAO, 2013) apresenta um método para reconhecimento de eventos baseados em papéis sociais (ex. gerente, pesquisador, visitante, etc.) utilizando grafos *And-Or*. Os papéis sociais podem ser inferidos a partir das atividades diárias dos agentes ao longo de várias sequências de vídeos de segurança. O método inclui interações entre os agentes e objetos da cena de forma que seja possível, por exemplo, inferir a intenção de uma pessoa em determinado vídeo em termos de seus papéis sociais.

A composição de *frameworks* tem sido comum na literatura quando trata-se da análise de vídeos usando grafos *And-Or*. Em (PEI et al., 2013) apresenta-se uma abordagem para prever eventos futuros em vídeos utilizando grafos *And-Or* temporários (T-AOG). O aprendizado das ações atômicas e das relações temporais ocorre de forma não supervisionada com base no modelo apresentado em (SI et al., 2011). Para prever ações futuras de agentes, o algoritmo de inferência procura o grafo sintático mais provável, de forma semelhante

aos algoritmos de análise sintática em imagens estáticas. O diferencial do algoritmo é sua capacidade de manipular eventos intercalados. Além disso, devido à ambiguidade decorrente nos vídeos têm-se “pontos de decisão”, nos quais os grafos sintáticos são podados para manter uma eficiência computacional. Na avaliação do modelo foram comparadas as predições do algoritmo com as predições de pessoas. Os resultados mostraram que a taxa de acerto do algoritmo foi de 90% e dos participantes de 87%.

O artigo (AMER et al., 2013) propõe um modelo de análise de vídeos sensível ao custo computacional, utilizando um grafo *And-Or* espaço-temporal (ST-AOG) para modelar atividades, relações espaço-temporais e relações de contexto de vídeos. O ST-AOG captura as ações individuais (movimentos repetitivos de uma mesma pessoa) dentro de atividades grupais. A melhor sequência de passos de inferência é aprendida para cada tipo de consulta a partir de exemplos de treinamento. No modelo, dada uma consulta (por exemplo: de onde e quando ocorre uma determinada atividade no vídeo), é formulada uma inferência do grafo com base em Árvores de Decisão Monte Carlo (*Monte Carlo Tree Search - MCTS*), o qual executa uma sequência de processos (α , β , γ e δ) para cada nó do grafo. O processo α detecta objetos, ações e atividades grupais a partir de características do vídeo, β prevê uma atividade a partir de seus nós filhos, o processo γ prevê ações ou objetos utilizando o contexto de seus nós pais, e δ faz acompanhamento de uma atividade ou uma ação. Segundo os autores, os resultados mostraram que o modelo melhora significativamente o reconhecimento de atividades e reduz o tempo computacional, uma vez que os vídeos são operados em blocos.

Em (LU; WU; ZHU, 2014) apresenta-se um *framework* para rastrear (localizar um determinado objeto na sequência de quadros em vídeo), aprender e analisar sintaticamente objetos em vídeo simultaneamente. Esse *framework* inclui três componentes: (i) construção da estrutura do grafo, (ii) formulação do algoritmo de rastreamento e análise sintática, e (iii) aprendizado do grafo. Para rastrear um determinado objeto, o grafo *And-Or* foi treinado a partir de instâncias dos objetos de forma incremental. Primeiro foi treinado um grafo inicial a partir de um conjunto pequeno de instâncias positivas e negativas. Nos *frames* subsequentes, o grafo *And-Or* é reaprendido de forma iterativa, atualizando a estrutura do grafo e, se necessário, reestimando os parâmetros.

Em (WANG; PEI; ZHU, 2014) apresenta-se um método para detectar se motoristas estão fazendo uso de celular enquanto dirigem. As atividades de ligação (levantar a mão, ligar e descer a mão) são representadas de forma hierárquica em um grafo *And-Or*, cuja estrutura foi desenhada manualmente e seus parâmetros foram aprendidos a partir de dados anotados. Além disso, foi utilizado um algoritmo de análise sintática para detectar as atividades de ligação. O método foi avaliado para detectar as atividades de ligação e o estado dos lábios na ligação. De acordo com os autores, os resultados mostraram que a taxa de detecção foi de 95%.

3.3 Discussões

Com os resultados obtidos a partir dos artigos incluídos, é possível perceber que uma das principais vantagens dos grafos *And-Or* é sua capacidade de aprender diferentes configurações possíveis de objetos a partir de pequenos conjuntos de imagens.

A abordagem sintática dos grafos *And-Or* permite representar de forma eficiente as estruturas hierárquicas de objetos e suas variações, facilitando o reconhecimento de imagens por meio delas. Assim, pode-se dizer que um dos principais avanços conquistados são algoritmos de análise sintática de grafos *And-Or*, como por exemplo os algoritmos apresentados em (WU; XIA; ZHU, 2007) e (WU; ZHU, 2011).

Uma das principais aplicações de grafos *And-Or* é o reconhecimento de eventos em vídeo, visto que uma das vantagens de utilizar grafos *And-Or*, nas tarefas de aprendizado, representação e reconhecimento, é sua capacidade de representar objetos e eventos de forma estruturada, hierárquica e composicional. Portanto, grafos *And-Or* podem ser aplicados em problemas reais, como por exemplo recuperação de vídeo baseado no contexto, vídeo-vigilância, interação humano computador, etc.

Verificou-se também que poucos trabalhos (15%) relataram como aprender a estrutura do grafo de forma automática, e como gerar imagens a partir de dados (15%), o que pode ser considerado como uma lacuna nesta área de investigação.

A partir do exposto neste capítulo, verifica-se que, apesar do aumento de trabalhos publicados na área, a quantidade e os resultados ainda são incipientes quando comparados, por exemplo, à área de processamento de imagens. Assim, acreditamos que a área de imagens é um campo rico para uso de gramáticas em geral e, mais especificamente, de grafos *And-Or* que conseguem representar imagens estruturadas com grande sucesso.

4 Metodologia

Neste trabalho é proposto um método capaz de identificar as posições de 30 pontos de controle faciais, previamente definidos na metodologia 1 (Figura 2.8 na Seção 2.3). Isso é feito por meio da aplicação de uma sequência de técnicas de processamento de imagens em todos os componentes faciais (sobrancelhas, olhos, nariz e boca), orientadas aos bancos de imagens utilizados neste projeto. Além disso, ao processar cada conjunto de imagens de uma mesma pessoa, essa abordagem calcula a variação de cada ponto de controle para cada emoção em relação à face neutra, para representação das emoções faciais. A metodologia proposta neste projeto (metodologia 2) é apresentada na Figura 4.1 e descrita nas seções a seguir.

Figura 4.1 – Metodologia de processamento de imagens para localizar pontos de controle faciais e identificar variações desses pontos para representar emoções faciais a partir de imagens de treinamento.

Fonte: Autoria própria

4.1 Bancos de imagens

As imagens processadas foram escolhidas dos seguintes bancos de expressões faciais: *MUG Facial Expressions Database*, *ESRC 3D Face Database*, e *LAPIS Facial Expressions Database*, apresentados a seguir.

O *MUG (Multimedia Understanding Group) Facial Expressions Database* ([AIFANTI; PAPACHRISTOU; DELOPOULOS, 2010](#)), apresenta uma coleção de imagens de expressões faciais (neutra, raiva, aversão, medo, satisfação, tristeza e surpresa) capturadas em um ambiente controlado de laboratório. Esse banco inclui imagens de 86 indivíduos, entre homens e mulheres, com origem caucasiana e idades entre 20 e 35 anos. As imagens são disponibilizadas em formato JPEG, com resolução de 896x896 pixels.

Apesar de seu nome, o *ESRC 3D Face Database* ([University of Starling, 2013](#)) é um banco que contém uma coleção de imagens 2D (bidimensionais) em alta resolução (1200x1200 pixels) capturadas com um sistema de câmeras DI3D, especiais para obter imagens 3D (tridimensionais). Cinquenta pessoas, entre homens e mulheres, na faixa etária de 19 a 50 anos, participaram das sessões fotográficas. É fornecido um conjunto variado de imagens de cada participante, representando expressões de uma face neutra, satisfação, tristeza, surpresa, medo, raiva e aversão. As imagens são disponibilizadas em formato JPEG.

O *LAPIS Facial Expressions Database* é um banco de imagens de expressões faciais ainda em formação do Laboratório de Aplicações de Informática em Saúde da Universidade de São Paulo, que tem como objetivo obter imagens de expressões faciais induzidas, com o intuito de auxiliar na investigação e desenvolvimento de pesquisas na área de reconhecimento e geração de imagens faciais. Parte do trabalho desse mestrado foi dedicado tanto ao planejamento como à execução do projeto de criação desse banco, que faz parte da iniciação científica do aluno Daniel Carvalho e foi aprovado pelo comitê de ética da da Universidade de São Paulo (Anexo A). No Apêndice G apresenta-se o protocolo formal que foi elaborado para aquisição de imagens. Foram pesquisados na Internet e em fontes já padronizadas ([MIKELS et al., 2005](#)), imagens e vídeos com o potencial de causar as emoções em questão, a fim de induzir os participantes a fazerem determinada emoção. Os participantes foram convidados, por meio de mensagens eletrônicas, a participarem das sessões de gravação. Um termo de consentimento livre e esclarecido (Anexo B) é solicitado para eles assinarem antes de participarem. Em uma fase inicial, foram disponibilizadas imagens de 20 indivíduos diferentes, entre homens e mulheres de diversas etnias, na faixa etária de 18 e 25 anos. Essas imagens estão em formato JPEG e alta resolução (4300x3250 pixels).

Desses três bancos de imagens, foram selecionadas imagens de 70 sujeitos, sendo sete imagens de cada um deles (uma para cada emoção), obtendo um total de 490 imagens

a serem processadas. A seleção das imagens foi realizada de forma a atender aos seguintes requisitos: (i) possuir imagens de todas emoções (alegria, tristeza, surpresa, medo, raiva, aversão e neutra) de uma mesma pessoa considerando uma intensidade máxima, (ii) imagens sem nenhum tipo de oclusão (óculos, bigode, cabelo, etc.), (iii) e imagens de homens caucasianos na faixa etária de 25 a 30 anos de idade.

4.2 Pré processamento

As imagens dos bancos utilizados podem variar em rotação e tamanho. Essas variações podem afetar de forma significativa a localização dos pontos de controle. Um método para resolver esse problema é alinhar as imagens geometricamente, rotacionando e alterando a escala da imagem, de modo a padronizar todas as imagens de uma mesma pessoa.

Para rotacionar uma imagem, o primeiro passo é localizar uma característica fixa, invariante entre as imagens. Para isso, foram utilizadas as coordenadas das pupilas. Para localizar as pupilas, primeiro localiza-se a face nas imagens (Figura 4.2.a, detalhes descritos na seção 4.3.1). Na sequência, assumindo que a face já foi localizada, localizam-se as regiões dos olhos (Figura 4.2.b, detalhes descritos na seção 4.3.2). A partir da localização da área de cada olho, é aplicada a técnica de detecção de bordas *Canny* (Figura 4.2.c), detalhes na seção 2.2. Finalmente, um algoritmo percorre a imagem binária obtida de cada olho (Figura 4.2.d) para calcular o centróide da área formada pelas bordas do olho. As coordenadas do centróide são consideradas as coordenadas da pupila (Figura 4.2.e).

Uma vez localizadas as coordenadas dos pontos P_1 e P_2 que representam as pupilas, um segmento de reta é traçado unindo-as, como mostrado na Figura 4.3. O ângulo θ de inclinação desse segmento de reta é calculado com base na Equação 4.1, na qual x_1 e y_1 são as coordenadas do ponto P_1 e x_2 e y_2 são as coordenadas do ponto P_2 . Se esse ângulo for diferente de zero, toda a imagem é rotacionada considerando o valor do ângulo encontrado. O sentido de rotação (no sentido horário ou anti-horário) é determinado pela diferença entre as distâncias verticais ($y_1 - y_2$) das coordenadas. Se a diferença for positiva, a imagem

Figura 4.2 – Processo de localização de pupilas: (a) localização da face em uma imagem original; (b) localização da região dos olhos; (c) região dos olhos; (d) imagem binária obtida após aplicação do algoritmo *Canny*; (e) localização das pupilas.

Fonte: Autoria própria

é rotacionada no sentido horário, caso contrário, é rotacionada no sentido anti-horário. Alguns resultados de imagens rotacionadas são mostrados na Figura 4.4.

$$\theta = \arctang\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (4.1)$$

Figura 4.3 – Representação das coordenadas das pupilas.

Fonte: Autoria própria

Figura 4.4 – Exemplos da aplicação de rotação nas imagens: (a) imagens originais; (b) imagens após rotação.

Fonte: Autoria própria

Para normalizar a escala das imagens de emoções faciais de uma mesma pessoa, primeiramente são localizadas e recortadas as faces de todas as imagens de uma mesma pessoa (Figura 4.5.a); na sequência, são localizadas as coordenadas das pupilas utilizando o processo descrito anteriormente (para rotação da imagem); a distância entre as pupilas, para todas as imagens, é calculada (Figura 4.5.b); após, identifica-se a menor distância entre elas (d_{min}); por último a escala de todas as imagens é alterada proporcionalmente ao valor de d_{min} (Figura 4.5.c).

4.3 Segmentação de imagens

Uma vez que as imagens estejam normalizadas, é aplicada a elas uma sequência de técnicas com a finalidade de obter imagens segmentadas dos componentes faciais. A

Figura 4.5 – Processo de ajuste de escala das imagens de uma mesma pessoa: (a) localização das faces; (b) cálculo das distâncias entre as pupilas; (c) ajuste de escala das imagens com respeito à distância mínima entre as pupilas.

Fonte: Autoria própria

descrição dessa sequência, e os passos prévios necessários, são descritos nas subseções a seguir.

4.3.1 Detecção da face

A fim de obter a posição exata e as coordenadas do rosto nas imagens, foi utilizado o algoritmo para detecção de objetos (motivado principalmente para detecção de faces) *Haar Cascade* (VIOLA; JONES, 2001), disponível na biblioteca do *OpenCV* (HARVEY, 2011) para localizar a face em cada uma das imagens processadas.

Esse algoritmo tenta encontrar em uma imagem alguma característica da região da face que está sendo procurada (por exemplo, em uma face os olhos localizam-se sempre na parte superior). Para tal tarefa, são utilizadas as chamadas características de *Haar*¹, responsáveis por codificar informações sobre a existência de contrastes entre regiões da imagem.

¹ As características de *Haar* definem-se sobre regiões retangulares de uma imagem em tons de cinza. Uma característica é formada por um número finito de retângulos e seu valor escalar será a soma dos *pixels* de cada retângulo com um peso associado.

As faces detectadas são marcadas com retângulos vermelhos e então recortadas para análise posterior, conforme mostrado na Figura 4.6.

Figura 4.6 – (a) Localização das áreas da face. (b) Recorte das áreas das faces.

Fonte: Autoria própria

4.3.2 Localização dos componentes faciais

Primeiramente, foram definidos os componentes faciais que estavam relacionados com as unidades de ação citadas no sistema FACS, a saber: sobrancelha direita, sobrancelha esquerda, olho direito, olho esquerdo, nariz e boca. Na sequência, aplicou-se novamente o algoritmo *Haar Cascade*, para encontrar na imagem resultante da fase anterior, características que codificassem informações dos componentes faciais citados. A biblioteca do *OpenCV* possui módulos específicos (*haarcascade-eye*, *haarcascade-mcs-mouth*, *haarcascade-msc-nose*) para detectar esses componentes (olhos, boca e nariz, respectivamente). A Figura 4.7.a apresenta exemplos após aplicados esses módulos nas imagens.

A partir da localização desses componentes é estimada a região das sobrancelhas. Partindo do conceito de que as sobrancelhas encontram-se acima dos olhos, o retângulo que delimita a área das sobrancelhas é colocado acima da região dos olhos (Figura 4.7.b). As dimensões (altura e largura) desse retângulo são as mesmas dimensões da região dos olhos, mas na largura é adicionada a cada lado a décima sexta parte da largura total.

Figura 4.7 – (a) Localização dos Componentes Faciais. (b) Estimativa da área das sobrancelhas a partir da área dos olhos.

Fonte: Autoria própria

4.3.3 Segmentação dos componentes faciais

Após a identificação dos componentes faciais, uma sequência de técnicas de processamento de imagens foi aplicada para obter uma versão segmentada desses componentes. Nos seguintes parágrafos, será descrito cada um desses passos.

Primeiramente, as imagens coloridas dos componentes faciais (Figura 4.8.a), previamente localizados, são transformados em imagens em níveis de cinza. Para isso, foi utilizado o algoritmo *Color2Gray* (GOOCH et al., 2005). Este algoritmo é composto por três etapas: i) converte-se imagens de entrada *RGB* para um espaço de cores *CIE-L*a*b**², ii) calculam-se as distâncias escalares de luminância³ e crominância⁴ para cada par de pixels vizinhos na imagem, e iii) resolve-se um problema de otimização desenhado para

² A CIE - *Commission Internationale Eclairage* é uma organização sem fins lucrativos considerada como a autoridade na ciência de luz e cor, a qual definiu três espaços de cores. Um deles é o espaço de cor *L*a*b**, no qual *L** indica a luminosidade e *a** (coordenada vermelho/verde) e *b** (coordenada amarelo/azul) são as coordenadas cromáticas.

³ Luminância é uma medida da densidade da intensidade de uma luz refletida em uma dada direção.

⁴ A crominância é um dos dois elementos que conformam um sinal de vídeo. Refere-se ao valor das cores.

modular seletivamente a representação em tons de cinza como uma função da variação de cromaticidade da imagem.

Na sequência, um filtro gaussiano (como detalhado na Seção 2.2) é aplicado nas imagens em níveis de cinza, com o intuito de suavizar as imagens (reduzir os ruídos). Os valores dos parâmetros utilizados foram $\sigma = 1$ (desvio padrão) e 5x5 pixels o tamanho da máscara.

Com base nos histogramas das imagens resultantes da fase anterior (imagens suavizadas), aplicou-se a técnica de *limiarização multiníveis* (descrita com maior detalhe na Seção 2.2), com o objetivo de detectar as bordas dos componentes faciais. Os valores dos limiares T_1 e T_2 variam entre 95 a 115, e 235 a 255, respectivamente. Esses valores foram calibrados em uma fase inicial. Uma vez estabelecidos os valores desses parâmetros, os mesmos foram utilizados de forma automática.

Algoritmos de detecção de bordas são tipicamente seguidos de procedimentos de ligação e de outros procedimentos de detecção de fronteiras, desenvolvidos para juntar e organizar os pixels de borda em fronteiras significativas. Várias técnicas são apropriadas para este propósito. A técnica utilizada neste projeto foi *processamento local*. Esta técnica consiste na análise das características dos pixels em uma pequena vizinhança (por exemplo, 3x3) em torno de cada ponto (x,y) . Todos os pontos que forem similares são ligados, formando uma fronteira de pixels que compartilham alguma propriedade em comum.

Na sequência, as imagens resultantes foram binarizadas (limiarização com um limiar único) com objetivo de isolar a área dos componentes faciais respeito ao fundo da imagem. Finalmente a transformação morfológica de fechamento (descrito na Seção 2.2) é utilizada com o intuito de preencher os espaços vazios dentro da área do componente facial.

Alguns resultados após a aplicação da sequência de técnicas no olho direito, sobrancelha direita e nariz são apresentados nas Figuras 4.8, 4.9 e 4.10, respectivamente.

4.4 Cálculo dos pontos de controle e de suas variações entre diferentes emoções

4.4.1 Localização dos pontos de controle

Nesta seção são descritos os algoritmos de localização dos pontos de controle para cada componente facial. Foi utilizada a mesma nomenclatura da antropometria definida

Figura 4.8 – Sequência de técnicas de processamento de imagens aplicada à região do olho direito: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.

Fonte: Autoria própria

Figura 4.9 – Sequência de técnicas de processamento de imagens aplicada à região da sobrancelha: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.

Fonte: Autoria própria

Figura 4.10 – Sequência de técnicas de processamento de imagens aplicada à região do nariz: (a) imagem original; (b) imagem após transformação para níveis de cinza e suavização com filtro gaussiano; (c) detecção de bordas com limiarização; (d) imagem após aplicação da segmentação por descontinuidade; (e) imagem binarizada; (f) imagem após aplicação da morfologia matemática.

Fonte: Autoria própria

em (TESTA et al., 2015). Uma descrição mais detalhada para cada ponto é encontrada no Apêndice B. Assim, a distância entre dois pontos é denotada pelo símbolo " " (traço inferior). Deste modo, se um traço inferior está entre dois pontos, significa que está sendo considerada a distância entre eles. Por exemplo, ex_en denota a distância euclidiana entre os pontos ex e en .

Região dos olhos: Cinco pontos de controle, a saber, ex , en , ps , pi e $pupil$ (Figura 2.8), são localizados em cada olho.

Um algoritmo percorre, em ordem crescente de coluna, a imagem binária do olho direito procurando a menor e a maior coluna que tem um pixel preto, e os cantos externo e interno do olho direito (ex e en , respectivamente). Para localizar os cantos superior e inferior (ps e pi), o mesmo algoritmo percorre a imagem em ordem crescente de linha procurando a menor e a maior linha que tem um pixel preto, respectivamente. O ponto $pupil$ é localizado como o pixel de intersecção entre as linhas ex_en e ps_pi .

A mesma lógica do algoritmo é utilizada na imagem binária do olho esquerdo, com a diferença que, para localizar os pontos *en* e *ex*, percorre-se a imagem procurando a menor e a maior coluna que tem um pixel preto (*en* e *ex*, respectivamente).

Região das sobrancelhas: Cinco pontos de controle, a saber, *p1*, *pas1*, *os*, *p2* e *pas2* (Figura 2.8), são localizados em cada sobrancelha.

Um algoritmo percorre, em ordem crescente de coluna, a imagem binária da sobrancelha direita procurando a menor e a maior coluna que tem um pixel preto (pontos *p2* e *p1*, respectivamente). Para localizar o ponto *os*, o mesmo algoritmo percorre a imagem, em ordem crescente de linha, procurando a menor linha que tem um pixel preto. Por sua vez, os pontos *pas1* e *pas2* são estimados a partir dos pontos *p2*, *p1*, e *os*: *pas1* define-se como o ponto central da linha que une os pontos *p1* e *os*. De forma semelhante, *pas2* define-se como o ponto central da linha que une os pontos *p2* e *os*.

A mesma lógica do algoritmo é utilizada na imagem binária da sobrancelha esquerda, com o diferencial que para localizar os pontos *p1* e *p2* percorre-se a imagem procurando a menor e a maior coluna que tem um pixel preto (*p1* e *p2*, respectivamente).

Região do nariz: Nesta região são localizados cinco pontos de controle, a saber: *aldir*, *alesq*, *acdir*, *acesq* e *sn* (Figura 2.8).

Na imagem binária do nariz pode-se observar que é detectada a área inferior do nariz, onde encontram-se as narinas. Um algoritmo percorre essa imagem, em ordem crescente de coluna, procurando a menor e a maior coluna que tem um pixel preto (pontos *aldir* e *alesq*, respectivamente) (Figura 4.11.a).

Um ponto auxiliar *aux* é definido, como o ponto central da linha *aldir_alesq* (Figura 4.11.b). A partir desse ponto, duas regiões *A* e *B* são definidas (Figura 4.11.c). Para localizar o ponto *acdir*, percorre-se a imagem da região *A*, em ordem crescente de linha, procurando a maior linha que tem um pixel preto. De forma semelhante, é localizado o ponto *acesq* na região *B*. Por último, o ponto central da linha *acdir_acesq* é definido como sendo a posição do ponto *sn*. É importante frisar que os pontos *mfesq* e *mfdir* não foram localizados na metodologia 2 (maiores detalhes são descritos na Seção 6.2).

Região da boca: Nesta região são localizados cinco pontos de controle, a saber: *chdir*, *chesq*, *ls*, *li* e *sto* (Figura 2.8). Um algoritmo percorre, em ordem crescente de coluna, a imagem binária da boca procurando a menor e maior coluna que tem um pixel preto,

Figura 4.11 – Localização dos pontos de controle na imagem binária do nariz.

Fonte: Autoria própria

pontos *chdir* e *chesq* respectivamente. O ponto *sto* é definido como o ponto central da linha *chdir_chesq*. Na mesma coluna onde encontra-se o ponto *sto*, procura-se a menor e a maior linha com pixels pretos. Eles são definidos como os pontos *ls* e *li*, respectivamente.

Na Figura 4.12 pode-se observar exemplos de localização dos pontos de controle. Esses pontos são colocados pelo algoritmo desenvolvido de processamento de imagens proposta neste projeto (metodologia 2).

Figura 4.12 – Exemplos de localização dos pontos de controle da metodologia 2.

Fonte: Autoria própria

4.4.2 Identificação das variações dos pontos de controle

No presente projeto, para representar a movimentação dos pontos de controle em cada emoção em relação às suas posições originais na face neutra, também empregou-se o conceito de envoltório (descrito na seção 2.3), seguindo os passos:

- um envoltório é definido ao redor de cada uma das curvas utilizadas para desenhar a caricatura da face neutra. Essas curvas são estimadas utilizando os posicionamentos de seus pontos de controle (Figura 4.13);
- na imagem de uma dada emoção, para cada ponto de controle é calculada a variação percentual (valor representado entre 0 e 1) de seu posicionamento em relação ao seu posicionamento original na face neutra em termos do tamanho do envoltório que o contém.
- no final, a variação de um ponto de controle, da face neutra para uma determinada emoção, corresponde à média de todos os valores de variação daquele ponto em todos os pares de imagens (face neutra, face da emoção sendo considerada) processados.

Figura 4.13 – (a) Envoltório da curva superior do olho direito, tendo essa curva sido estimada a partir dos pontos de controle que a compõem, a saber *ex*, *ps* e *en*. (b) Envoltório da curva inferior do olho direito, tendo essa curva sido estimada a partir dos pontos de controle que a compõem, a saber *ex*, *pi* e *en*.

Fonte: Autoria própria

4.5 Comparação entre as duas metodologias

As seções anteriores descreveram uma metodologia para estimar parâmetros de variação dos pontos de controle de cada emoção em relação à face neutra aprendidos a partir de imagens de treinamento, como uma alternativa à metodologia 1 de utilizar tais parâmetros definidos empiricamente.

Para poder comparar as duas metodologias quanto à qualidade da representação de cada emoção, foram construídos formulários *online*. As imagens foram aleatoriamente distribuídas gerando dez versões diferentes de formulários.

Os formulários continham imagens misturadas das duas metodologias, sendo disponibilizada uma única imagem em cada página a fim de evitar que o avaliador se confundisse. Ao lado de cada imagem foram disponibilizadas as opções de sete emoções faciais (neutra, satisfação, tristeza, aversão, surpresa, medo e raiva), sendo solicitado ao avaliador que selecionasse aquela que ele considerasse que é representada na imagem⁵.

Cada formulário possuía três imagens com diferentes níveis de intensidade (25%, 50% e 100%) de cada uma das seis emoções das duas metodologias, mais uma face neutra. Como mecanismo de conferência, foram repetidas três imagens com intensidade 50%, das emoções de satisfação, tristeza e surpresa, de cada metodologia, em posições aleatórias, a fim de permitir estimar o nível de coerência de cada avaliador. Cada formulário totalizou, assim, 43 imagens.

A participação dos avaliadores foi voluntária e anônima. Os formulários foram distribuídos por mensagens eletrônicas pelos participantes do projeto a pessoas consideradas saudáveis. Os dados de idade e sexo também foram coletados com o intuito de permitir verificar eventuais diferenças de percepção em função dessas variáveis.

Após a coleta dos formulários, os resultados foram analisados e discutidos, conforme apresentado no capítulo a seguir.

⁵ Um formulário de demonstração está disponível em: <http://goo.gl/forms/gJJGa7OOoE>

5 Resultados e Discussões

Neste capítulo são apresentados os resultados e discussões da metodologia 2 neste projeto. Na Seção 5.1 são discutidos os resultados do reconhecimento das caricaturas a partir do experimento conduzido com voluntários. A coerência das respostas desse experimento é descrita na Seção 5.2. A metodologia de processamento de imagens da metodologia 2 é avaliada na Seção 5.3. Finalmente, na Seção 5.4 são apresentadas as considerações finais.

5.1 Reconhecimento das caricaturas geradas

Com a finalidade de permitir a comparação das caricaturas geradas por ambas metodologias, as Figuras 5.1 a 5.6 apresentam as caricaturas geradas pelas duas metodologias para cada uma das emoções faciais e suas diferentes intensidades (25%, 50% e 100%). Para cada emoção e intensidade, a imagem do lado esquerdo é a gerada pela metodologia 1 e a imagem do lado direito é a gerada neste projeto (metodologia 2). A face neutra, por ser independente de metodologia e intensidade, é apresentada na Figura 5.7.

Conforme apresentado na Seção 4.5, foram convidados voluntários para responderem o questionário a respeito do reconhecimento das emoções, com o objetivo de avaliar a qualidade da representação das emoções nas imagens geradas por ambas metodologias.

As Tabelas 1 e 2 apresentam as matrizes de confusão do reconhecimento das emoções para as duas metodologias, metodologia 1 e metodologia 2, respectivamente. Em cada linha é apresentado como as imagens de uma dada emoção, de todas as intensidades testadas, foram classificadas pelos voluntários. Mais especificamente, cada posição (i, j) fornece a percentagem das imagens da i -ésima emoção que foram classificadas como sendo da j -ésima emoção. Desta forma, a diagonal principal fornece as taxas de acerto, enquanto as demais posições as taxas de erro.

Uma comparação apenas dos valores das diagonais principais das duas tabelas de confusão pode ser melhor visualizada na Figura 5.8. Nesta figura, pode-se observar que a taxa de reconhecimento das imagens geradas com a metodologia 1, em relação às imagens geradas com a metodologia 2, foi inferior para cinco emoções (medo, raiva, satisfação, surpresa e tristeza), igual para a face neutra e superior para uma emoção (aversão).

Figura 5.1 – Expressões faciais de satisfação geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

Figura 5.2 – Expressões faciais de tristeza geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

A emoção de **satisfação** foi a que apresentou a maior queda da taxa de reconhecimento na metodologia 2: 38 pontos percentuais de diferença (queda de 92% para 54%). Ao

Figura 5.3 – Expressões faciais de medo geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

analisar a Figura 5.1 verifica-se que, à medida que aumenta a intensidade, o arqueamento para cima da boca é maior nas caricaturas geradas pela metodologia 1 do que as geradas

Figura 5.4 – Expressões faciais de surpresa geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

pela metodologia 2. Nas imagens geradas na metodologia 2, esse arqueamento foi mais sutil devido à presença, na amostra de treinamento, de imagens de sujeitos cujas expressões de

Figura 5.5 – Expressões faciais de raiva geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

satisfação eram muito sutis, e em alguns casos parecem até não representar a emoção de satisfação (Figura 5.9).

Figura 5.6 – Expressões faciais de aversão geradas por ambas as metodologias para as três intensidades consideradas.

Fonte: Autoria própria

Figura 5.7 – Expressão facial neutra.

Fonte: Autoria própria

Tabela 1 – Metodologia 1: matriz de confusão das respostas obtidas. Em negrito estão as taxas de acerto.

	Aversão	Medo	Neutra	Raiva	Satisfação	Surpresa	Tristeza
Aversão	62%	1%	2%	33%	1%	1%	0%
Medo	6%	33%	25%	10%	6%	11%	9%
Neutra	1%	1%	91%	1%	6%	0%	0%
Raiva	5%	8%	24%	46%	12%	4%	2%
Satisfação	0%	0%	6%	0%	92%	1%	0%
Surpresa	1%	6%	14%	0%	7%	69%	2%
Tristeza	2%	7%	6%	1%	0%	2%	83%

Fonte: Autoria própria

Verifica-se, ainda, que a emoção satisfação foi majoritariamente confundida com a emoção neutra (6% na metodologia 1 e 41% na metodologia 2). Observando a Figura 5.1, as caricaturas geradas pela metodologia 2 diferem da emoção neutra (Figura 5.7) no arqueamento da boca e na inclusão de rugas nos contornos dos olhos. No entanto, essas alterações, especialmente na boca, são muito sutis nas intensidades de 25 e 50%, fornecendo indícios de que essas alterações sutis dificultam o reconhecimento.

A segunda maior queda da taxa de reconhecimento ocorreu para a emoção de **tristeza**: 24 pontos percentuais de diferença (queda de 83% para 59%). Quando comparadas as caricaturas geradas por ambas as metodologias (Figura 5.2), pode-se perceber que, na metodologia 1, à medida que aumenta a intensidade da emoção, os cantos internos das sobrancelhas apresentam um esticamento para cima, mais pronunciado na intensidade de

Tabela 2 – Metodologia 2: matriz de confusão das respostas obtidas. Em negrito estão as taxas de acerto.

	Aversão	Medo	Neutra	Raiva	Satisfação	Surpresa	Tristeza
Aversão	73%	1%	2%	21%	2%	1%	0%
Medo	7%	10%	12%	56%	1%	10%	4%
Neutra	1%	1%	91%	1%	6%	0%	0%
Raiva	8%	4%	24%	43%	3%	1%	17%
Satisfação	0%	1%	41%	1%	54%	1%	2%
Surpresa	6%	8%	17%	20%	2%	46%	1%
Tristeza	5%	8%	10%	15%	0%	3%	59%

Fonte: Autoria própria

Figura 5.8 – Taxas de acerto das emoções em cada metodologia.

Fonte: Autoria própria

100%, enquanto as sobrancelhas da metodologia 2 não se mexem. Isso sugere que essa característica exerce um importante papel no reconhecimento da emoção citada. Logo, há indícios de que as técnicas de processamento de imagens podem ter apresentado falhas no processamento das sobrancelhas ou, ainda, que as imagens dos bancos utilizadas podem ter influenciado negativamente a geração deste componente nas caricaturas, como discutido mais adiante.

Tanto a emoção **medo** quanto **surpresa** tiveram uma queda de 23 pontos percentuais na taxa de reconhecimento. No entanto, a emoção **medo** foi a que apresentou a menor taxa de reconhecimento nas duas metodologias: 33% na metodologia 1 e 10% na metodologia 2. Enquanto na metodologia 1 foi mais confundida com a emoção neutra

Figura 5.9 – Exemplos de faces sutis representando a emoção satisfação no banco de dados.

Fonte: Autoria própria

(25%), na metodologia 2 foi confundida majoritariamente com a emoção raiva (56%) (Tabelas 1 e 2). Ao analisar as imagens de emoção medo geradas pela metodologia 2 (coluna direita da Figura 5.3), verifica-se que, à medida que a intensidade aumenta, as sobrancelhas tendem a fazer uma curva em forma de “S” com pronunciado arqueamento para baixo nos cantos internos das sobrancelhas. Essa mesma característica é observada também nas caricaturas da emoção **surpresa** geradas pela metodologia 2 (coluna direita da Figura 5.4). De fato, a emoção de surpresa gerada pela metodologia 2 também foi mais confundida com a emoção raiva (20%, sexta linha da Tabela 2).

A emoção **raiva** obteve taxas de reconhecimento semelhantes nas duas metodologias (1 e 2), com valores de 46% e 43%, respectivamente. Verifica-se, ainda, que nos dois casos, a emoção raiva foi mais confundida com a face neutra (24% nas duas metodologias - quarta linha nas Tabelas 1 e 2). Em segundo lugar, esta foi mais confundida com a emoção de satisfação na metodologia 1 (12%) e com a emoção de tristeza na metodologia 2 (17%). Na Figura 5.5, pode-se perceber que as partes superiores das faces são parecidas mas, à medida que aumenta a intensidade da emoção raiva, cresce o arqueamento para baixo da boca nas caricaturas geradas pela metodologia 2, o que explica uma maior confusão com a emoção de tristeza. Isso sugere que a forma atual de desenho das curvas que representam a boca pode ser inadequada ou insuficiente para representar a emoção raiva e que outros pontos

deveriam ser adicionados à caricatura. Considerando as imagens usadas, verifica-se que há pontos em torno da boca que se enrugam, mas que não foram considerados na metodologia 2. Esses poderiam ser futuramente processados a fim de melhorar a representação desta emoção.

Apesar da queda das taxas de reconhecimento de cinco emoções, foi obtida uma melhora na taxa de reconhecimento da emoção de **aversão**, subindo de 62% para 73%. Ao analisar as imagens das duas metodologias (Figura 5.6), percebe-se que, no caso das caricaturas geradas pela metodologia 2, principalmente para uma intensidade de 100%, os componentes faciais (olhos, sobrancelhas, nariz e boca) juntam-se mais para o meio da face, e a nariz torna-se mais pontiagudo na parte inferior. Isso sugere que os voluntários relacionam essas características com a emoção aversão. Essas diferenças entre as caricaturas de 100% de aversão das duas metodologias ocorrem justamente devido a variações em pontos de controles não considerados pelas Unidades de Ação definidas pelo FACS para a emoção aversão (metodologia 1). Logo, a ideia de aprender as variações de todos os pontos de controle, metodologia 2, melhora a representação da emoção citada e aumenta a taxa de reconhecimento.

A Figura 5.10 apresenta as taxas de acerto para cada emoção da metodologia 1 e da metodologia 2 separadas por intensidade. Cada curva representa a percentagem de acertos para cada uma das intensidades (25%, 50% e 100%) consideradas de cada emoção.

Pode-se perceber que, na metodologia 1, duas emoções apresentaram taxas de reconhecimento constantes para todas as intensidades (neutra e aversão) e o restante das emoções apresentaram aumento na taxa de reconhecimento com o aumento da intensidade (Figura 5.10). Já na metodologia 2, uma emoção permaneceu praticamente constante (medo, que decresceu apenas em 1% da intensidade de 25% para 100%) e uma emoção, surpresa, apresentou uma queda de reconhecimento com o aumento da intensidade. A emoção surpresa decresceu em 4% (de 47% para 43%, da intensidade de 25% para 100%). Como discutido linhas acima, ao analisar as Figuras 5.3 e 5.4 percebe-se que, à medida que aumenta a intensidade das emoções, as caricaturas de medo e surpresa geradas pela metodologia 2 tendem a incrementar o arqueamento das sobrancelhas em forma de “S”, o que faz confundir essas emoções com a emoção raiva.

Verifica-se, também, que à medida que a intensidade aumenta a taxa de reconhecimento da emoção aversão manteve-se constante em 61% na metodologia 1, e aumentou de 70% para 76% na metodologia 2. Ao analisar a Figura 5.6, percebe-se que, nas caricaturas

Figura 5.10 – Taxas de acerto para cada emoção da metodologia 1 (linhas retas) e da metodologia 2 (linhas tracejadas).

Fonte: Autoria própria

geradas pela metodologia 1 a parte superior (olhos e sobrancelhas) e central (nariz) da face são muito parecidas, já na parte inferior (boca) tem-se um incremento sutil na abertura da boca. Por sua vez, nas caricaturas geradas pela metodologia 2, percebe-se que, à medida que a intensidade aumenta a nariz pontua-se em baixo dando a sensação do nariz ter subido ou enrugado mais.

5.2 Coerência das respostas

Considerando que a resposta dos voluntários poderia sofrer influência de fatores como fadiga ou emissão de respostas aleatórias, uma análise foi conduzida sobre as respostas a fim de verificar a coerência das mesmas. Conforme descrito na Seção 4.5, seis imagens foram duplicadas nos formulários para verificar a coerência das respostas de um mesmo voluntário, repetindo três imagens com intensidade 50% (das emoções de satisfação, tristeza e surpresa) de cada metodologia.

Um resultado foi considerado coerente quando o avaliador indicou a mesma emoção para duas caricaturas idênticas, mesmo que esta emoção não correspondesse àquela gerada pela metodologia 2. A coerência de um indivíduo é aqui definida como sendo a percentagem desses seis pares de imagens duplicadas que foram classificadas pelo indivíduo de forma

coerente. A Tabela 3 apresenta a percentagem de indivíduos que atingiram determinada faixa de coerência.

Tabela 3 – Percentagem de indivíduos que atingiram determinada faixa de coerência referente às imagens duplicadas em ambas metodologias.

	<25%	26 a 50%	51 a 75%	>75%
Metodologia 1	2%	6%	40%	53%
Metodologia 2	6%	33%	35%	26%

Fonte: Autoria própria

A Tabela 3 indica que, enquanto 93% dos voluntários atingiram mais de 50% de coerência na metodologia 1, apenas 61% atingiram a mesma faixa de coerência na metodologia 2, apresentando assim uma queda de 32 pontos percentuais. Considerando especificamente a faixa de coerência acima de 75%, verifica-se uma queda de 53% para 26% (27 pontos percentuais) dos voluntários que atingiram essa faixa. Assim, há indícios que os voluntários estão se confundindo mais nas caricaturas geradas pela metodologia 2. Logo, essas caricaturas não estão simplesmente dando a ideia de uma emoção errada (por exemplo, representando consistentemente a emoção raiva quando deveria ser medo), mas também não estão representando claramente algumas das emoções.

5.3 Avaliação da metodologia de processamento de imagens

Um problema recorrente presente na avaliação das caricaturas (Seção 5.1) foi na correta variação das sobrancelhas nas várias emoções, ou seja, em várias emoções as sobrancelhas das caricaturas geradas a partir da metodologia 2 não foram condizentes com a emoção. Por exemplo, as caricaturas de medo têm sobrancelhas que induzem o reconhecimento de raiva. A única emoção que apresentou aumento na taxa de acerto com a metodologia 2 foi justamente uma que não parece depender da variação das sobrancelhas (aversão). Algumas possíveis causas desse problema são descritas nos seguintes parágrafos.

Como descrito na Seção 4.3.2, a metodologia 2 não identifica a região propriamente dita na qual estão as sobrancelhas, mas estima a localização dessa região assumindo que ela encontra-se acima da região dos olhos (Figura 5.11.a). Dessa forma, o retângulo que define essa área pode não conter toda a sobrancelha ou não separar corretamente as sobrancelhas dos olhos (Figura 5.11.b). Logo, a localização dos pontos de controle das sobrancelhas pode não ser adequada (Figura 5.11.c).

Figura 5.11 – (a) Estimação da área das sobrancelhas (retângulo verde) a partir da localização da área dos olhos (retângulo amarelo). (b) Imagem binária da sobrancelha direita. (c) Localização dos pontos de controle da face.

Fonte: Autoria própria

Além do problema da estimação da região das sobrancelhas, detectou-se também um outro problema, agora no reconhecimento das sobrancelhas contidas na região estimada. Ao analisar as imagens de treinamento, verificou-se que, alguns sujeitos possuem sobrancelhas muito claras (contraste), dificultando assim o processo de segmentação (Figura 5.12.a e 5.12.b). Nestes casos, a localização dos pontos de controle das sobrancelhas é comprometida (Figura 5.12.c).

Figura 5.12 – (a) Imagem segmentada (binária) da sobrancelha direita. (b) Imagem segmentada (binária) da sobrancelha esquerda. (c) Localização dos pontos de controle da face.

Fonte: Autoria própria

Um caso especial são as sobrancelhas das caricaturas de medo e surpresa (coluna direita das Figuras 5.3 e 5.4, respectivamente), as quais têm um arqueamento na forma de uma “S” mais pronunciado nos cantos internos, que fazem essas duas emoções parecerem raiva. Na Seção 4.4.1 é descrito, que dentre os 5 pontos de controle definidos nas sobrancelhas, dois deles (*pas1* e *pas2*) são estimados a partir da localização dos outros três pontos. Quando comparados os valores das variações desses dois pontos de controle com os valores empíricos definidos na metodologia 1, verifica-se que o valor estimado é quase o dobro do empírico. É justamente esse valor elevado nas variações estimadas dos pontos de controle *pas1* e *pas2*, que fazem desenhar as curvas das sobrancelhas de forma inconsistente. Logo, a estimação desses pontos pode não representar fielmente as características das imagens reais. Esse valor é elevado porque as posições dos outros três pontos de controle, utilizados para estimar *pas1* e *pas2*, variam significativamente com respeito à face neutra. Uma possível forma de contornar esse problema é definir regiões que possam facilitar a localização desses pontos, como foi feito para o componente facial da boca (Seção 4.4.1). A definição dessas regiões faria com que os pontos de controle fossem localizados de forma mais adequada (nas extremidades das sobrancelhas), minimizando o problema de falhas da segmentação. Como discutido, tais falhas ocorrem principalmente em relação aos indivíduos com sobrancelhas muito claras e/ou com sobrancelhas apresentando imperfeições que causam descontinuidades na imagem.

Outro problema encontrado nas caricaturas geradas pela metodologia 2 foi com relação à boca. Em emoções como surpresa e medo (Figuras 5.4 e 5.3), o lábio superior tem um arqueamento para cima mais pronunciado do que o arqueamento para baixo do lábio inferior. Já na metodologia 1, o arqueamento dos dois lábios são mais simétricos. Ao analisar as imagens de treinamento, percebe-se que os sujeitos fazem uma abertura da boca maior, quase simétrica em relação à parte superior e inferior, nessas emoções. O fato de a boca apresentar uma abertura maior com respeito ao eixo vertical, faz com que o tamanho do retângulo que delimita a sua área durante a localização desse componente facial seja também maior no eixo vertical e, por isso, em muitos casos esse retângulo contém também uma pequena área da nariz. Logo, as imagens binárias da boca apresentam ruídos na parte superior, o que traz erros na localização dos pontos de controle. Esse problema poderá ser abordado adicionando-se outras técnicas de processamento de imagens para eliminação de ruídos ou aperfeiçoando-se o processamento já empregado que visa a eliminar ruídos (por exemplo, as operações já empregadas da morfologia matemática).

Como discutido na Seção 5.1, as caricaturas da emoção satisfação geradas pela metodologia 2 apresentam variações muito sutis na boca à medida que aumenta a intensidade (Figura 5.1). As imagens de treinamento indicam que quando pede-se para uma pessoa comum (não ator) fazer determinadas emoções, talvez ela se expresse de uma forma muito mais sutil do que as expressões representadas na metodologia 1. A presença de indivíduos com emoções sutis na base de imagens pode gerar caricaturas sutis, uma vez que a metodologia 2 encontra os pontos de controle por meio do cálculo de média. Para exemplificar esse problema, mais fortemente observado na emoção de satisfação, a Figura 5.13 apresenta caricaturas geradas para pessoas de forma individual, algumas mais expressivas, outras mais sutis.

5.4 Considerações Finais

O tipo de avaliação utilizado neste trabalho foi o mesmo que utilizou-se em (TESTA et al., 2015). A avaliação das imagens contou com 163 voluntários supostamente saudáveis, cujas idades variaram de 15 a 70 anos, sendo 53% homens e 47% mulheres do total de avaliadores. Foram considerados três grupos de acordo com a faixa etária: menos de 25, 26 a 55 e maior de 55 anos. O percentual de participantes em cada grupo foi 30% (49 voluntários), 65% (106 voluntários) e 5% (5 voluntários), respectivamente.

As Tabelas 4 a 7 apresentam as taxas de acerto para cada emoção considerando os gêneros feminino e masculino, e os três grupos de idade, para cada metodologia.

Tabela 4 – Taxa de acerto para cada gênero da metodologia 1.

	Satisfação	Tristeza	Medo	Surpresa	Raiva	Aversão	Neutra
Masculino	92%	83%	31%	67%	44%	61%	91%
Feminino	92%	83%	29%	72%	48%	63%	92%

Fonte: Autoria própria

Tabela 5 – Taxa de acerto para cada gênero da metodologia 2.

	Satisfação	Tristeza	Medo	Surpresa	Raiva	Aversão	Neutra
Masculino	52%	53%	7%	42%	43%	71%	91%
Feminino	56%	66%	13%	49%	42%	75%	92%

Fonte: Autoria própria

A partir das Tabelas 4 e 5, verifica-se que, as taxas de acerto das emoções são parecidas nas duas metodologias, mas há indícios que as mulheres reconhecem mais as emoções surpresa e aversão.

Tabela 6 – Taxa de acerto para cada faixa etária da metodologia 1.

	Satisfação	Tristeza	Medo	Surpresa	Raiva	Aversão	Neutra
<25 anos	91%	84%	50%	78%	56%	64%	92%
26 a 55 anos	93%	84%	22%	65%	41%	60%	92%
>55 anos	94%	69%	17%	72%	46%	67%	75%

Fonte: Autoria própria

Tabela 7 – Taxa de acerto para cada faixa etária da metodologia 2.

	Satisfação	Tristeza	Medo	Surpresa	Raiva	Aversão	Neutra
<25 anos	57%	59%	14%	57%	48%	78%	92%
26 a 55 anos	51%	58%	7%	39%	42%	72%	92%
>55 anos	69%	78%	13%	66%	25%	58%	75%

Fonte: Autoria própria

A partir das Tabelas 6 e 7, verifica-se que, as taxas de acerto das emoções são parecidas nas duas metodologias, mas há indícios que os mais jovens (<25 anos) reconhecem mais as emoções medo e raiva, e o mais velhos (>55 anos) reconhecem mais a emoções satisfação.

Figura 5.13 – Exemplos de caricaturas geradas individualmente pela metodologia 2. Na coluna esquerda encontra-se as imagens originais dos sujeitos para emoção satisfação. Na coluna direita encontra-se sua correspondente caricatura gerada pela metodologia 2.

Fonte: Autoria própria

6 Conclusões

6.1 Principais resultados obtidos e contribuições

Este projeto define, implementa e avalia uma abordagem que gera caricaturas de várias emoções faciais (satisfação, tristeza, surpresa, medo, aversão, raiva e neutra) utilizando parâmetros aprendidos a partir de imagens de treinamento. Para isso foi desenvolvida uma metodologia de processamento de imagens que localiza pontos de controle faciais em imagens de pessoas expressando as diferentes emoções, e aprende as variações dos pontos de controle, da face neutra para uma determinada emoção, para caracterização das expressões faciais.

Em uma abordagem anterior (metodologia 1), essas caricaturas são geradas utilizando os conceitos de antropometria e FACS sendo as variações dos pontos de controle de uma face neutra para uma determinada emoção definidos de forma empírica. Com o intuito de melhorar essas caricaturas, a proposta apresentada neste projeto consiste em: i) aprender as variações dos pontos de controle a partir de imagens de pessoas reais, e ii) variar potencialmente todos os pontos de controle de todos os componentes faciais em cada emoção, já que pessoas reais normalmente não conseguem isolar apenas alguns músculos faciais em uma expressão.

Para poder comparar a fidelidade da representação da emoção desejada nas imagens geradas, com respeito à metodologia 1, um experimento com voluntários foi conduzido. Os resultados mostraram que a taxa de reconhecimento das imagens geradas com a metodologia 2, em relação às imagens geradas com a metodologia 1, foi inferior em cinco emoções (medo, raiva, satisfação, surpresa e tristeza), igual para a face neutra e superior para uma emoção (aversão).

Para a emoção de aversão, cuja taxa de reconhecimento foi superior na metodologia 2 neste projeto, conseguiu-se aprender variações significativas de pontos de controle que não estão presentes nas unidades de ação que o FACS prevê para essa emoção, como os pontos de controle localizados no nariz (Figura 5.6). Esses dados sugerem que a proposta de aprender as variações de todos os pontos de controle da face pode ajudar em representar melhor as emoções faciais. Para as emoções cuja taxa de reconhecimento foi inferior em

relação à metodologia 1, foram identificadas possíveis causas tanto na seleção da amostra de treinamento quanto nas etapas de processamento de imagens. Desta forma, esses resultados inferiores não invalidam a proposta desenvolvida neste projeto, mas indicam que melhorias devem ser desenvolvidas em algumas etapas do processo.

A geração de imagens sintéticas pode contribuir para imprimir flexibilidade às ferramentas computacionais para o contexto deste projeto (jogo para auxílio ao diagnóstico e terapia). Pacientes diferentes podem apresentar necessidades diferentes para terapia. Por exemplo, um paciente pode reconhecer emoções de satisfação com facilidade, mas não consegue reconhecer emoções de raiva se esta não estiver muito intensa. Em geral, as bases de imagens fornecem imagens com emoções induzidas por pessoas. Estudos indicam que alguns músculos da face não conseguem ser movidos voluntariamente. Adicionalmente, não há como controlar a intensidade da emoção ao adquirir uma imagem de um indivíduo. Por exemplo, não há meios de definir, de forma objetiva, como uma pessoa deveria fazer uma expressão de 75% de aversão. Assim, ainda que bases de imagens sejam construídas com metodologias adequadas, as limitações humanas impedem que imagens reais representem certas características desejáveis para diagnóstico e treinamento de pacientes. A geração de imagens (caricaturas, no caso deste projeto) pode contribuir para o terapeuta obtenha imagens com as características adequadas para cada paciente.

Sendo assim, as principais contribuições deste trabalho são: a proposta e a implementação de uma nova abordagem de geração de caricaturas com indícios de possibilitar uma melhor representação das emoções, principalmente após melhorias sugeridas em 6.2; e uma revisão sistemática sobre o uso de grafos *And-Or* para geração e reconhecimento de imagens. Vale a pena destacar que uma das contribuições mais importantes deste projeto é a formação do banco de imagens *LApIS Database*, que contém conjuntos de imagens de faces de uma mesma pessoa expressando mais de uma emoção facial como descrito na seção 4.1.

Além do mais, a integração com a gramática permitirá uma flexibilização do modelo, e, desta maneira, a evolução do trabalho no futuro se dará de uma forma mais dinâmica, conforme se agregar mais variáveis como etnia, faixa etária e gênero.

Por fim, cabe mencionar que parte do esforço inicial deste projeto foi dedicado também à participação da concepção das ideias presentes no projeto geral (que contempla as duas metodologias). O artigo (TESTA et al., 2015), publicado durante este mestrado,

é fruto deste trabalho (Anexo ??). Além deste, um segundo artigo especificamente da revisão sistemática está em fase de submissão.

6.2 Trabalhos futuros

O componente facial que mais necessita de ajustes na geração das caricaturas é o par de sobrancelhas. O fato das sobrancelhas serem desenhadas utilizando uma única curva *spline* pode não representar de forma adequada suas variações nas diferentes emoções faciais, quando são aprendidas a partir de imagens reais. Outra variável a ser considerada é a quantidade de pontos de controle que definiria cada curva deste componente. Embora este fato possa parecer uma limitação da abordagem, essas características podem ser facilmente alteradas na metodologia 2. Um trabalho futuro, portanto, é desenhá-las utilizando mais curvas (por exemplo deixando-as mais grossas) e mais pontos de controle, no intuito de melhorar o desenho e a representação dos movimentos das sobrancelhas nas caricaturas.

Verificou-se na Seção 5.3 que a metodologia de processamento de imagens proposta neste projeto (metodologia 2), em alguns casos, não localiza corretamente a região das sobrancelhas e que este componente facial pode ter exercido grande influência no reconhecimento de algumas expressões pelos voluntários. Logo, a falha no processamento de imagens e a consequente localização errada de seus correspondentes pontos de controle, pode comprometer os resultados. Em trabalhos futuros, portanto, pode-se pesquisar uma maneira alternativa de localizar a região deste componente facial. Verificou-se, ainda, que alguns sujeitos possuem sobrancelhas muito claras (loiras) dificultando assim o processamento de segmentação. Esse problema poderá ser abordado adicionando outras técnicas de processamento de imagens.

Pode-se considerar também a inclusão de mais pontos de controle na boca (por exemplo, cinco pontos nas curvas superior e inferior) para melhorar o desenho deste componente facial, uma vez que verificou-se nas imagens dos bancos de dados (Figura 4.7) que pessoas reais podem “torcer” a boca ou arqueá-la de diferentes formas.

A proposta mostrou que nos bancos de dados utilizados existem sujeitos com expressões faciais sutis, o que prejudicou a expressividade nas emoções. Portanto, uma possível forma de contornar esse problema é fazer um treinamento baseado em um subconjunto das imagens dos bancos de dados. Ou seja, será necessário criar um critério de escolha das amostras a fim de selecionar apenas pessoas mais expressivas.

Outra abordagem que pode ser testada no intuito de melhorar a expressividade das imagens de treinamento seria utilizar imagens de voluntários capturadas, de alguma forma, em situações naturais, isto é, sem solicitar que o voluntário force as emoções, e verificar se os resultados sofrem alguma variação significativa que possa ajudar na representação das emoções faciais.

Vale a pena mencionar que na metodologia 2 não foram aprendidas as variações dos pontos de controle que formam as rugas na face. Outro trabalho futuro, portanto, seria estender a metodologia desenvolvida neste projeto para o reconhecimento de rugas e outros sinais gerados pelas emoções – principalmente ao redor dos olhos, no nariz e ao redor da boca.

Os pontos *mfdir* e *mfesq* (Figura 2.8), os quais também são definidos no nariz na metodologia 1, não foram localizados já que as imagens binárias não apresentam informações suficientes para poder localizar esses pontos de controle. Considerando que esses pontos não estão presentes na imagem binária, os mesmos poderiam ser obtidos a partir do processamento das imagens originais e incorporados à metodologia 2. Como pode ser observado, as caricaturas geradas pelo presente projeto não mostram as narinas em nenhuma emoção, mas talvez elas pudessem auxiliar, por exemplo, na diferenciação entre as emoções de aversão e raiva. Na emoção aversão, as narinas ficariam mais aparentes do que na raiva. Portanto, a inclusão destes novos traços poderiam melhorar a representação das emoções faciais.

Referências

- AHO, A. V.; ULLMAN, J. D. *The theory of parsing, translation, and compiling*. [S.l.]: Prentice-Hall, Inc., 1972. Citado na página 26.
- AIFANTI, N.; PAPACHRISTOU, C.; DELOPOULOS, A. The mug facial expression database. In: IEEE. *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*. [S.l.], 2010. p. 1–4. Citado na página 55.
- AMER, M. et al. Monte carlo tree search for scheduling activity recognition. In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. [S.l.: s.n.], 2013. p. 1353–1360. ISSN 1550-5499. Citado na página 52.
- BEKELE, E. et al. Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. *Journal of Autism and Developmental Disorders*, Jan 2014. Citado na página 20.
- BOYLE, E. A.; ANDERSON, A. H.; NEWLANDS, A. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and speech*, SAGE Publications, v. 37, n. 1, p. 1–20, 1994. Citado na página 20.
- CALDER, A. J. et al. Caricaturing facial expressions. *Cognition*, Elsevier, v. 76, n. 2, p. 105–146, 2000. Citado na página 20.
- CANNY, J. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, n. 6, p. 679–698, 1986. Citado na página 35.
- CHEN, H. et al. Composite templates for cloth modeling and sketching. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. [S.l.: s.n.], 2006. v. 1, p. 943–950. ISSN 1063-6919. Citado 2 vezes nas páginas 21 e 45.
- CHENG, Y.; LING, S. 3d animated facial expression and autism in taiwan. In: IEEE. *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*. [S.l.], 2008. p. 17–19. Citado na página 20.
- DONG, J. et al. Towards unified human parsing and pose estimation. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. [S.l.: s.n.], 2014. p. 843–850. Citado 2 vezes nas páginas 44 e 50.
- DONG, J. et al. A deformable mixture parsing model with parselets. In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. [S.l.: s.n.], 2013. p. 3408–3415. ISSN 1550-5499. Citado na página 49.
- DYCK, M. et al. Recognition profile of emotions in natural and virtual faces. *PLOS One*, v. 3, n. 11, p. e3628, 2008. Citado 2 vezes nas páginas 20 e 23.
- EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, American Psychological Association, v. 17, n. 2, p. 124, 1971. Citado na página 20.

EKMAN, P.; FRIESEN, W. V. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, Springer, v. 1, n. 1, p. 56–75, 1976. Citado 2 vezes nas páginas 21 e 38.

EKMAN, P.; FRIESEN, W. V.; HAGER, J. C. Facs investigator's guide. *A human face*, 2002. Citado na página 96.

FARKAS, L. *Anthropometry of the head and face*. Raven Press, 1994. ISBN 9780781701594. Disponível em: <<https://books.google.com.br/books?id=bNJqAAAAMAAJ>>. Citado 4 vezes nas páginas 21, 36, 37 e 38.

FOLEY, J. D. *Computer graphics: Principles and practice, in C.* [S.l.]: Addison-Wesley Professional, 1996. v. 12110. Citado na página 21.

GONZALEZ, R.; WOODS, R. *Digital Image Processing*. [S.l.]: Pearson Education, 2011. ISBN 9780133002324. Citado 4 vezes nas páginas 30, 33, 34 e 35.

GOOCH, A. A. et al. Color2gray: salience-preserving color removal. In: ACM. *ACM Transactions on Graphics (TOG)*. [S.l.], 2005. v. 24, n. 3, p. 634–639. Citado na página 61.

GRYNSZPAN, O.; MARTIN, J.-C.; NADEL, J. Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human-Computer Studies*, Elsevier, v. 66, n. 8, p. 628–639, 2008. Citado na página 20.

HAN, F.; ZHU, S.-C. Bottom-up/top-down image parsing by attribute graph grammar. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. [S.l.: s.n.], 2005. v. 2, p. 1778–1785 Vol. 2. ISSN 1550-5499. Citado na página 46.

HARVEY, A. *OpenCV Face Detection: Visualized'*. 2011. Citado na página 59.

HOPCROFT, J.; ULLMAN, J. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979. (Addison-Wesley Series in Computer Science and Information Processing). ISBN 9780201029888. Disponível em: <https://books.google.com.br/books?id=i_BQAAAAMAAJ>. Citado na página 26.

JIANG, K. et al. Video event representation and inference on and-or graph. *Computer Animation and Virtual Worlds*, Wiley Online Library, v. 23, n. 3-4, p. 145–154, 2012. Citado 2 vezes nas páginas 44 e 51.

KIM, H.-C.; KIM, D.; BANG, S.-Y. A pca mixture model with an efficient model selection method. In: IEEE. *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*. [S.l.], 2001. v. 1, p. 430–435. Citado na página 41.

KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele University*, v. 33, n. 2004, p. 1–26, 2004. Citado na página 42.

KU, J. et al. Experimental results of affective valence and arousal to avatar's facial expressions. *Cyber Psychology & Behavior*, v. 8, n. 5, p. 493–503, 2005. Citado na página 20.

LAHIRI, U. et al. Design of a virtual reality based adaptive response technology for children with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, IEEE, v. 21, n. 1, p. 55–64, 2013. Citado na página 20.

LEE, H. et al. Automatic face and facial features detection. In: IEEE. *Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on*. [S.l.], 2001. v. 1, p. 254–259. Citado na página 41.

LEVENSON, R. W. et al. Emotion, physiology, and expression in old age. *Psychology and aging*, American Psychological Association, v. 6, n. 1, p. 28, 1991. Citado na página 20.

LIN, L. et al. Object categorization with sketch representation and generalized samples. *Pattern Recognition*, Elsevier, v. 45, n. 10, p. 3648–3660, 2012. Citado 2 vezes nas páginas 44 e 50.

LIN, L. et al. Discriminatively trained and-or graph models for object shape detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 37, n. 5, p. 959–972, May 2015. ISSN 0162-8828. Citado 2 vezes nas páginas 44 e 48.

LU, Y.; WU, T.; ZHU, S.-C. Online object tracking, learning, and parsing with and-or graphs. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. [S.l.: s.n.], 2014. p. 3462–3469. Citado na página 52.

MENEZES, P. *Linguagens formais e autômatos*. Bookman, 2005. (Série livros didáticos). ISBN 9788577802661. Disponível em: <<https://books.google.com.br/books?id=a5zCPgAACAAJ>>. Citado 2 vezes nas páginas 26 e 27.

MIKELS, J. et al. Emotional category data on images from the international affective picture system. *Behavior Research Methods*, Springer-Verlag, v. 37, n. 4, p. 626–630, 2005. ISSN 1554-351X. Disponível em: <<http://dx.doi.org/10.3758/BF03192732>>. Citado na página 55.

NEWSCHAFER, C. J.; CURRAN, L. K. Autism: an emerging public health problem. *Public health reports*, Association of Schools of Public Health, v. 118, n. 5, p. 393, 2003. Citado na página 20.

NUNES, F. L. S. *Introdução ao Processamento de Imagens Médicas para Auxílio ao Diagnóstico*. 1ed. ed. Rio de Janeiro: PUC-Rio: In: Karin Breitman, 2006. Citado 2 vezes nas páginas 30 e 32.

PEI, M.; DONG, Z.; ZHAO, M. Event recognition based-on social roles in continuous video. In: *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. [S.l.: s.n.], 2013. p. 1–6. ISSN 1945-7871. Citado na página 51.

PEI, M. et al. Learning and parsing video events with goal and intent prediction. *Computer Vision and Image Understanding*, Elsevier, v. 117, n. 10, p. 1369–1383, 2013. Citado 2 vezes nas páginas 48 e 51.

PHIMOLTARES, S.; LURSINSAP, C.; CHAMNONGTHAI, K. Locating essential facial features using neural visual model. In: IEEE. *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on*. [S.l.], 2002. v. 4, p. 1914–1919. Citado na página 41.

ROTHROCK, B.; PARK, S.; ZHU, S.-C. Integrating grammar and segmentation for human pose estimation. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. [S.l.: s.n.], 2013. p. 3214–3221. ISSN 1063-6919. Citado na página 49.

RUSS, C. J. *The Image Processing Handbook*. [S.l.]: Data Translation, Incorporated, 1995. Citado na página 30.

SALOMAA, A. *Formal languages*. Academic Press, 1973. (ACM monograph series). Disponível em: <<http://books.google.com.br/books?id=aOhQAAAAMAAJ>>. Citado na página 26.

SERRA, J. *Image analysis and mathematical morphology, v. 1*. [S.l.]: Academic press, 1982. Citado na página 35.

SI, Z. et al. Unsupervised learning of event and-or grammar and semantics from video. In: IEEE. *Computer Vision (ICCV), 2011 IEEE International Conference on*. [S.l.], 2011. p. 41–48. Citado 2 vezes nas páginas 48 e 51.

SI, Z.; ZHU, S.-C. Learning and-or templates for object recognition and detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 35, n. 9, p. 2189–2205, Sept 2013. ISSN 0162-8828. Citado na página 47.

SOHAIL, A. S. M.; BHATTACHARYA, P. Detection of facial feature points using anthropometric face model. In: *Signal Processing for Image Enhancement and Multimedia Processing*. [S.l.]: Springer, 2008. p. 189–200. Citado na página 41.

SONG, X. et al. Discriminatively trained and-or tree models for object detection. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. [S.l.: s.n.], 2013. p. 3278–3285. ISSN 1063-6919. Citado na página 48.

SUO, J. et al. A compositional and dynamic model for face aging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 32, n. 3, p. 385–401, March 2010. ISSN 0162-8828. Citado 3 vezes nas páginas 21, 45 e 46.

TESTA, R. et al. Generating facial emotions for diagnosis and training. In: *Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on*. [S.l.: s.n.], 2015. p. 304–309. Citado 7 vezes nas páginas 21, 22, 23, 39, 64, 84 e 88.

TU, Z. et al. Image parsing: Unifying segmentation, detection, and recognition. *International Journal of computer vision*, Springer, v. 63, n. 2, p. 113–140, 2005. Citado 2 vezes nas páginas 29 e 30.

University of Starling. <http://pics.stir.ac.uk/ESRC/>. 2013. Citado na página 55.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: IEEE. *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. [S.l.], 2001. v. 1, p. I–511. Citado na página 59.

WANG, D.; PEI, M.; ZHU, L. Detecting driver use of mobile phone based on in-car camera. In: *Computational Intelligence and Security (CIS), 2014 Tenth International Conference on*. [S.l.: s.n.], 2014. p. 148–151. Citado na página 53.

WANG, X. et al. Incorporating structural alternatives and sharing into hierarchy for multiclass object recognition and detection. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. [S.l.: s.n.], 2013. p. 3334–3341. ISSN 1063-6919. Citado na página 51.

WU, T.; ZHU, S.-C. A numerical study of the bottom-up and top-down inference processes in and-or graphs. *International journal of computer vision*, Springer, v. 93, n. 2, p. 226–252, 2011. Citado 2 vezes nas páginas 47 e 53.

WU, T.-F.; XIA, G.-S.; ZHU, S.-C. Compositional boosting for computing hierarchical image structures. In: *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*. [S.l.: s.n.], 2007. p. 1–8. ISSN 1063-6919. Citado 3 vezes nas páginas 44, 47 e 53.

XU, Z. et al. A hierarchical compositional model for face representation and sketching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 30, n. 6, p. 955–969, June 2008. ISSN 0162-8828. Citado 3 vezes nas páginas 21, 45 e 46.

ZHU, L. L. et al. Max margin learning of hierarchical configurational deformable templates (hcdts) for efficient object parsing and pose estimation. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 93, n. 1, p. 1–21, maio 2011. ISSN 0920-5691. Disponível em: <<http://dx.doi.org/10.1007/s11263-010-0375-1>>. Citado na página 49.

ZHU, L. L. et al. Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion. In: *Computer vision–eccv 2008*. [S.l.]: Springer, 2008. p. 759–773. Citado na página 49.

ZHU, S.-C.; MUMFORD, D. *A stochastic grammar of images*. [S.l.]: Now Publishers Inc, 2007. Citado 2 vezes nas páginas 28 e 29.

APÊNDICE A – Descrições da aparência de cada unidade de ação

A Tabela 8 contém uma descrição textual do movimento e da aparência gerada por cada unidade de ação baseados em (EKMAN; FRIESEN; HAGER, 2002).

Tabela 8 – Descrições da aparência de cada UA.

Unidade de ação	Nome	Descrição
1	Inner Brow Raiser	Eleva os cantos internos da sobrancelha
		Pode causar enrugamento horizontal na testa
2	Outer Brow Raiser	Eleva a porção exterior das sobrancelhas
		Produz uma forma arqueada para as sobrancelhas
		Permite gerar rugas laterais entre as sobrancelhas
4	Brow Lowerer	Diminui a sobrancelha
		Empurra as pálpebras estreitando a altura dos olhos
		Aproxima as sobrancelhas
		Produz rugas verticais entre as sobrancelhas
		Pode produzir um enrugamento do nariz
5	Upper Lid Raiser	Amplia a abertura dos olhos
		Eleva a pálpebra superior sem que ela desapareça por completo
		Permite expor esclera por cima da íris
		Apresenta um olhar fixo
		Amplia pálpebra inferior
6	Cheek Raiser and Lid Compressor	Puxa as bochechas em direção aos olhos
		Empurra a pele ao redor dos olhos em direção à cavidade ocular
		Aprofunda o sulco da pálpebra inferior
		Pode diminuir porção lateral das sobrancelhas
		Provoca rugas nos cantos externos da abertura dos olhos (Pés de galinha)
		Aperta as pálpebras

Tabela 8 – Descrições da aparência de cada UA.

Unidade de ação	Nome	Descrição
		Diminui a abertura dos olhos
		Eleva a pálpebra inferior
		Pode causar rugas nos olhos
9	Nose Wrinkler	Puxa a pele do nariz, causando rugas laterais
		Reduz a parte medial das sobrancelhas
		Limita a abertura dos olhos
		Puxa o centro do lábio superior para cima
		Pode ampliar e levantar as narinas
12	Lip Corner Puller	Puxa os cantos dos lábios para trás e para cima, de forma oblíqua
		Aprofunda o sulco nasogeniano, puxando-o lateralmente e para cima
		Pode limitar a abertura dos olhos
		Pode aumentar e alargar as narinas
		Pode achatam e esticar a pele do queixo
15	Lip Corner Depressor	Puxa os cantos dos lábios para baixo
		O lábio inferior é ligeiramente esticado horizontalmente
		Pode produzir enrugamentos nos cantos dos lábios
		Pode achatam/ produzir uma depressão no queixo
16+25	Lower Lip Depressor	Puxa o lábio inferior para baixo
		Esticam os lábios lateralmente
		Normalmente expõe mais os dentes
		Estica o queixo
		Pode causar rugas no queixo ou no lábio inferior
20	Lip Stretcher	Estica os lábio lateralmente
		Alonga a boca
		Achata os lábios
		Pode aparecer rugas nos cantos dos lábios
		Puxa a porção inferior do sulco naso-labial, e puxa a pele na área inferior do sulco nasolabiais lateralmente

Tabela 8 – Descrições da aparência de cada UA.

Unidade de ação	Nome	Descrição
20	Lip Stretcher	Estica lateralmente a pele do queixo
		Pode alongar a abertura das narinas
23	Lip Tightener	Estreita os lábios
		Pode haver uma leve separação dos lábios
		Pode produzir enrugamento dos lábios
25	Lips Part	Expõe uma parte da mucosa da boca
		Pode expor os dentes
26	Jaw Drop	A mandíbula é abaixada
		Pode separar os lábios a fim de expor os dentes

APÊNDICE B – Descrição de cada ponto de controle

Uma descrição mais detalhada para ponto de controle definido na abordagem anterior é encontrada na Tabela 9.

Tabela 9 – Inferências feitas a partir da antropometria.

Parte do Rosto	Pontos	Considerado
Sobrancelha	$p1$ e $p2$	A distância entre as extremidades das duas sobrancelhas é média da diferença de ex_ex e ft_ft
		A distância entre as extremidades internas das sobrancelhas é igual a en
		O nível mais baixo da sobrancelha tem a mesma altura que n
	os	Fica na metade da distância de $p1$ e $p2$
Olho	ps e pi	Ficam na metade da distância de en_ex
	ex e en	Ficam na mesma altura
Pálpebra inferior	$pp1$	Fica logo abaixo de en
	$pp2$	Fica a 90% de en_ex em relação a en
	$pp1$ e $pp2$	Ficam a 15% de pi_or em relação a pi
Nariz	n	Fica na metade de en_en
	mf	Fica 5% de n_sn abaixo do olho (en)
	al	Fica 90% de n_sn abaixo de n
Lábios	ls e sto	Ficam na mesma linha vertical que n
Enrugamento do olho	$po1_po3$, $po4_po6$ e $po7_po9$	As rugas possuem em média 2 mm
	$po1_po3$, $po4_po6$ e $po7_po9$	As rugas esticam até 2+1 mm
	$po4$	$po4$ fica na altura de ex
	$po1$	$po1$ fica 1mm acima de $po4$

Tabela 9 – Inferências feitas a partir da antropometria.

Parte do Rosto	Pontos	Considerado
	<i>po9</i>	<i>p9</i> fica 1mm abaixo de <i>po4</i>
Enrugamento da testa	<i>pt3</i>	Fica no meio de <i>ft_ft</i> e <i>tr_g</i> (no meio da testa)
	<i>pt1</i>	Fica na mesma posição que <i>ft</i> (no final da testa)
	<i>pt2</i>	Fica à um quarto a esquerda de <i>ft_ft</i> e <i>tr_g</i> (um quarto da testa)
Enrugamento da testa	<i>pt4</i>	Fica a um quarto a direita de <i>ft_ft</i> e <i>tr_g</i> (um quarto da testa)
	<i>pt5</i>	Fica na mesma posição que <i>ft(d)</i> (no final da testa)
	<i>pt3(c)</i>	Fica a 25% de <i>tr_g</i> em relação a <i>tr</i>
	<i>pt3(b)</i>	Fica a 75% de <i>tr_g</i> em relação a <i>tr</i>
Contorno	<i>zy</i>	Possui o mesmo posicionamento vertical que <i>or</i>
	<i>eu</i>	Possui o mesmo posicionamento vertical que <i>tr</i>
	<i>gn, tr</i> e <i>v</i>	Possui o mesmo posicionamento horizontal que <i>n</i>
	<i>t</i>	Possui o posicionamento vertical na metade de <i>pi_or</i>
Auxiliares	<i>n</i>	Fica na metade da distância de <i>en_en</i> e <i>mf_mf</i>
	<i>se</i>	Fica na mesma linha que <i>n</i>

Além deste ponto, os pontos *p2* da sobrancelha, *pas1* e *pas2* da pálpebra inferior, bem como os pontos do enrugamento não apresentam nenhuma descrição pela antropometria e, por isso, foram inferidos com base em outras medidas antropométricas para que fosse possível a representação de suas curvas.

APÊNDICE C – Posicionamento dos pontos de controle na abordagem inicial

Na Tabela 10 é apresentado o cálculo das posições dos pontos de controle nas unidades de ação seguindo a descrição do envoltório.

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
1	Sobrancelha esquerda	$p1eX = p1eX + 0.07 \cdot i \cdot (p2eX - p1eX)$
		$p1eY = p1eY + 0.09 \cdot i \cdot (p1eY - oseY)$
		$pas1eY = pas1eY + 0.035 \cdot i \cdot (p1eY - oseY)$
		$oseY = oseY - 0.02 \cdot i \cdot (p1eY - oseY)$
	Sobrancelha direita	$p1dX = p1eX - 0.07 \cdot i \cdot (p2dX - p1dX)$
		$p1dY = p1dY + 0.09 \cdot i \cdot (p1dY - osdY)$
		$pas1dY = pas1dY + 0.035 \cdot i \cdot (p1dY - osdY)$
		$osdY = osdY - 0.02 \cdot i \cdot (p1dY - osdY)$
	Enrugamentos Testa central	$pt2Y = pt2Y + (i-1) \cdot (pt3Y - pt2Y)$
		$pt3Y = pt3Y + (i-1) \cdot (pt3Y - pt2Y)$
		$pt4Y = pt4Y + (i-1) \cdot (pt3Y - pt2Y)$
2	Sobrancelha esquerda	$p1eX = p1eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$p2eX = p2eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$pas1eX = pas1eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
	Sobrancelha esquerda	$pas2eX = pas2eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$oseX = oseX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$p1eY = p1eY + 0.01 \cdot i \cdot (p1eY - oseY)$
		$p2eY = p2eY + 0.015 \cdot i \cdot (p1eY - oseY)$
		$pas1eY = pas1eY + 0.015 \cdot i \cdot (p1eY - oseY)$
		$pas2eY = pas2eY + 0.06 \cdot i \cdot (p1eY - oseY)$
		$oseY = oseY - 0.025 \cdot i \cdot (p1eY - oseY)$
	Sobrancelha direita	$p1dX = p1dX + 0.01 \cdot i \cdot (p2dX - p1dX)$
		$p2dX = p2dX + 0.01 \cdot i \cdot (p2dX - p1dX)$
		$pas1dX = pas1dX + 0.01 \cdot i \cdot (p2dX - p1dX)$
		$pas2dX = pas2dX + 0.01 \cdot i \cdot (p2dX - p1dX)$
		$osdX = osdX + 0.01 \cdot i \cdot (p2dX - p1dX)$

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
		$p1dY = p1dY + 0.01 \cdot i \cdot (p1dY - osdY)$
		$p2dY = p2dY + 0.015 \cdot i \cdot (p1dY - osdY)$
		$pas1dY = pas1dY + 0.015 \cdot i \cdot (p1dY - osdY)$
		$pas2dY = pas2dY + 0.06 \cdot i \cdot (p1dY - osdY)$
		$osdY = osdY - 0.025 \cdot i \cdot (p1dY - osdY)$
	Olho: pálpebras superiores	$pseY = pseY + 0.2 \cdot i \cdot (pieY - pseY)$
		$pseY = pseY + 0.2 \cdot i \cdot (pidY - psdY)$
	Enrugamentos Testa central	$pt1cY = pt1cY + (i-1) \cdot (pt3cY - pt2cY)$
		$pt2cY = pt2cY + (i-1) \cdot (pt3cY - pt2cY)$
		$pt3cY = pt3cY + (i-1) \cdot (pt3cY - pt2cY)$
		$pt4cY = pt4cY + (i-1) \cdot (pt3cY - pt2cY)$
		$pt5cY = pt5cY + (i-1) \cdot (pt3cY - pt2cY)$
		$pt1mY = pt1mY + (i-1) \cdot (pt3mY - pt2mY)$
		$pt2mY = pt2mY + (i-1) \cdot (pt3mY - pt2mY)$
		$pt3mY = pt3mY + (i-1) \cdot (pt3mY - pt2mY)$
2	Enrugamentos Testa central	$pt4mY = pt4mY + (i-1) \cdot (pt3mY - pt2mY)$
		$pt5mY = pt5mY + (i-1) \cdot (pt3mY - pt2mY)$
		$pt1bY = pt1bY + (i-1) \cdot (pt3bY - pt2bY)$
		$pt2bY = pt2bY + (i-1) \cdot (pt3bY - pt2bY)$
		$pt3bY = pt3bY + (i-1) \cdot (pt3bY - pt2bY)$
		$pt4bY = pt4bY + (i-1) \cdot (pt3bY - pt2bY)$
		$pt5bY = pt5bY + (i-1) \cdot (pt3bY - pt2bY)$
4	Sobrancelha esquerda	$p1eX = p1eX - 0.05 \cdot i \cdot (p2eX - p1eX)$
		$p2eX = p2eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$pas1eX = pas1eX - 0.04 \cdot i \cdot (p2eX - p1eX)$
		$pas2eX = pas2eX - 0.01 \cdot i \cdot (p2eX - p1eX)$
		$oseX = oseX - 0.02 \cdot i \cdot (p2eX - p1eX)$
		$p1eY = p1eY - 0.03 \cdot i \cdot (p1eY - oseY)$
		$p2eY = p2eY - 0.01 \cdot i \cdot (p1eY - oseY)$
		$pas1eY = pas1eY - 0.025 \cdot i \cdot (p1eY - oseY)$
		$pas2eY = pas2eY - 0.01 \cdot i \cdot (p1eY - oseY)$
		$oseY = oseY - 0.04 \cdot i \cdot (p1eY - oseY)$
		$p1dX = p1dX + 0.05 \cdot i \cdot (p2dX - p1dX)$
		$p2dX = p2dX + 0.01 \cdot i \cdot (p2dX - p1dX)$

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
		$\text{pas1dX} = \text{pas1dX} + 0.04 \cdot i \cdot (\text{p2dX} - \text{p1dX})$
		$\text{pas2dX} = \text{pas2dX} + 0.01 \cdot i \cdot (\text{p2dX} - \text{p1dX})$
		$\text{osdX} = \text{osdX} + 0.02 \cdot i \cdot (\text{p2dX} - \text{p1dX})$
		$\text{p1dY} = \text{p1dY} - 0.03 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{p2dY} = \text{p2dY} - 0.01 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{pas1dY} = \text{pas1dY} - 0.025 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{pas2dY} = \text{pas2dY} - 0.01 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{osdY} = \text{osdY} - 0.04 \cdot i \cdot (\text{p1dY} - \text{osdY})$
	Olho: pálpebras superiores	$\text{pseY} = \text{pseY} - 0.2 \cdot i \cdot (\text{pieY} - \text{pseY})$
		$\text{psdY} = \text{psdY} - 0.2 \cdot i \cdot (\text{pidY} - \text{psdY})$
1+2	Sobrancelha	$\text{pleY} = \text{pleY} - 0.02 \cdot i \cdot (\text{pleY} - \text{oseY})$
1+2	Sobrancelha esquerda	$\text{pleY} = \text{pleY} - 0.02 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{p2eY} = \text{p2eY} + 0.01 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{pas1eY} = \text{pas1eY} + 0.015 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{pas2eY} = \text{pas2eY} + 0.02 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{oseY} = \text{oseY} + 0.04 \cdot i \cdot (\text{pleY} - \text{oseY})$
	Sobrancelha direita	$\text{p1dY} = \text{p1dY} - 0.02 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{p2dY} = \text{p2dY} + 0.01 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{pas1dY} = \text{pas1dY} + 0.015 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{pas2dY} = \text{pas2dY} + 0.02 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{osdY} = \text{osdY} + 0.04 \cdot i \cdot (\text{p1dY} - \text{osdY})$
1+4	Sobrancelha esquerda	$\text{pleX} = \text{pleX} + 0.02 \cdot i \cdot (\text{p2eX} - \text{pleX})$
		$\text{pas1eX} = \text{pas1eX} + 0.1 \cdot i \cdot (\text{p2eX} - \text{pleX})$
		$\text{pas1eY} = \text{pas1eY} + 1.5 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{pas2eY} = \text{pas2eY} - 0.5 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{oseY} = \text{oseY} + i \cdot (\text{pleY} - \text{oseY})$
	Sobrancelha direita	$\text{p1dX} = \text{p1dX} - 0.02 \cdot i \cdot (\text{p2dX} - \text{p1dX})$
		$\text{pas1dX} = \text{pas1dX} - 0.1 \cdot i \cdot (\text{p2dX} - \text{p1dX})$
		$\text{pas1dY} = \text{pas1dY} + 1.5 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{pas2dY} = \text{pas2dY} - 0.5 \cdot i \cdot (\text{p1dY} - \text{osdY})$
		$\text{osdY} = \text{osdY} + i \cdot (\text{p1dY} - \text{osdY})$
	Sobrancelha esquerda	$\text{pleY} = \text{pleY} + 0.01 \cdot i \cdot 3.5\text{cm}(\text{pleY} - \text{oseY})$
		$\text{p2eY} = \text{p2eY} + 0.01 \cdot i \cdot (\text{pleY} - \text{oseY})$
		$\text{pas1eY} = \text{pas1eY} - 0.025 \cdot i \cdot (\text{pleY} - \text{oseY})$

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
		$pas2eY = pas2eY + 0.01 \cdot i \cdot (pleY - oseY)$
		$oseY = oseY + 0.01 \cdot i \cdot (pleY - oseY)$
	Sobrancelha direita	$p1dY = p1dY + 0.01 \cdot i \cdot (p1dY - osdY)$
		$p2dY = p2dY + 0.01 \cdot i \cdot (p1dY - osdY)$
		$pas1dY = pas1dY - 0.025 \cdot i \cdot (p1dY - osdY)$
		$pas2dY = pas2dY + 0.01 \cdot i \cdot (p1dY - osdY)$
		$osdY = osdY + 0.01 \cdot i \cdot (p1dY - osdY)$
1+2+4	Olho: pálpebras superiores	$pseY = pseY - 0.2 \cdot i \cdot (pieY - pseY)$
		$psdY = psdY - 0.2 \cdot i \cdot (pidY - psdY)$
5	Olho: pálpebras superiores	$pseY = pseY + 0.6 \cdot i \cdot (pieY - pseY)$
		$psdY = psdY + 0.6 \cdot i \cdot (pidY - psdY)$
6	Olhos e pálpebras esquerda	$pseY = pseY - 0.3 \cdot i \cdot (pieY - pseY)$
		$pieY = pieY + 0.3 \cdot i \cdot (pieY - pseY)$
		$oreY = oreY + 0.3 \cdot i \cdot (oreY - pp1eY)$
	Olhos e pálpebras direita	$psdY = psdY - 0.3 \cdot i \cdot (pidY - psdY)$
		$pidY = pidY + 0.3 \cdot i \cdot (pidY - psdY)$
		$ordY = ordY + 0.3 \cdot i \cdot (ordY - pp1dY)$
	Enrugamentos olho esquerdo	$p1eY = p1eY + (i-1) \cdot (p8eY - p6eY)$
		$p1eX = p1eX + (i-1) \cdot (p8eX - p6eX)$
		$p2eX = p2eX + (i-1) \cdot (p8eX - p6eX)$
		$p3eX = p3eX + (i-1) \cdot (p8eX - p6eX)$
		$p4eX = p4eX + (i-1) \cdot (p8eX - p6eX)$
		$p5eX = p5eX + (i-1) \cdot (p8eX - p6eX)$
		$p6eX = p6eX + (i-1) \cdot (p8eX - p6eX)$
		$p7eX = p7eX + (i-1) \cdot (p8eX - p6eX)$
		$p8eX = p8eX + (i-1) \cdot (p8eX - p6eX)$
		$p9eX = p9eX + (i-1) \cdot (p8eX - p6eX)$
		$p10eX = p10eX + (i-1) \cdot (p8eX - p6eX)$
		$p11eX = p11eX + (i-1) \cdot (p8eX - p6eX)$
	Enrugamentos olho direito	$p1dY = p1dY + (i-1) \cdot (p8dY - p6dY)$
		$p1dX = p1dX + (i-1) \cdot (p8dX - p6dX)$
		$p2dX = p2dX + (i-1) \cdot (p8dX - p6dX)$
		$p3dX = p3dX + (i-1) \cdot (p8dX - p6dX)$
		$p4dX = p4dX + (i-1) \cdot (p8dX - p6dX)$

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
		$p5dX = p5dX + (i-1) \cdot (p8dX - p6dX)$
		$p6dX = p6dX + (i-1) \cdot (p8dX - p6dX)$
		$p7dX = p7dX + (i-1) \cdot (p8dX - p6dX)$
		$p8dX = p8dX + (i-1) \cdot (p8dX - p6dX)$
		$p9dX = p9dX + (i-1) \cdot (p8dX - p6dX)$
		$p10dX = p10dX + (i-1) \cdot (p8dX - p6dX)$
		$p11dX = p11dX + (i-1) \cdot (p8dX - p6dX)$
7	Olhos e pálpebras	$pseY = pseY - 0.5 \cdot i \cdot (pieY - pseY)$
7	Olhos e pálpebras esquerda	$pieY = pieY + 0.5 \cdot i \cdot (pieY - pseY)$
		$oreY = oreY + 0.5 \cdot i \cdot (oreY - pp1eY)$
	Olhos e pálpebras direita	$psdY = psdY - 0.5 \cdot i \cdot (pidY - psdY)$
		$pidY = pidY + 0.5 \cdot i \cdot (pidY - psdY)$
		$ordY = ordY + 0.5 \cdot i \cdot (ordY - pp1dY)$
9	Nariz esquerda	$aceX = aceX + 0.03 \cdot i \cdot (mfeX - snX)$
		$aceY = aceY + 0.06 \cdot i \cdot (snY - mfeY)$
		$aleY = aleY + 0.06 \cdot i \cdot (snY - mfeY)$
	Nariz direita	$aceX = aceX - 0.03 \cdot i \cdot (mfeX - snX)$
		$aceY = aceY + 0.06 \cdot i \cdot (snY - mfeY)$
		$aleY = aleY + 0.06 \cdot i \cdot (snY - mfeY)$
	Lábio superior	$lsY = lsY + 0.06 \cdot i \cdot (liY - lsY)$
		$stoY = stoY + 0.06 \cdot i \cdot (liY - lsY)$
12	Lábios	$cheX = cheX + 0.1 \cdot i \cdot (cheX - chX)$
		$chdX = chdX - 0.1 \cdot i \cdot (cheX - chX)$
		$cheY = cheY + 1.0 \cdot i \cdot (liY - lsY)$
		$chdY = chdY + 1.0 \cdot i \cdot (liY - lsY)$
15	Lábios	$cheY = cheY - 1.0 \cdot i \cdot (liY - lsY)$
		$chdY = chdY - 1.0 \cdot i \cdot (liY - lsY)$
16+25	Lábio inferior	$stoY = stoY - 1.2 \cdot i \cdot (liY - lsY)$
		$liY = liY - 1.2 \cdot i \cdot (liY - lsY)$
20	Lábios	$cheX = cheX + 0.15 \cdot i \cdot (cheX - chX)$
		$chdX = chdX - 0.15 \cdot i \cdot (cheX - chX)$
		$lsY = lsY - 0.1 \cdot i \cdot (liY - lsY)$
		$liY = liY + 0.1 \cdot i \cdot (liY - lsY)$
		$cheX = cheX + 0.004 \cdot i \cdot (cheX - chX)$

Tabela 10 – Cálculo das posições de cada ponto na abordagem inicial.

Unidade de ação	Parte do rosto	Equação
		$\text{chdX} = \text{chdX} - 0.004 \cdot i \cdot (\text{cheX} - \text{chX})$
		$\text{lsY} = \text{lsY} - 0.27 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{liY} = \text{liY} + 0.27 \cdot i \cdot (\text{liY} - \text{lsY})$
25	Lábios	$\text{cheX} = \text{cheX} + 0.01 \cdot i \cdot (\text{cheX} - \text{chX})$
		$\text{chdX} = \text{chdX} - 0.01 \cdot i \cdot (\text{cheX} - \text{chX})$
	Lábio superior	$\text{lsY} = \text{lsY} + 0.3 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{stoY} = \text{stoY} + 0.3 \cdot i \cdot (\text{liY} - \text{lsY})$
	Lábio inferior	$\text{stoY} = \text{stoY} - 0.3 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{liY} = \text{liY} - 0.3 \cdot i \cdot (\text{liY} - \text{lsY})$
26	Lábios	$\text{cheX} = \text{cheX} + 0.05 \cdot i \cdot (\text{cheX} - \text{chX})$
		$\text{chdX} = \text{chdX} - 0.05 \cdot i \cdot (\text{cheX} - \text{chX})$
	Lábio superior	$\text{lsY} = \text{lsY} + 1.0 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{stoY} = \text{stoY} + 1.0 \cdot i \cdot (\text{liY} - \text{lsY})$
	Lábio inferior	$\text{stoY} = \text{stoY} - 1.0 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{liY} = \text{liY} - 1.0 \cdot i \cdot (\text{liY} - \text{lsY})$
27	Lábios	$\text{cheX} = \text{cheX} + 0.1 \cdot i \cdot (\text{cheX} - \text{chX})$
		$\text{chdX} = \text{chdX} - 0.1 \cdot i \cdot (\text{cheX} - \text{chX})$
	Lábio superior	$\text{lsY} = \text{lsY} + 2.0 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{stoY} = \text{stoY} + 2.0 \cdot i \cdot (\text{liY} - \text{lsY})$
	Lábio inferior	$\text{stoY} = \text{stoY} - 2.0 \cdot i \cdot (\text{liY} - \text{lsY})$
		$\text{liY} = \text{liY} - 2.0 \cdot i \cdot (\text{liY} - \text{lsY})$

APÊNDICE D – Estrutura do Grafo *And-Or*

Figura D.1 – Estrutura do grafo *And-Or*.

APÊNDICE E – Valores das variações dos pontos de controle obtidos neste projeto

Ponto de Controle	Satisfação	Tristeza	Surpresa	Medo	Raiva	Aversão
Sobrancelha Direita						
p2dX	0.000	-0.006	-0.011	-0.016	0.015	0.015
p2dY	0.023	-0.096	0.686	0.159	-0.032	-0.015
osdX	-0.005	-0.015	-0.003	-0.019	0.018	0.014
osdY	0.027	-0.185	0.741	0.177	-0.468	-0.034
p1dX	0.000	-0.019	-0.005	-0.004	0.058	0.015
p1dY	0.041	-0.139	0.778	0.133	-0.395	-0.053
pas1dX	0.001	0.002	-0.003	0.002	0.003	0.002
pas1dY	0.003	0.003	0.040	0.090	0.002	0.001
pas2dX	0.000	0.001	0.001	0.001	0.001	0.003
pas2dY	0.002	0.001	0.037	0.080	0.002	0.003
Sobrancelha Esquerda						
p2eX	0.001	0.000	0.009	0.007	-0.014	-0.018
p2eY	0.025	-0.086	0.624	0.167	-0.335	-0.014
oseX	-0.001	0.027	0.003	0.012	-0.020	-0.023
oseY	0.043	-0.194	0.774	0.151	-0.428	-0.048
p1eX	-0.001	0.018	0.003	0.000	-0.057	-0.012
p1eY	0.057	-0.147	0.735	0.185	-0.320	-0.077
pas1eX	0.000	0.001	0.002	0.000	0.001	0.000
pas1eY	0.002	0.001	0.039	0.028	0.002	0.001
pas2eX	0.001	0.000	0.001	0.001	0.000	0.001
pas2eY	0.003	0.002	0.040	0.030	0.001	0.002
Olho Direito						
exdX	-0.008	-0.008	0.001	0.008	-0.019	-0.010
exdY	-0.043	0.217	-0.004	-0.056	0.039	0.258
endX	0.009	-0.005	-0.001	-0.003	0.016	0.004
endY	-0.038	0.206	-0.005	-0.078	0.043	0.265
psdX	0.001	-0.000	-0.003	0.008	0.006	0.009
psdY	-0.412	-0.170	0.918	0.547	-0.747	-0.099
pidX	0.001	-0.001	-0.001	0.004	0.005	0.008

Continua na seguinte pagina.

Ponto de Controle	Satisfação	Tristeza	Surpresa	Medo	Raiva	Aversão
pidY	0.390	0.156	-0.901	-0.486	0.547	0.245
pupildX	0.000	0.001	0.000	0.001	0.001	0.000
pupildY	-0.001	0.000	0.007	0.006	0.000	0.001
Olho Esquerdo						
exeX	0.009	0.008	-0.003	-0.007	0.015	0.004
exeY	-0.032	0.258	-0.009	-0.098	0.036	0.205
eneX	-0.002	0.001	0.002	0.007	-0.012	-0.000
eneY	-0.036	0.217	-0.004	-0.059	0.044	0.289
pseX	-0.005	0.000	-0.001	0.003	0.007	0.005
pseY	-0.387	-0.162	0.919	0.537	-0.793	-0.089
pieX	-0.000	0.005	-0.007	0.005	0.005	0.001
pieY	0.423	0.146	-0.886	-0.486	0.583	0.194
pupileX	0.000	0.000	0.000	0.001	0.002	0.002
pupileY	-0.000	-0.001	0.008	0.009	0.002	0.001
Nariz						
aldX	-0.058	-0.016	0.001	-0.003	-0.021	-0.017
aldY	0.044	-0.116	0.002	0.000	0.151	0.281
aleX	0.054	0.008	0.001	0.003	0.034	0.018
aleY	0.034	-0.120	0.000	0.005	0.128	0.272
acdX	-0.044	-0.003	0.001	-0.001	-0.017	-0.018
acdY	0.035	-0.129	0.002	0.002	0.154	0.253
aceX	0.040	0.004	0.000	0.001	0.016	0.019
aceY	0.041	-0.139	0.001	0.002	0.173	0.244
snX	0.002	-0.012	0.000	-0.001	-0.000	-0.006
snY	0.003	0.109	0.000	0.000	0.105	0.067
Boca						
chdX	-0.150	-0.021	-0.006	-0.125	-0.006	-0.013
chdY	0.959	-0.825	-0.187	0.219	0.314	-0.381
cheX	0.133	0.002	0.007	0.117	0.007	0.010
cheY	0.924	-0.851	-0.166	0.257	0.318	-0.345
lsX	-0.008	-0.007	-0.001	-0.001	-0.005	-0.003
lsY	-0.091	0.154	0.918	-0.567	-0.476	0.410
liX	-0.009	-0.008	-0.003	-0.009	-0.007	-0.001
liY	-0.086	0.062	-0.997	0.570	0.290	0.528
stoX	-0.006	-0.010	-0.001	-0.006	-0.004	-0.011
stoY	-0.021	0.022	0.003	-0.504	-0.234	0.251

Continua na seguinte pagina.

Ponto de Controle	Satisfação	Tristeza	Surpresa	Medo	Raiva	Aversão
-------------------	------------	----------	----------	------	-------	---------

Tabela 11 – Valores das variações dos pontos de controle
obtidos neste projeto.

APÊNDICE F – Utilização de grafos *And-Or* para reconhecimento de padrões em imagens e geração de imagens: Protocolo de Revisão

Objetivo

Identificar os algoritmos de grafos *And-Or* utilizados para reconhecimento de padrões e geração de imagens, a serem implementados e aplicados.

Questão de Pesquisa

Quais são os algoritmos de grafos *And-Or* utilizados para reconhecimento de padrões e geração de imagens?

Palavras-chaves e Strings de busca

- “*And-Or* Graph” AND “Algorithm” AND (“Inference” OR “Learning” OR “Estimation” OR “Syntactic Analysis” OR “Parsing”)
- “*And-Or* Graph” AND “Image” AND (“Generation” OR “Design”)

Intervenção

Identificação e compreensão dos algoritmos de grafos *And-Or*.

Controle

- Livro: Zhu, Song-Chun, and David Mumford. A stochastic grammar of images. *Now Publishers Inc*, 2007.
- Artigo: Pedro, Ricardo Wandré Dias, Fátima LS Nunes, and Ariane Machado-Lima. "Using grammars for pattern recognition in images: A systematic review." *ACM Computing Surveys* (CSUR) 46.2 (2013): 26.

População

Livros e artigos científicos relacionados a reconhecimento de padrões em imagens e geração de imagens, que descrevam algoritmos que manipulem grafos *And-Or*.

Resultados:

Conhecimento aprofundado e abrangente dos algoritmos que envolvem reconhecimento de imagens e geração de imagens utilizando grafos *And-Or*, assim como, identificação das lacunas da área para projetos futuros.

Aplicação

Para os pesquisadores nas áreas de processamento de imagens e linguagens formais, que tenham interesse em aplicar gramáticas de imagens (particularmente grafos *And-Or*), para reconhecimento de padrões em imagens e geração de imagens.

Idiomas dos artigos

Inglês, Português e Espanhol.

Listagem de Fontes

- Biblioteca Digital do IEEE (<http://ieeexplore.ieee.org>)
- Biblioteca Digital da ACM (<http://dl.acm.org/>)
- ISI Web of Knowledge (<http://apps.webofknowledge.com>)
- SCOPUS (<http://www.scopus.com>)
- Teses da USP (<http://www.teses.usp.br/>)
- Banco de teses da Capes (<http://servicos.capes.gov.br/capesdw/>)

Critérios de Inclusão

- CI-1: Serão incluídos os trabalhos que proponham algoritmos de manipulação de grafos *And-Or* (algoritmos de aprendizado do grafo, estimação de parâmetros e análise sintática) para geração e reconhecimento de imagens.
- CI-2: Serão incluídos todos os estudos encontrados, independentemente do ano de publicação.

Critérios de Exclusão

- CE-1: Serão excluídos trabalhos que não estejam disponíveis integralmente em bases de dados científicas ou em versões impressas.
- CE-2: Serão excluídos trabalhos escritos em outro idioma que não seja inglês, português ou espanhol.

Tipos de Documentos

Serão considerados dissertações de mestrado, teses de doutorado, artigos e publicações em revistas, conferências ou periódicos.

Critérios de Qualidade

- CQ-1: Se a proposta do estudo foi descrito de forma clara e adequada.
- CQ-2: Se os métodos ou técnicas utilizadas em os estudos primários foram relatados com clareza.
- CQ-3: Se os algoritmos de manipulação de grafos *And-Or* utilizados em estudos primários foram relatados com clareza.

- CQ-4: Se a proposta do estudo foi avaliada/validada.
- CQ-5: Se os resultados do estudo foram apresentados com clareza.

Processo de Seleção

Serão construídas *strings* utilizando como base as palavras-chave e seus sinônimos. Estas *strings* serão submetidas às máquinas de busca. Após a leitura do resumo e aplicação dos critérios de inclusão e exclusão, o trabalho será selecionado se confirmada a sua relevância pelo principal revisor (aluno). Se houver dúvida da relevância os demais revisores serão consultados (orientador e coorientador).

Extração de Informação

Serão preenchidos formulários de extração de dados para cada um dos estudos considerados como válidos. Além das informações básicas (dados bibliográficos, data de publicação, *abstract*, entre outros), estes formulários deverão conter a síntese do trabalho, destacando não só os métodos e as técnicas utilizadas, mas também os resultados obtidos, discussões e conclusões.

Sumarização dos Resultados

Após a leitura e o resumo dos trabalhos selecionados, será realizado uma análise quantitativa e qualitativa. Estas deverão ser documentadas na forma de um relatório técnico ou documento equivalente. Ao término da revisão bibliográfica, deverão ser levantadas as possíveis lacunas existentes no estado da arte que poderão ser exploradas.

APÊNDICE G – Aquisição de Imagens

Objetivo:

Construir uma base de imagens de expressões faciais com o intuito de auxiliar na investigação e desenvolvimento de pesquisas na área de reconhecimento e geração de imagens para diagnóstico e treinamento de transtornos psiquiátricos,

Procedimento:

O participante será filmado ao executar expressões faciais. A execução da expressão será estimulada de duas formas: (i) solicitação para expressar uma emoção e (ii) exibição de vídeos que visam a induzir uma emoção. A coleta de imagens obedecerá aos seguintes passos:

- 1.Serão lidos o objetivo e o procedimento do experimento.
- 2.O participante será convidado a ler e assinar o Termo de Consentimento Livre e Esclarecido.
- 3.Será solicitado para o participante tirar qualquer tipo de oclusão do rosto (óculos, brincos, etc.) se tiver.
- 4.O participante será convidado a se sentar em uma cadeira (regulável em altura), posicionado em linha reta diante da câmera, e de frente para um fundo branco proporcionado por uma tela de projeção.
- 5.A sessão começa solicitando ao participante fazer uma face neutra (sem nenhum tipo de emoção).
- 6.Na sequência, ou fase de indução, o pesquisador irá exibir uma série de vídeos para estimular ou induzir uma determinada emoção no participante. A ordem dos vídeos deverá estimular as seguintes emoções: satisfação, tristeza, aversão, surpresa, medo e finalmente raiva.
- 7.Na fase seguinte, ao qual chamaremos de expressão controlada, será solicitado para o participante fazer ou imitar uma determinada emoção, na mesma ordem mencionada no passo 6.
- 8.No final, o gênero e a idade do participante serão registrados.

ANEXO A – Aprovação do Comitê de Ética do Projeto de Formação de um Banco de Dados de Expressões Faciais

DETALHAR PROJETO DE PESQUISA

- DADOS DA VERSÃO DO PROJETO DE PESQUISA

Título da Pesquisa: Formação de bases de imagens de expressões faciais para diagnóstico e treinamento de transtornos psiquiátricos.

Pesquisador Responsável: Liseth Ury Segundo Carpio

Área Temática:

Versão: 1

CAAE: 47947115.2.0000.0065

Submetido em: 02/08/2015

Instituição Proponente: UNIVERSIDADE DE SAO PAULO

Situação da Versão do Projeto: Aprovado

Localização atual da Versão do Projeto: Pesquisador Responsável

Patrocinador Principal: UNIVERSIDADE DE SAO PAULO

Comprovante de Recepção: PB_COMPROVANTE_RECEPCAO_549585

- DOCUMENTOS DO PROJETO DE PESQUISA

- ▼ Versão Atual Aprovada (PO) - Versão 1
 - ▼ Projeto Original (PO) - Versão 1
 - ▼ Documentos do Projeto
 - Folha de Rosto - Submissão 1
 - Informações Básicas do Projeto - Subm
 - Outros - Submissão 1
 - Projeto Detalhado / Brochura Investigac
 - TCLE / Termos de Assentimento / Justif
 - ▼ Apreciação 1 - Faculdade de Medicina da U
 - ▼ Projeto Completo

Tipo de Documento

Situação

Arquivo

Postagem

Ações

- LISTA DE APRECIACÕES DO PROJETO

Apreciação ⁺	Pesquisador Responsável ⁺	Versão ⁺	Submissão ⁺	Modificação ⁺	Situação ⁺	Exclusiva do Centro Coord. ⁺	Ações
PO	Liseth Ury Segundo Carpio	1	02/08/2015	10/09/2015	Aprovado	Não

- HISTÓRICO DE TRÂMITES

Apreciação	Data/Hora	Tipo Trâmite	Versão	Perfil	Origem	Destino	Informações
PO	10/09/2015 18:55:23	Parecer liberado	1	Coordenador	Faculdade de Medicina da Universidade de São Paulo - FMUSP	PESQUISADOR	
PO	08/09/2015 21:42:35	Parecer do colegiado emitido	1	Coordenador	Faculdade de Medicina da Universidade de São Paulo - FMUSP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	
PO	31/08/2015 14:19:37	Parecer do relator emitido	1	Membro do CEP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	
PO	31/08/2015 14:15:05	Aceitação de Elaboração de Relatoria	1	Membro do CEP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	
PO	06/08/2015 22:54:18	Confirmação de Indicação de Relatoria			Faculdade de Medicina da Universidade de São Paulo - FMUSP		
PO	06/08/2015 18:03:14	Indicação de Relatoria			Faculdade de Medicina da Universidade de São Paulo - FMUSP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	
PO	06/08/2015 18:02:16	Aceitação do PP			Faculdade de Medicina da Universidade de São Paulo - FMUSP	Faculdade de Medicina da Universidade de São Paulo - FMUSP	

ANEXO B – Termo de Consentimento Livre e Esclarecido

Termo de Consentimento Livre e Esclarecido

Prezado(a),

O(A) Sr(a) está sendo convidado(a) a participar, como voluntário(a), em uma pesquisa acadêmica intitulada **“Formação de bases de imagens de expressões faciais para diagnóstico e treinamento de transtornos psiquiátricos”**, que tem como objetivo construir uma base de imagens de expressões faciais com o intuito de auxiliar na investigação e desenvolvimento de pesquisas na área de reconhecimento e geração de imagens para diagnóstico e treinamento de transtornos psiquiátricos.

Os experimentos estão divididos nas seguintes fases: **1** - Fornecimento de instruções sobre a tarefa a ser realizada, além de informações sobre os riscos ao(a) voluntário(a) (tempo estimado de 5 minutos); **2** - Efetiva participação do(a) voluntário(a), participação da sessão de gravação (tempo estimado de 30 minutos, considerando a fase de estimulação e de expressão controlada); **3** - Coleta do perfil do(a) voluntário(a) (tempo estimado de 5 minutos).

Os dados obtidos (vídeos) serão empregados para obter imagens de faces representando as várias emoções faciais (neutra, satisfação, tristeza, raiva, nojo, medo e surpresa). Cada um dos vídeos serão armazenados em computador para processamento posterior. No processamento posterior, serão selecionadas algumas imagens de cada vídeo, que melhor representem cada uma das emoções a fim de que sejam identificados os pontos da face que se movimentam ao expressar uma emoção. A identidade do participante não será divulgada na fase do processamento de imagens.

A participação efetiva na sessão consiste, numa primeira fase, assistir vídeos com o intuito de estimular a expressão de uma determinada emoção. Os movimentos serão capturados por uma câmera situada em linha reta em frente ao participante. Em uma segunda fase, ao qual chamaremos de expressão controlada, será solicitado para o participante fazer ou imitar uma determinada emoção, na mesma ordem mencionada cima.

Alguns riscos e desconfortos ao(a) voluntário(a) podem ser: **Constrangimento pelos conteúdos dos vídeos que irá assistir** – a sessão será acompanhada pelo pesquisador, que estará disponível para esclarecimento de quaisquer dúvidas ou no caso de alguma situação imprevisível; **Constrangimento pela violação do anonimato** - os dados coletados em cada sessão de experimentos receberão uma identificação associada ao Termo de Consentimento Livre e Esclarecido, para rastreabilidade da identidade do(a) voluntário(a), ressaltando os casos de interesse do(a) voluntário(a) no acesso aos dados de seus testes ou ocorrência de retirada do consentimento por parte deste(a). Para todos os demais efeitos, os dados serão tratados sem qualquer associação à identidade do(a) voluntário(a); **Sentimento de obrigação de participar** – a participação na pesquisa é de caráter voluntário e o participante pode se retirar em qualquer momento que desejar; **Medidas de higiene** - visando à prevenção de contágio de doenças de qualquer natureza entre os(as) voluntários(as), todo o equipamento que tiver contato direto com os mesmos, como a cadeira utilizada, será higienizado a cada sessão de teste; **Cansaço visual ou de outra natureza** – devido à participação em diversos testes para representar diferentes emoções, o(a) voluntário(a) poderá apresentar cansaço visual ou de outra natureza. Nesses casos, poderá solicitar a interrupção da coleta das

RUBRICA DO SUJEITO DE PESQUISA

RUBRICA DO PESQUISADOR

imagens temporária ou definitivamente, a qual será atendida pelo pesquisador.

Informações sobre a pesquisa ou dúvidas sobre a utilização do sistema computacional poderão ser solicitadas a qualquer momento, direto ao pesquisador responsável, que acompanhará os experimentos. A participação na pesquisa não acarretará nenhum tipo de despesa ao(à) voluntário(a). Indenização por danos comprovadamente causados pelo experimento será de incumbência do pesquisador responsável. Os endereços e telefones de contato dos pesquisadores encontram-se no final deste documento. Deve-se ressaltar que a participação é voluntária e que não haverá nenhuma espécie de remuneração.

Pelo presente instrumento que atende às exigências legais, o(a) Sr.(a),

portador(a) da R.G/R.N.E. nº _____, sexo _____,
nascido(a) em ____/____/_____, telefone (____) _____, e-mail
_____, profissão _____, declara
que foi realizada leitura minuciosa das informações constantes neste Termo de Consentimento
Livre e Esclarecido, com explicações detalhadas, ficando ciente dos procedimentos aos quais
será submetido(a). Deve-se ressaltar que o(a) voluntário(a) pode a qualquer momento retirar
seu Consentimento Livre e Esclarecido, deixando de participar da pesquisa e ciente de que
todas as informações prestadas tornar-se-ão confidenciais e guardadas por força de sigilo
profissional.

O voluntário(a) terá acesso aos resultados da pesquisa por meio dos artigos publicados
ou diretamente com o pesquisador.

O pesquisador responsável compromete-se a cumprir todas as exigências contidas no
item IV. 3 da Resolução do CNS/MS nº 466, de dezembro de 2.012, publicada em 13 de junho
de 2.013.

Por fim, o(a) voluntário(a) declara que, após convenientemente esclarecido(a) pelo
pesquisador e ter entendido o que foi explicado, aceita participar da pesquisa **“Formação de
bases de imagens de expressões faciais para diagnóstico e treinamento de transtornos
psiquiátricos”**, firmando o termo em duas vias (uma via para o(a) voluntário(a) e outra para
o pesquisador responsável), que serão rubricadas em todas as suas páginas e assinadas ao seu
término.

Local: _____

Data: ____/____/____

Assinatura do pesquisador

Assinatura do(a) voluntário(a)

Pesquisador Responsável: Liseth Urpy Segundo Carpio

Orientador: Profa. Dra. Arianne Machado Lima

Escola Artes, Ciências e Humanidades da Universidade de São Paulo: Av. Arlindo Béttio, 1000
Ermelino Matarazzo - CEP: 03828-000 - São Paulo - SP - Telefone: (11) 2648-0067 – e-mail:
liseth.segundo@usp.br.

Observação: Em caso de dúvida, sugestão ou reclamação, o(a) Sr.(a) também pode procurar o
Comitê de Ética e Pesquisa – da Escola de Artes, Ciências e Humanidades da Universidade de
São Paulo - Av. Arlindo Béttio, 1000 Ermelino Matarazzo - CEP: 03828-000 - São Paulo - Sala
4 - I1 - e-mail: cep-each@usp.br – Telefone: (11) 3091-1046

RUBRICA DO SUJEITO DE PESQUISA

RUBRICA DO PESQUISADOR