

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**

**CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA**

**Coordenação de Pós-Graduação em Informática**

**DISSERTAÇÃO DE MESTRADO**

**GERAÇÃO DE REPRESENTAÇÕES NÃO FOTO-REALÍSTICAS**

**A PARTIR DE FOTOGRAFIAS DIGITAIS:**

**UM ESTUDO DE CASO COM FACES HUMANAS**

**FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA**

**ORIENTADORES:**

**JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ**

**HERMAN MARTINS GOMES**

**CAMPINA GRANDE, PARAÍBA**

**AGOSTO – 2009**

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE**

**CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA**

**Coordenação de Pós-Graduação em Informática**

**GERAÇÃO DE REPRESENTAÇÕES NÃO FOTO-REALÍSTICAS  
A PARTIR DE FOTOGRAFIAS DIGITAIS:  
UM ESTUDO DE CASO COM FACES HUMANAS**

**FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA**

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Informática do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários à obtenção do grau de Mestre em Ciência da Computação (MSc).

Área de concentração: Ciência da Computação

Linha de pesquisa: Modelos Computacionais e Cognitivos

José Eustáquio Rangel de Queiroz

Herman Martins Gomes

Orientadores

Campina Grande - Paraíba

Agosto – 2009

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A779g  
2009

Arruda, Francisco de Assis Pereira Vasconcelos de

Geração de representações não foto-realísticas a partir de fotografias digitais: um estudo de caso com faces humanas / Francisco de Assis Pereira Vasconcelos de Arruda – Campina Grande, 2009.

120 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Referências.

Orientadores: Prof. Dr. José Eustáquio Rangel de Queiroz.

Prof. Ph.D. Herman Martins Gomes.

1. Inteligência Artificial – Aprendizagem de Máquina 2. Processamento de imagens 3 Representações não foto-realísticas I. Título

CDU 004.855.5(043)

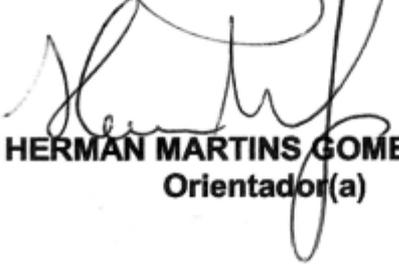
**"GERAÇÃO DE REPRESENTAÇÕES NÃO FOTO-REALÍSTICAS A PARTIR DE FOTOGRAFIAS DIGITAIS: UM ESTUDO DE CASO COM FACES HUMANAS"**

**FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA**

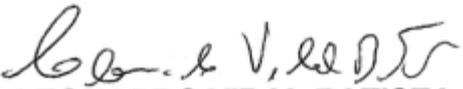
**DISSERTAÇÃO APROVADA EM 28.08.2009**



**JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ, D.SC**  
Orientador(a)



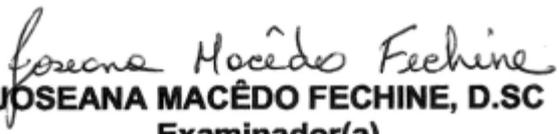
**HERMAN MARTINS GOMES, PH.D**  
Orientador(a)



**LEONARDO VIDAL BATISTA, D.SC**  
Examinador(a)



**JOÃO MARQUES DE CARVALHO, PH.D**  
Examinador(a)



**JOSEANA MACÊDO FECHINE, D.SC**  
Examinador(a)

**CAMPINA GRANDE - PB**

## **AGRADECIMENTOS**

À minha família, pelo apoio.

Aos orientadores Herman Martins Gomes e José Eustáquio Rangel de Queiroz, pela dedicação e paciência durante a orientação.

Aos membros da banca examinadora, pelas críticas e sugestões que contribuíram para o enriquecimento deste trabalho.

Aos *chickens* do projeto iPhotoBot e dissidentes do Laboratório de Visão Computacional: Brito, Bruno Alexandre, Saulo, Cláudio, Luciana, Eanes, Einstein, Paulo, Eduardo, Odilon, Vinas, Fernando, Bosco, Luiz Gonzaga, Thiago *Harry Potter*, Caio, Alexandre e Kimmi.

Aos desenvolvedores do *software* SyncToy.

A todos que fazem a COPIN, dentre professores e funcionários, em especial Aninha e Vera, sempre prontas a atender aos pedidos urgentíssimos, em prazos curtíssimos.

O presente trabalho de dissertação foi alcançado em cooperação com a Hewlett-Packard Brasil Ltda. e com recursos provenientes da Lei de Informática (Lei nº 8.248, de 1991).

## RESUMO

A renderização não foto-realística é uma área comum à computação gráfica e ao processamento digital de imagens que visa à geração de imagens simples e compreensíveis a partir de cenas complexas em diferentes níveis de representações. Nesse contexto, esboços constituem uma das formas de representação não foto-realística mais comum e efetiva, sendo usados como base para animações, no entretenimento e na publicidade. O presente trabalho teve como objetivo o desenvolvimento de uma técnica automática para a geração de representações não foto-realísticas robustas que se assemelham a esboços semi-detalhados, com avaliação centrada em faces humanas.

Os seguintes problemas foram investigados: (i) *pré-processamento* (correção e conversão cromática, suavização e transformação histogrâmica) de imagens contendo faces e verificação do impacto na melhoria da representação gerada; (ii) *utilização de redes neurais* para a detecção de contornos; (iii) *pós-processamento* para a melhoria do resultado final e redução de artefatos espúrios; (iv) *avaliação experimental* das representações produzidas pelo sistema, usando métricas objetivas (PSNR, FOM, SSIM); e (v) *validação* dos resultados produzidos pelo sistema, a partir de um processo experimental de avaliação subjetiva fundamentado em um esquema de votação dos resultados por usuários de teste.

Palavras-chave: renderização não foto-realística, geração de representações não foto-realísticas, detecção de bordas.

## ABSTRACT

Non-photorealistic rendering (NPR) is an area common to both computer graphics and digital image processing, which focuses on generating simple and comprehensible images from complex scenes, using various levels of abstraction. In this context, sketches are one of the most common and efficient illustration abstractions, being used in many fields, such as animation schemes, entertainment, and advertisement. In this work a technique to automatically enable NPR renderings, which resemble semi-detailed sketches of human frontal faces, from digital images is presented and evaluated.

The following aspects of the sketch generation have been addressed: (i) *pre-processing* (color constancy, blurring and histogram transformation) of face images and verification of image quality improvement; (ii) *neural network use* for edge and boundary detection; (iii) *post-processing* for improving final results and removing spurious artifacts; (iv) *experimental assessment* of the produced renderings by objective metrics (PSNR, FoM, SSIM); (v) and *validation* of the results by an experimental subjective assessment, based on a user voting scheme.

Keywords: Non-photorealistic rendering, semi-detailed sketch generation, edge detection.

# CONTEÚDO

<b>1. INTRODUÇÃO</b> .....	<b>1</b>
<b>1.1. MOTIVAÇÃO</b> .....	<b>2</b>
<b>1.2. OBJETIVOS E RELEVÂNCIA</b> .....	<b>5</b>
<b>1.3. ESTRUTURA DA DISSERTAÇÃO</b> .....	<b>7</b>
<b>2. TRABALHOS RELACIONADOS</b> .....	<b>8</b>
<b>2.1. BREVE HISTÓRICO</b> .....	<b>10</b>
<b>2.2. GERAÇÃO DE ESBOÇOS</b> .....	<b>13</b>
<b>2.3. CONSIDERAÇÕES FINAIS DO CAPÍTULO</b> .....	<b>23</b>
<b>3. ABORDAGEM PROPOSTA</b> .....	<b>25</b>
<b>3.1. PRÉ-PROCESSAMENTO</b> .....	<b>27</b>
<b>3.1.1. Correção Cromática</b> .....	<b>27</b>
<b>3.1.2. Conversão de Espaço Cromático</b> .....	<b>28</b>
<b>3.1.3. Suavização</b> .....	<b>29</b>
<b>3.2. DETECÇÃO DE BORDAS</b> .....	<b>30</b>
<b>3.2.1. Treinamento Usando Ground-truth</b> .....	<b>37</b>
<b>3.2.2. Treinamento Usando Padrões Sintéticos</b> .....	<b>40</b>
<b>3.3. PÓS-PROCESSAMENTO</b> .....	<b>43</b>
<b>3.3.1. Transformação de Histograma</b> .....	<b>44</b>
<b>3.4. DETALHES DE PROJETO E IMPLEMENTAÇÃO</b> .....	<b>46</b>
<b>3.5. CONSIDERAÇÕES FINAIS DO CAPÍTULO</b> .....	<b>47</b>
<b>4. APRESENTAÇÃO E DISCUSSÃO DOS EXPERIMENTOS E RESULTADOS</b> ...	<b>49</b>
<b>4.1. CURVA ROC</b> .....	<b>50</b>
<b>4.2. EXPERIMENTOS OBJETIVOS</b> .....	<b>52</b>
<b>4.2.1. Relação Sinal/Ruído Máxima (PSNR)</b> .....	<b>55</b>
<b>4.2.2. Similaridade Estrutural (SSIM)</b> .....	<b>56</b>
<b>4.2.3. Figura de Mérito (FoM)</b> .....	<b>57</b>
<b>4.2.4. Comparação Objetiva entre Detectores</b> .....	<b>59</b>
<b>4.3. EXPERIMENTO SUBJETIVO</b> .....	<b>65</b>
<b>4.3.1. Primeira Parte do Experimento</b> .....	<b>66</b>
<b>4.3.2. Segunda Parte do Experimento</b> .....	<b>69</b>
<b>4.4. CONSIDERAÇÕES FINAIS DO CAPÍTULO</b> .....	<b>73</b>
<b>5. CONSIDERAÇÕES FINAIS</b> .....	<b>76</b>
<b>5.1. SUMÁRIO DA DISSERTAÇÃO</b> .....	<b>76</b>
<b>5.2. CONTRIBUIÇÕES DO TRABALHO REALIZADO</b> .....	<b>77</b>

<b>5.3. TRABALHOS FUTUROS .....</b>	<b>78</b>
<b>APÊNDICE A. ALGUMAS TÉCNICAS DE PDI ADOTADAS NESTA PESQUISA</b>	<b>89</b>
<b>A.1. TÉCNICAS PARA PRÉ-PROCESSAMENTO.....</b>	<b>89</b>
<b>A.1.1. Correção Cromática .....</b>	<b>90</b>
<b>A.1.2. Conversão de Espaço Cromático .....</b>	<b>96</b>
<b>A.1.3. Suavização .....</b>	<b>99</b>
<b>A.2. TÉCNICAS PARA PÓS-PROCESSAMENTO .....</b>	<b>103</b>
<b>A.2.1. Transformação Histogrâmica.....</b>	<b>103</b>
<b>APÊNDICE B. RESULTADOS ADICIONAIS OBTIDOS A PARTIR DO USO DA ABORDAGEM PROPOSTA .....</b>	<b>106</b>
<b>APÊNDICE C. IMAGENS UTILIZADAS NO EXPERIMENTO OBJETIVO.....</b>	<b>108</b>
<b>APÊNDICE D. TESTE DE TUKEY PARA VALORES DE FOM .....</b>	<b>111</b>
<b>APÊNDICE E. MÉTRICAS PARA AVALIAÇÃO DE DETECTORES DE BORDAS</b>	<b>113</b>
<b>E.1. RELAÇÃO SINAL/RUÍDO MÁXIMA (PSNR).....</b>	<b>114</b>
<b>E.2. SIMILARIDADE ESTRUTURAL (SSIM).....</b>	<b>115</b>
<b>E.3. FIGURA DE MÉRITO (FOM) .....</b>	<b>116</b>
<b>APÊNDICE F. IMAGENS UTILIZADAS NO EXPERIMENTO SUBJETIVO .....</b>	<b>118</b>

## LISTA DE TABELAS

Tabela 2.1	– Sumário das técnicas de geração de esboços. ....	18
Tabela 3.1	– Número de padrões usados na determinação do vetor médio para ... PCA. ....	34
Tabela 3.2	– Distribuição das amostras.....	39
Tabela 3.3	– Distribuição de padrões de borda sintéticos.....	42
Tabela 4.1	– Métricas usadas na análise ROC.....	50
Tabela 4.2	– AUC para curvas ROC. ....	52
Tabela 4.3	– Avaliação objetiva usando PSNR.....	55
Tabela 4.4	– Teste F ANOVA para valores de PSNR. ....	55
Tabela 4.5	– Avaliação objetiva usando SSIM. ....	56
Tabela 4.6	– Teste F ANOVA para valores de SSIM.....	56
Tabela 4.7	– Avaliação objetiva usando FoM. ....	57
Tabela 4.8	– Teste F ANOVA para valores de FoM.....	57
Tabela 4.9	– Comparação objetiva entre a abordagem proposta, DoG e FDoG...	60
Tabela 4.10	– Teste ANOVA para valores de PSNR. ....	61
Tabela 4.11	– Teste ANOVA para valores de SSIM.....	61
Tabela 4.12	– Teste ANOVA para valores de FoM. ....	61
Tabela 4.13	– Teste de Tukey ( <i>Honestly Significant Differences</i> ) para valores de PSNR.....	62
Tabela 4.14	– Teste de Tukey ( <i>Honestly Significant Differences</i> ) para valores de SSIM. ....	62
Tabela 4.15	– Teste de Tukey ( <i>Honestly Significant Differences</i> ) para valores de FoM. ....	62
Tabela 4.16	– Comparação visual – PSNR. ....	63
Tabela 4.17	– Comparação visual – SSIM.....	63
Tabela 4.18	– Comparação visual – FoM.....	64
Tabela 4.19	– Categorias de fotos para experimento subjetivo.....	66
Tabela 4.20	– Melhores parâmetros por detector.....	68
Tabela 4.21	– Desempenho relativo por detector.....	71
Tabela 4.22	– Teste ANOVA para o experimento subjetivo – parte 2.....	72
Tabela 4.23	– Teste de Tukey para o experimento subjetivo – parte 2.....	72
Tabela D.1	– Grupos comparados no teste de Tukey. ....	111
Tabela D.2	– Teste de Tukey ( <i>Honestly Significant Differences</i> ) para valores de FoM. ....	111
Tabela F.1	– Imagens utilizadas no experimento subjetivo. ....	119

## LISTA DE FIGURAS

Figura 1.1	– Esboços renderizados a partir de cenas 3D sintéticas. ....	4
Figura 1.2	– Esboços gerados a partir de imagens digitais 2D.....	5
Figura 2.1	– Delimitação do escopo do trabalho: geração de esboços. ....	8
Figura 2.2	– Exemplo de renderização foto-realística. ....	9
Figura 2.3	– Exemplos de renderizações não foto-realísticas.....	12
Figura 2.4	– Uso de NPR para melhorar a comunicação em mapas: (A) Mapa completo; (B) Mapa feito por uma pessoa; e (C) Mapa gerado pelo sistema proposto por Agrawala e Stolte (2001). ....	13
Figura 2.5	– Exemplo de uma representação semi-detalhada em um diagrama técnico. ....	14
Figura 2.6	– Importância das bordas: (A) Aparato usado para rastreamento dos olhos; (B) Imagem segmentada sem bordas; (B) Bordas detectadas; e (C) Resultado final mais expressivo. ....	15
Figura 2.7	– Esboços gerados a partir de argumentos estatísticos: (A) Entrada; e (B) Resultado. ....	16
Figura 2.8	– Exemplo de esboço gerado usando a técnica proposta por Tresset e Fol-Leymarie. ....	17
Figura 2.9	– Filtro DoG aprimorado: (A) Entrada; e (B) FDoG .....	18
Figura 2.10	– Arredondamento de arestas na detecção de bordas: (A) Canny (B) Neural. ....	21
Figura 2.11	– Bordas em imagem com ruído: (A) Entrada; (B) Sobel; (C) Canny; e (D) Neural.....	22
Figura 2.12	– Bordas usando abordagem neural: (A) Entrada; (B) Laplaciano; e (C) Neural. ....	23
Figura 3.1	– Macro-arquitetura da técnica proposta.....	25
Figura 3.2	– Etapa 1 – Correção cromática: (A) Imagem de entrada; e (B) <i>GrayWorld</i> . ....	28
Figura 3.3	– Etapa 2 – Conversão para o espaço YCbCr: (A) Original; (B) Canal Y; (C) Canal Cb; e (D) Canal Cr.....	29
Figura 3.4	– Etapa 3 – Suavização: (A) Canal Y; e (B) Imagem Suavizada.....	30
Figura 3.5	– Resultado com rede de Hopfield: (A) Original; (B) Bordas. ....	36
Figura 3.6	– Topologia da Rede Neural.....	36
Figura 3.7	– Imagem de treinamento: (A) Original; e (B) <i>Ground-truth</i> . ....	37
Figura 3.8	– Seleção dos padrões de borda e não-borda. ....	38
Figura 3.9	– Etapa 4 – Bordas usando treinamento com <i>Ground-truth</i> : (A) Suavizada; e (B) Bordas.....	40

<b>Figura 3.10</b>	<b>– Exemplos de padrões sintéticos: (A) Padrões de borda; e (B) Padrões de não-borda. ....</b>	<b>41</b>
<b>Figura 3.11</b>	<b>– Etapa 4 – Detecção de bordas usando treinamento com padrões sintéticos: (A) Imagem Suavizada; e (B) Bordas.....</b>	<b>43</b>
<b>Figura 3.12</b>	<b>– Etapa 5 - Ajuste de brilho e Contraste: (A) Mapa de Bordas; e (B) Imagem modificada. ....</b>	<b>44</b>
<b>Figura 3.13</b>	<b>– Etapa 6 – Transformação histogrâmica: (A) Bordas após ajuste de brilho e Contraste; e (B) Imagem final.....</b>	<b>45</b>
<b>Figura 3.14</b>	<b>– Comparação: (A) Treinamento com GT; e (B) Treinamento com bordas sintéticas.....</b>	<b>45</b>
<b>Figura 3.15</b>	<b>– Diagrama de classes do software desenvolvido.....</b>	<b>47</b>
<b>Figura 4.1</b>	<b>– Curva ROC para os classificadores estudados.....</b>	<b>51</b>
<b>Figura 4.2</b>	<b>– Etapa 1 do experimento subjetivo: pré-seleção dos parâmetros. .</b>	<b>67</b>
<b>Figura 4.3</b>	<b>– Imagens mais e menos votadas – Experimento subjetivo - Etapa 1.....</b>	<b>69</b>
<b>Figura 4.4</b>	<b>– Experimento subjetivo - Etapa 2 – Pontuação. ....</b>	<b>71</b>
<b>Figura 4.5</b>	<b>– Comparação visual do melhor desempenho por detector. ....</b>	<b>73</b>
<b>Figura A.1</b>	<b>– Exemplo de falha do algoritmo <i>GrayWorld</i>. ....</b>	<b>92</b>
<b>Figura A.2</b>	<b>– Caminho entre <math>j</math> e <math>i</math>. A luminosidade relativa é computada como a razão entre as resposta visuais dos pontos <math>x</math> e <math>x+1</math>. ....</b>	<b>93</b>
<b>Figura A.3</b>	<b>– Exemplos de caminhos randômicos (<math>j_1 \rightarrow j_n</math>) para a computação da luminosidade em um ponto <math>i</math>. ....</b>	<b>94</b>
<b>Figura A.4</b>	<b>– Correção cromática: (A) Original; (B) Expansão; (C) <i>GrayWorld</i>; e (D) Retinex. ....</b>	<b>96</b>
<b>Figura A.5</b>	<b>– Conversão para o espaço cromático YCbCr.....</b>	<b>98</b>
<b>Figura A.6</b>	<b>– Cone HSV.....</b>	<b>98</b>
<b>Figura A.7</b>	<b>– Conversão para o espaço cromático HSV.....</b>	<b>99</b>
<b>Figura A.8</b>	<b>– Visualização da distribuição gaussiana em 3D.....</b>	<b>101</b>
<b>Figura A.9</b>	<b>– Suavização: (A) Original; (B) Gaussiana 3x3; (C) Gaussiana 5x5; (D) Duas iterações do filtro da mediana 5x5; e (E) Duas iterações do filtro da mediana 3x3. ....</b>	<b>103</b>
<b>Figura A.10</b>	<b>– Curvas de potência para operador exponencial. ....</b>	<b>104</b>
<b>Figura A.11</b>	<b>– Operador exponencial: (A) Imagem cinza; e (B) Após transformação.....</b>	<b>104</b>

## LISTA DE SIGLAS E ABREVIATURAS

<b>ANOVA</b>	<b>AN</b> alysis <b>O</b> f <b>VA</b> riance
<b>API</b>	<b>A</b> pplication <b>P</b> rogram <b>I</b> nterface
<b>AUC</b>	<b>A</b> rea <b>U</b> nder the <b>C</b> urve
<b>DoG</b>	<b>D</b> ifference <b>o</b> f <b>G</b> aussians
<b>F</b>	<b>F</b> ixation indices (Estatística F)
<b>FDoG</b>	<b>F</b> low-Based <b>D</b> ifference <b>o</b> f <b>G</b> aussians
<b>FoM</b>	<b>F</b> igure <b>o</b> f <b>M</b> erit
<b>gl</b>	<b>g</b> raus de <b>l</b> iberdade
<b>GT</b>	<i><b>G</b>round-<b>T</b>ruth</i>
<b>HSV</b>	<b>H</b> ue, <b>S</b> aturation, <b>V</b> alue
<b>IDE</b>	<b>I</b> ntegrated <b>D</b> evelopment <b>E</b> nvironment
<b>JPEG</b>	<b>J</b> oint <b>P</b> hotographic <b>E</b> xperts <b>G</b> roup
<b>MATLAB</b>	<b>MAT</b> rix <b>LAB</b> oratory
<b>MLP</b>	<b>M</b> ulti- <b>L</b> ayer <b>P</b> erceptron
<b>MQ</b>	<b>M</b> édia dos <b>Q</b> uadrados
<b>MSE</b>	<b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>NPR</b>	<b>N</b> on- <b>P</b> hotorealistic <b>R</b> endering
<b>OpenCV</b>	<b>O</b> pen <b>S</b> ource <b>C</b> omputer <b>V</b> ision (Intel Library)
<b>PCA</b>	<b>P</b> rincipal <b>C</b> omponent <b>A</b> nalysis
<b>PCM</b>	<b>P</b> ixel <b>C</b> orrespondence <b>M</b> etric
<b>PDI</b>	<b>P</b> rocessamento <b>D</b> igital de <b>I</b> magens
<b>PSNR</b>	<b>P</b> eak <b>S</b> ignal-to- <b>N</b> oise <b>R</b> atio
<b>RGB</b>	<b>R</b> ed, <b>G</b> reen, <b>B</b> lue
<b>ROC</b>	<b>R</b> eceiver <b>O</b> perating <b>C</b> haracteristics
<b>SQ</b>	<b>S</b> oma dos <b>Q</b> uadrados
<b>SSIM</b>	<b>S</b> tructural <b>S</b> IMilarity

# Capítulo 1

## Introdução

Enquanto vertente ativa na área do Processamento Digital de Imagens (PDI), a *renderização não foto-realística*, usualmente conhecida como NPR (*Non-Photorealistic Rendering*), visa à geração de imagens simples e compreensíveis, a partir de cenas complexas (Markosian, 2000). Mais precisamente, NPR reúne uma série de técnicas que enfatizam a comunicação visual e a ilustração conceitual, ao invés da fidelidade à realidade (Sayeed e Howard, 2006), envolvendo conceitos artístico-científicos e guiando-se mais pela percepção humana do que pela precisão física (Gooch, 2001). O termo *renderização estilizada* também pode ser usado para definir de forma mais expressiva tal classe de técnicas (Lee et al., 2007).

No contexto da confecção de ilustrações, vários níveis de representação da imagem podem ser gerados a partir de técnicas NPR, desde linhas elementares a formas mais elaboradas de representação. Enquanto representação não foto-realística, o *esboço* é uma forma de desenho simples, constituído de linhas salientes que definem características essenciais do objeto retratado. Assim sendo, pode-se, então, considerar uma representação não foto-realística como **robusta** quando, apesar de reunir apenas um número reduzido de linhas (ou traços), esta possibilita aos observadores a identificação do objeto retratado. Ao longo deste documento, o termo *robusto* será empregado com esta conotação, também compartilhada por Kang et al. (2005).

No contexto da NPR, o traçado de esboços ou a delineação de contornos constitui uma das formas mais comuns e efetivas de ilustração, os quais permitem expressar a informação de forma concisa, e.g., em representações gráficas técnicas e desenhos arquitetônicos preliminares,

cenar de pré-produção de quadrinhos (*comics*) e desenhos animados (*cartoons*). O objetivo primário de um esboço é delinear objetos, a partir do traçado de formas, com o propósito de capturar a idéia geral do objeto retratado (Sayeed e Howard, 2006).

Diante do exposto, nesta dissertação são discutidos os trabalhos no âmbito da renderização não foto-realística para a estilização de imagens, dando-se ênfase à confecção de representações não foto-realísticas, quer sob a forma de esboços preliminares ou mais elaborados. Em especial, no contexto da geração de representações desta natureza, buscou-se o delineamento robusto de aspectos marcantes em faces humanas frontais.

As seções remanescentes deste capítulo estão divididas como segue. Na Seção 1.1, argumenta-se sobre a motivação para esta pesquisa. Os objetivos a serem alcançados e a descrição do problema a ser resolvido são apresentados na Seção 1.2. Finalmente, na Seção 1.3, é apresentada a estrutura desta dissertação.

## **1.1. Motivação**

Técnicas de renderização *foto-realística* são focadas na simulação da realidade e na modelagem de aspectos físicos que a permeiam, enquanto técnicas de renderização *não foto-realística* (NPR) buscam a imitação de estilos artísticos. Em contraste a métodos tradicionais de renderização, o estilo dos traços e linhas usados em NPR pode agregar informação ao objeto retratado com o uso eficiente da orientação, espessura e tonalidade, e.g., ao ilustrar modelos de anatomia, NPR pode melhorar a visualização do modelo por meio da diversificação na orientação do traço, demonstrando variações na curvatura ou demonstrando variações na iluminação por meio de mudanças na espessura do traço (Lee et. al, 2007).

Na década de 90, a área de NPR começou a despertar interesse da comunidade científica. Exemplos do interesse crescente nesta área são: (i) a criação de uma seção dedicada ao tema no evento SIGGRAPH<sup>1</sup>, em 1998;

---

<sup>1</sup> SIGGRAPH: The 25<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques. (<http://www.siggraph.org/s98/>)

(ii) a criação de conferências tais como a Eurographics<sup>2</sup>, em 1999; e sobretudo (iii) o surgimento do primeiro simpósio voltado exclusivamente ao tema NPR - o NPAR<sup>3</sup>, em 2000. Outro marco importante é a publicação de dois livros-texto que abordam o tema, de autoria de Gooch e Gooch (2001) e Strothotte e Schlechtweg (2002).

Ressalta-se também a ênfase dada, nos últimos anos, a animações criadas a partir de cenas ou tomadas de vídeos com atores reais, a exemplo dos filmes *Avenue Amy* (1999), *Waking Life* (2001) e *A Scanner Darkly* (2006). Adicionalmente, a indústria de jogos eletrônicos tem encontrado e explorado algumas oportunidades no uso do aspecto visual típico dos *cartoons*, o que tem possibilitado a revitalização de séries consagradas como *Team Fortress 2* (2007), *Grand Theft Auto IV* (GTA4 - 2008) e *Street Fighter 4* (2009).

Dentre as aplicações de NPR destacam-se: (i) a transformação automática de vídeos ou fotos reais em representações que imitam desenhos ou pinturas (Wang et al., 2004; Kang, Lee e Chui, 2007; Kyprianidis & Döllner, 2008); (ii) o suporte a desenhistas de histórias em quadrinhos (Hwang et al., 2006); (iii) a animação de estilos clássicos de NPR (Gooch, 2001); (iv) a geração de caricaturas, com fins ao entretenimento e apelo publicitário (Gooch, Reinhard e Gooch, 2004; Shet et al., 2005); e (v) a transferência e automatização de estilos de desenho (Barla et al., 2006).

No contexto da geração de representações não foto-realísticas, a detecção de bordas é uma etapa relevante no contexto da geração de esboços. Adicionalmente, a detecção de bordas destaca-se como método coadjuvante em processos de extração de características (Wang et al., 2006), segmentação e reconhecimento (Jiang et al., 2006) e análise de imagens (Chabrier et al., 2008), sendo amplamente usada em vários algoritmos de visão computacional e de PDI, uma vez que possibilita o

---

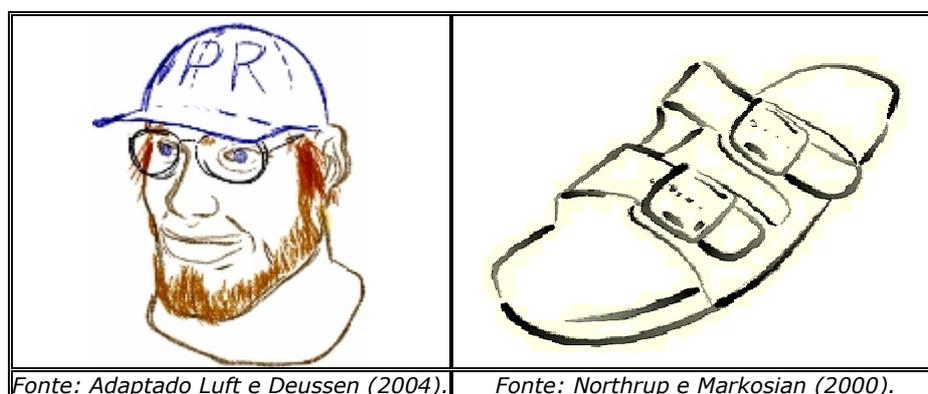
<sup>2</sup> Eurographics: 20<sup>th</sup> Annual Conference of the European Association of Computer Graphics. (<http://www.cg.tuwien.ac.at/conferences/VisSym99/>)

<sup>3</sup> NPAR: The 1<sup>st</sup> International Symposium on Non-Photorealistic Animation and Rendering. (<http://www.npar.org/>)

agrupamento de pixels que delimitam áreas que dividem a imagem em regiões (Chabrier et al., 2008). A delimitação em regiões serve de base para as etapas de pré-processamento e, por sua vez, o mapa de bordas serve de entrada para outros algoritmos destinados à detecção de objetos, à renderização não foto-realística e/ou à busca, detecção e reconhecimento de objetos.

É importante ressaltar que, nos primórdios das pesquisas em NPR, várias técnicas foram desenvolvidas, com o intuito de gerar cenas estáticas a partir de informações sintéticas em três dimensões (3D). Meier (1996) apresenta um dos primeiros trabalhos focado na renderização não foto-realística, a partir de animações em cenas 3D, seguido por Litwinowicz (1997), com um trabalho propondo renderizações estilizadas destinadas ao processamento de vídeos.

A criação de esboços e traços fundamentada na descrição sintética de uma cena 3D, ao invés do uso de informação puramente 2D, também foi tratada por vários autores nos últimos anos (e.g. Hertzmann, 1999; Northrup e Markosian, 2000; Kalnins et al., 2002; Luft e Deussen, 2004; Lewis et al., 2005; Lee, Kwon e Lee, 2006; Kolliopoulos, Wang e Hertzmann, 2006). Na Figura 1.1, estão contidos exemplos de esboços gerados a partir de cenas 3D sintéticas.



**Figura 1.1 – Esboços renderizados a partir de cenas 3D sintéticas.**

Todavia, sendo NPR uma área ampla e, embora constituindo um tópico de pesquisa promissor, a geração de esboços usando técnicas de *renderização* em 3D, conforme anteriormente mencionado, **não** faz parte do

escopo desta pesquisa. O foco desta pesquisa é restrito, pois, à geração de representações não foto-realísticas a partir de fotografias digitais estáticas, não fazendo uso de informações 3D (vide Figura 1.2). Deve-se ressaltar ainda que a técnica usada para gerar a Figura 1.2C é manual, sendo o resultado classificado pelo autor do esboço como esboço foto-realístico (*photorealistic sketch*). Trata-se da geração manual de uma representação elaborada da face, ressaltado pela presença de muitas gradações de cinza no resultado, a qual se define como um esboço detalhado da face.

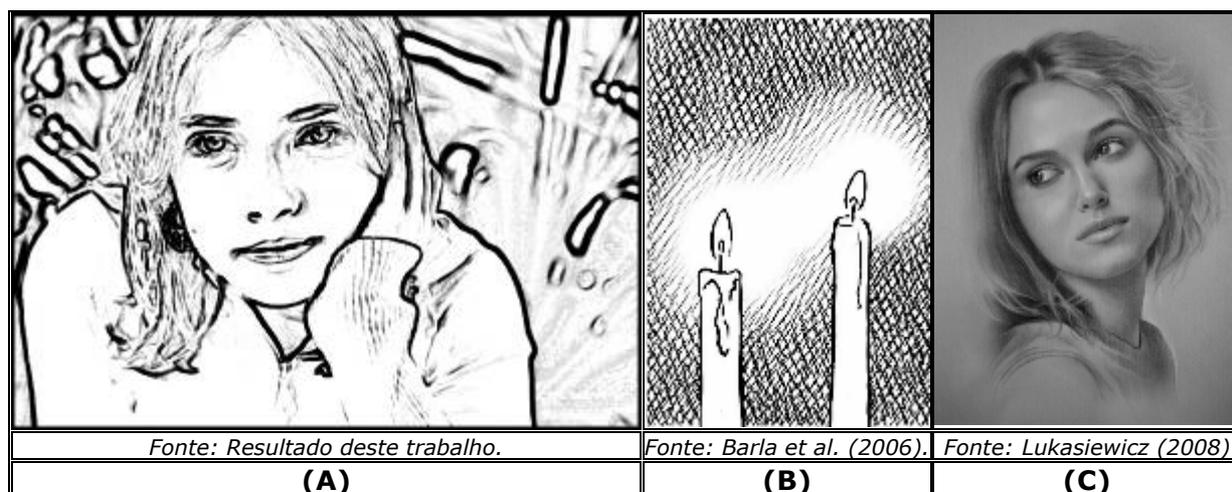


Figura 1.2 – Esboços gerados a partir de imagens digitais 2D.

Apesar de toda a pesquisa já desenvolvida até então, a detecção de silhuetas e geração de esboços - enquanto representações não foto-realísticas - centrados em faces humanas, de forma automática, precisa, robusta e com poucos artefatos indesejáveis, ainda constitui um problema cuja solução não foi esgotada. Dificuldades surgem devido a vários fatores, dentre os quais podem ser destacados: (i) variações nas condições de iluminação; (ii) resolução da imagem; (iii) dificuldades na geração de representações de forma automática; (iv) geração não aprimorada de esboços de faces; e (v) problemas com detecção de bordas usando algoritmos clássicos.

## 1.2. Objetivos e Relevância

Diante do contexto apresentado na seção anterior, nesta dissertação é proposta uma técnica robusta e automática de geração de representações

não foto-realísticas bidimensionais, a partir de fotografias digitais. Para validar a técnica, um estudo de caso foi conduzido, considerando-se fotografias de faces humanas frontais. Em face dos problemas apresentados pelas abordagens utilizadas atualmente, a técnica desenvolvida incorpora: (a) algoritmos para a compensação de iluminação, os quais visam aumentar a robustez da detecção de contornos e silhuetas fidedignas que formarão a base para o resultado gerado; (b) a detecção semi-detalhada de características de faces humanas; e (c) uma abordagem para a remoção de artefatos indesejados. A detecção semi-detalhada de características faciais compreende a detecção de aspectos faciais marcantes, e.g., olhos e boca, sem a presença de detalhes minuciosos, como evidenciado na Figura 1.2A, contrapondo-se ao nível de detalhes presente em uma representação não foto-realística detalhada (presente na Figura 1.2C).

A técnica tem como diferencial a agregação de módulos tanto para atenuar os efeitos advindos da falta de controle das condições de iluminação ao capturar a imagem quanto para utilizar técnicas de aprendizagem de máquina destinadas à detecção de bordas. Os objetivos específicos são:

(a) Aplicar técnicas de pré-processamento a imagens contendo faces e avaliar o impacto na melhoria das representações não foto-realísticas geradas;

(b) Aplicar técnicas de aprendizagem de máquina ao problema de detecção de bordas;

(c) Aplicar técnicas de pós-processamento ao resultado obtido, visando melhorar o resultado final e reduzir artefatos espúrios;

(d) Validar o estudo, a partir de uma aplicação de software capaz de gerar representações não foto-realísticas robustas<sup>4</sup> de imagens contendo faces humanas;

(e) Validar os resultados produzidos pelo sistema a partir de métricas objetivas; e

(f) Validar os resultados produzidos pelo sistema mediante uma

---

<sup>4</sup> Considera-se uma representação robusta quando, apesar de reunir um número reduzido de linhas (ou traços), é permitida aos observadores a identificação do objeto retratado.

avaliação subjetiva, realizada a partir de experimentos com usuários.

### **1.3. Estrutura da Dissertação**

O presente documento é composto por um total de cinco capítulos. Neste capítulo, contextualiza-se a pesquisa ora documentada. No **Capítulo 2**, apresenta-se um panorama das técnicas de renderização não foto-realísticas, a partir de uma revisão de trabalhos importantes e um breve histórico da área. Descrevem-se, adicionalmente, técnicas usadas na tarefa específica de geração de esboços, enquanto representações não foto-realísticas, bem como se delimita o escopo do presente trabalho.

No **Capítulo 3**, descreve-se a abordagem proposta para a geração de esboços, sendo apresentado o fluxo de processamento, desde a imagem de entrada até a geração do resultado final. O funcionamento de cada módulo da técnica proposta é ilustrado, incluindo imagens de passos intermediários. Adicionalmente, são apresentados alguns detalhes de implementação e de organização interna do *software* desenvolvido para validar a abordagem proposta.

No **Capítulo 4**, são reunidos os experimentos realizados e a discussão dos resultados obtidos. A apresentação dos experimentos contempla os testes realizados nos vários caminhos disponíveis até que fosse atingida a abordagem descrita no Capítulo 3. Adicionalmente, o Capítulo 4 contém gráficos de validação do classificador usado e análises comparativas, uma objetiva e outra subjetiva, envolvendo outras técnicas existentes.

Finalmente, no **Capítulo 5**, são apresentadas as considerações finais resultantes deste estudo, as contribuições e trabalhos futuros que podem ser derivados do tema abordado nesta dissertação.

# Capítulo 2

## Trabalhos Relacionados

Um breve histórico da área de renderização não foto-realística (NPR) é apresentado neste capítulo, no qual são ressaltados os trabalhos pioneiros e de maior relevância, reunidos após pesquisa bibliográfica. Em seguida, apresenta-se uma descrição panorâmica das técnicas usadas especificamente na tarefa de confecção de esboços para a geração de representações não foto-realísticas. A revisão de trabalhos relacionados à geração de esboços visa contextualizar o trabalho ora proposto, que tem como objetivo a geração de representações não foto-realísticas que contêm gradações de níveis de cinza e variações na espessura dos traços. Uma vez que a detecção de bordas surge como aspecto importante na geração de esboços, algumas técnicas de geração de bordas são também apresentadas neste capítulo.

NPR configura-se como uma área ampla que provê vários modos de expressar a realidade (Tresset e Fol Leymarie, 2005). Portanto, o trabalho ora proposto limita-se a estudos em uma fração desta área. As áreas que compõem, mas não restringem, a Computação Gráfica, são delineadas na Figura 2.1. Note-se que tal agrupamento visa tão somente facilitar a inserção da presente pesquisa em um contexto mais amplo.

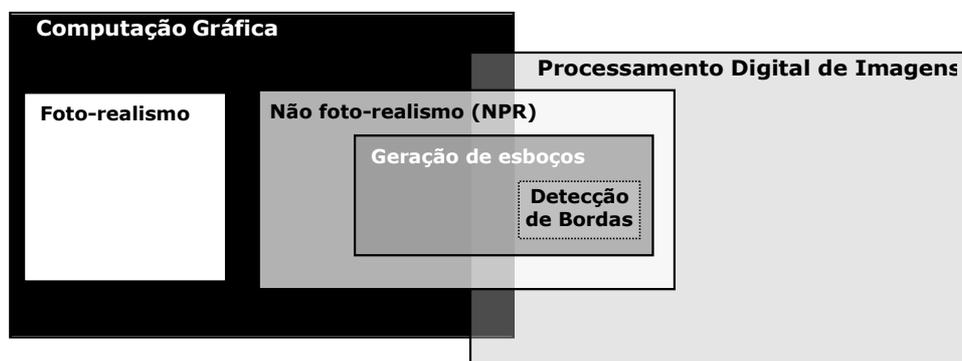


Figura 2.1 – Delimitação do escopo do trabalho: geração de esboços.

O Foto-realismo, no contexto da Computação Gráfica, consiste de um conjunto de técnicas destinadas à geração de imagens idênticas a fotografias reais, a partir de algoritmos e leis da Física envolvidas no processo fotográfico (Strothotte e Schlechtweg, 2002). Na década de 60, a Computação Gráfica despontava com o objetivo de gerar cenas que imitavam imagens geradas por uma câmera fotográfica, dando grande impulso a pesquisas na área. Hoje em dia, os trabalhos na área de Foto-realismo se concentram na adição de efeitos especiais para aumentar o realismo, e.g., a modelagem de fenômenos climáticos numa cena ou efeitos físicos de materiais específicos. Na Figura 2.2, é apresentada uma cena foto-realística gerada a partir de técnicas de Computação Gráfica.



**Figura 2.2 – Exemplo de renderização foto-realística.**

Todavia, Foley et al. (1990) observaram que, se o objetivo de uma imagem é transmitir informação, então uma imagem livre de complicações, tais como reflexos, a depender do contexto da aplicação, pode obter mais êxito para cumprir tal objetivo do que uma imagem foto-realística. Fundamentada nesta premissa, a Computação Gráfica Não Foto-realística surgiu com o propósito de gerar imagens e animações, de forma digital, que pareçam ter sido produzidas à mão. Mais precisamente, NPR visa à geração de imagens com a correta distribuição de contornos, textura e sombra em

algumas regiões, mas conferindo um aspecto artístico à renderização (Strothotte e Schlechtweg, 2002).

## 2.1. Breve Histórico

De acordo com Strothotte e Schlechtweg (2002), os autores Lansdown e Schofield (1995) foram os primeiros a estruturar os problemas a serem tratados pela NPR. Particularmente, de acordo com Strothotte e Schlechtweg (2002), Lansdown e Schofield (1995) foram os primeiros a observar que as renderizações geradas na época eram muito grosseiras e possuíam artefatos que poderiam ser atribuídos exclusivamente à imagem resultante e ao processamento realizado, mas não ao objeto retratado. Lansdown e Schofield (1995) também afirmaram que as técnicas de renderização foto-realística baseadas em modelos geométricos, nuances de iluminação, sombras, texturas e aspectos puramente físicos não são necessariamente a melhor escolha de representação, i.e., o foto-realismo é apenas um dentre vários estilos adotados para retratar a realidade.

As imagens geradas a partir de técnicas de NPR usualmente lembram artefatos gerados por arquitetos ou ilustradores de livros científicos que, ao produzi-los, objetivam comunicar mais ou menos detalhes, em situações nas quais tais artefatos comumente acompanham descrições textuais. As características-chave de imagens renderizadas por tais técnicas são a aleatoriedade, ambiguidade e arbitrariedade de alguns traços, mais do que a consideração das propriedades exibidas pelo objeto retratado (Strothotte e Schlechtweg, 2002).

Embora Strothotte e Schlechtweg (2002) comentem que o termo NPR foi utilizado pela primeira vez por Lansdown e Schofield (1995), as raízes das técnicas aparecem em artigos da década de 80, como Strassmann (1986) ou Sasada (1987). Artigos influentes foram publicados no evento SIGGRAPH<sup>5</sup> 1990 (Saito e Takahashi, 1990 e Haeberli, 1990). Em 1994, no evento SIGGRAPH, os contornos da área delinearam-se com a publicação de Winkenbach e Salesin (1994), e Salisbury et al. (1994), bem como a

---

<sup>5</sup> SIGGRAPH: The 17th Annual Conference on Computer Graphics and Interactive Techniques. (<http://www.siggraph.org/>)

publicação de Strothotte et al. (1994), no evento Eurographics<sup>6</sup> do mesmo ano.

Dentre os artigos pioneiros, um dos primeiros que se adéquam ao que é chamado hoje de NPR, cujo objetivo era a renderização não foto-realística, é o artigo de Strassmann (1986). No referido artigo, o autor apresenta a idéia de “*path-and-stroke*”, a qual consiste na implementação de estruturas (*strokes*) que simulam uma pincelada ou um risco a lápis e permitem gerar traços, pinceladas ou riscos ao longo de um caminho (*path*).

Haeberli (1990) apresentou uma abordagem para que, de posse de uma cena natural ou sintética, se possa convertê-la interativamente em uma imagem abstrata no estilo impressionista, a partir da manipulação ou seleção de parâmetros de renderização, resultando numa exploração de diversas representações de uma única fonte. Um exemplo de resultado obtido pelo autor ser visto na Figura 2.3A.

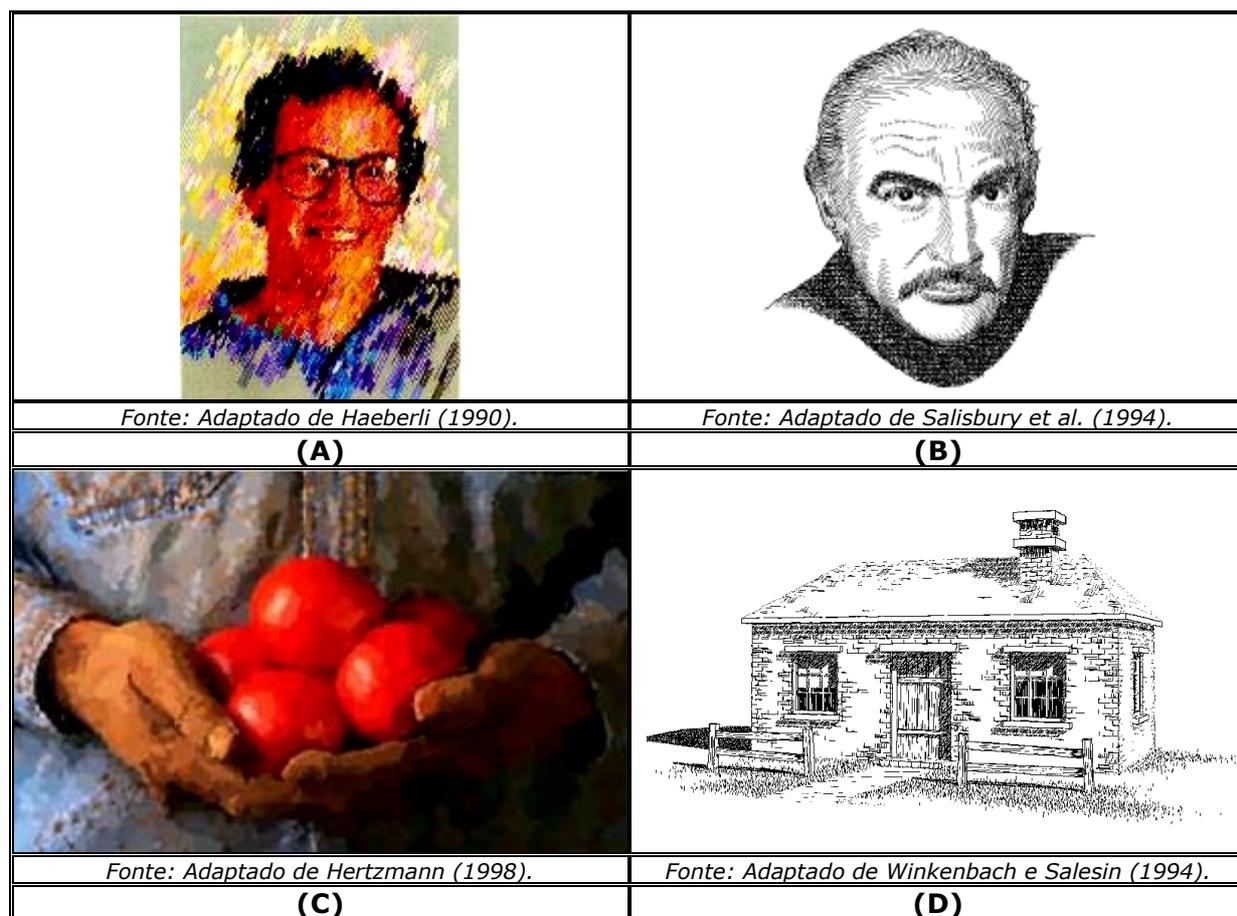
Salisbury et al. (1994) introduziram algoritmos e estruturas de dados usados na geração de imagens que imitam desenhos com nanquim, onde, em um sistema de ilustração 2D, padrões de riscos são utilizadas para fornecer textura e tonalidade às ilustrações. As ilustrações são criadas a partir de imagens guia que o usuário utiliza para aplicar texturas. Neste artigo, Salisbury et al. (1994) apresentaram o arcabouço algorítmico necessário para seleção das posições e orientações das pinceladas. Um exemplo de resultado obtido pelos autores pode ser visto na Figura 2.3B.

Vários métodos que visam à simulação de estilos artísticos foram relatados na literatura, tais como ilustrações técnicas (Gooch et. al, 1998 e Gooch et. al, 1999), ilustrações com nanquim (Salisbury et. al, 1994; Winkenbach e Salesin, 1994; Winkenbach e Salesin, 1996), ilustrações com grafite (Sousa e Buchanan, 1999; Sousa e Buchanan, 2000), pintura em aquarela (Bousseau et. al 2006; Curtis et. al, 1997), pintura impressionista (Litwinowicz, 1997; Meier, 1996), mosaicos (Di Blasi e Gallo, 2005; Dobashi et. al, 2002; Faustino e De Figueiredo, 2005), *cartoons* (Lake et. al, 2000; Wang et. al, 2004), vitrais (Mould, 2003), dentre outros. Exemplos de

---

<sup>6</sup> Eurographics: 15th Annual Conference of the European Association of Computer Graphics. (<http://www.eg.org/>)

renderizações não foto-realísticas são mostrados na Figura 2.3.

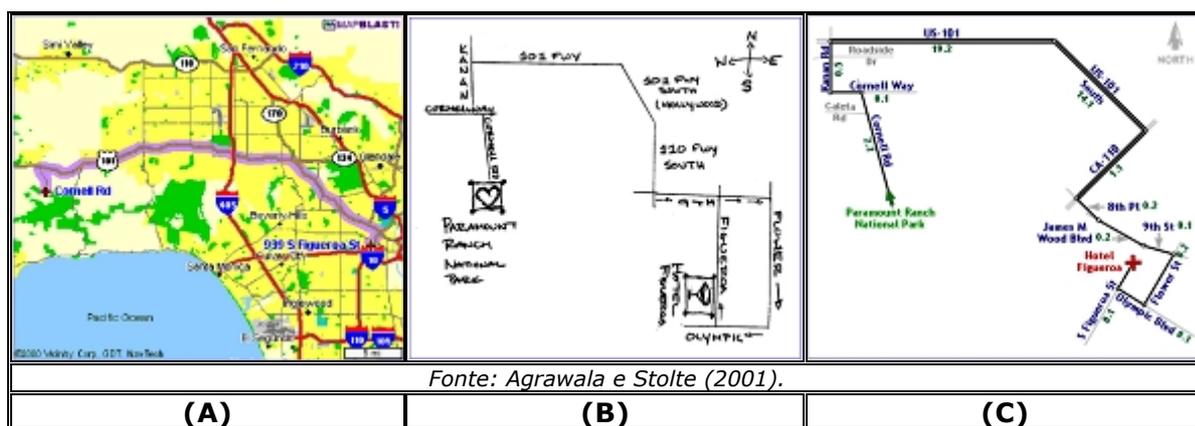


**Figura 2.3 – Exemplos de renderizações não foto-realísticas.**

Antes do termo NPR, várias outras designações foram dadas à área: (i) *renderização ilustrativa (illustrative rendering)*, por Dooley e Cohen (1990); (ii) *renderização compreensível (comprehensible rendering)*, por Saito e Takahashi (1990); (iii) *renderização por esboço (sketch rendering)*, por Strothotte et al. (1994); (iv) *ilustração em nanquim (pen-and-ink illustration)*, por Winkenbach e Salesin (1994); (v) *renderização não foto-realística (non-photorealistic rendering)*, por Lansdown e Schofield (1995); (vi) *renderização artística (artistic rendering)*, por Lansdown e Schofield (1995); (vii) *renderização por pontilhismo (stipple rendering)*, por Deussen et al. (1999); (viii) *apresentação elástica (elastic presentation)*, por Carpendale (1999); e, ainda, (ix) *renderização não realística (non-realistic rendering)*, usados pelos organizadores da conferência Eurographics<sup>7</sup> em 1999.

<sup>7</sup>Eurographics: 20<sup>th</sup> Annual Conference of the European Association of Computer Graphics. (<http://www.cg.tuwien.ac.at/conferences/VisSym99/>)

No contexto da motivação para o uso de imagens no estilo NPR, Agrawala e Stolte (2001) notaram que mapas para uma determinada rota são uma das formas mais comuns de comunicação por meio de imagens. No entanto, mesmo sendo uma tarefa trivial, em alguns casos há problemas como o excesso ou a falta de informação relevante para a tarefa a ser desempenhada. Assim, técnicas de NPR podem ser aplicadas à visualização concisa da informação, como mostrado na Figura 2.4, para melhorar a percepção, cognição e o objetivo da imagem.



**Figura 2.4 – Uso de NPR para melhorar a comunicação em mapas: (A) Mapa completo; (B) Mapa feito por uma pessoa; e (C) Mapa gerado pelo sistema proposto por Agrawala e Stolte (2001).**

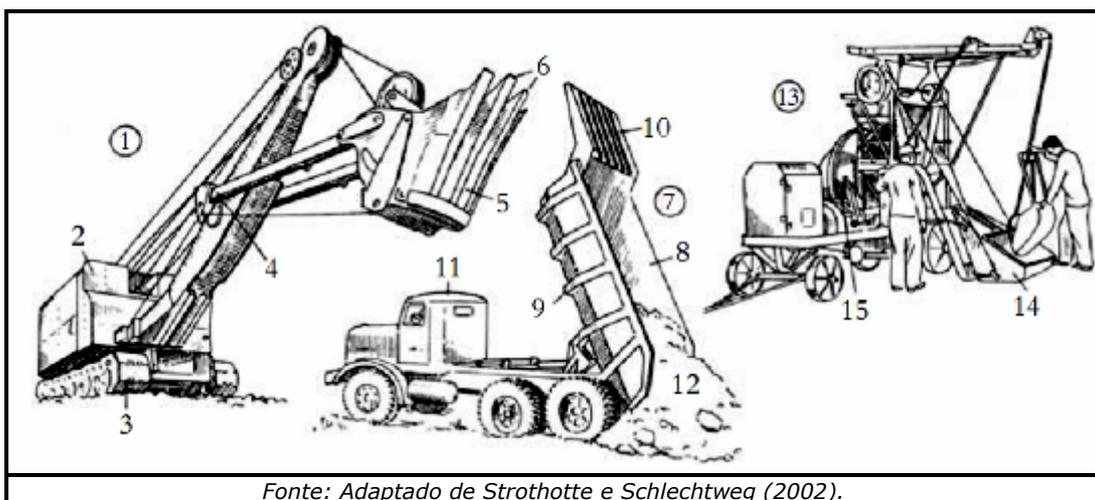
Além da visualização concisa da informação, Tateosian e Healey (2003) elencaram vários usos para NPR, atentando para o fato de que técnicas descritas e os sistemas computacionais construídos não são apenas empregados na geração de arte por não-artistas, mas que existem sistemas construídos especialmente para artistas que, por meio de novas técnicas computacionais, dispõem de maior ferramental para expressão criativa. A área de NPR provê também técnicas de visualização alternativa, aplicações na geração de quadrinhos (*comics*) e desenhos animados (*cartoons*), pinturas artísticas, desenhos técnicos, comunicação publicitária, visualização de desenhos científicos e outros usos que não priorizam o realismo.

## 2.2. Geração de Esboços

No contexto da NPR, o desenho de esboços ou contornos constitui uma das formas mais comuns e efetivas de ilustração, expressando a informação de forma concisa, e.g., em ilustrações técnicas, desenhos arquitetônicos preliminares, cenas de pré-produção de filmes. Markosian et al. (1997)

deram destaque especial à confecção de esboços, citando arquitetos que imprimem modelos computadorizados e criam a impressão de esboço sobrepondo rabiscos a tais impressões, para não criar, nos clientes, uma falsa impressão de completude do projeto.

O objetivo primário de uma representação desta natureza é delinear objetos, com um nível menor ou maior de detalhes, de modo a permitir o traçado de formas e transmitir a idéia geral do objeto retratado (Sayeed e Howard, 2006), como pode ser observado na Figura 2.5.



**Figura 2.5 – Exemplo de uma representação semi-detalhada em um diagrama técnico.**

Fekete et al. (1995), assim como Chen et al. (2001, 2002a, 2002b, 2004), focaram esforços na geração de representações estilizadas de faces usando uma abordagem com compatibilização (*matching*) de gabaritos ou *templates* de elementos-chave da face, tais como boca e nariz, para gerar desenhos de faces. A técnica proposta pelos autores faz uso de *Active Shape Model* (Cootes et al., 1995) para a busca de pontos fiduciais das faces, por meio de modelos de forma. Adicionalmente, usam a técnica dos *k*-vizinhos mais próximos (*kNN*) para cálculo de proporções entre atributos faciais que irão compor o esboço vetorial final. No entanto, tal técnica é restrita a faces de mulheres asiáticas.

DeCarlo e Santella (2002) fizeram a estilização de fotografias a partir de um mecanismo de rastreamento do movimento dos olhos do observador, obtendo informações que geram um mapa de atenção visual em uma imagem sob avaliação. A partir do referido mapa, foram inferidas

informações acerca de zonas de grande impacto, as quais foram usadas para criar a renderização estilizada. Adicionalmente, foi dado um destaque às bordas usadas na composição da imagem estilizada, pois estas, segundo os autores, formam os elementos constituintes do estilo de renderização obtido. Nesta abordagem, os autores usaram uma variante robusta do detector de bordas Canny (Meer e Georgescu, 2001), ressaltando que, apesar de algumas limitações do detector, foram geradas bordas que capturavam uma quantidade razoável de aspectos importantes da imagem. Tais bordas foram usadas para as outras etapas da técnica proposta, em especial, a inclusão de bordas na renderização adiciona expressividade ao resultado final, como pode ser observado na Figura 2.6.



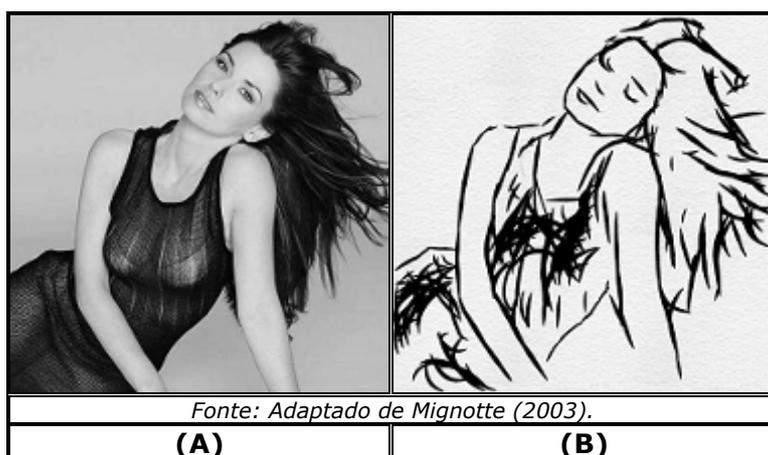
**Figura 2.6 – Importância das bordas: (A) Aparato usado para rastreamento dos olhos; (B) Imagem segmentada sem bordas; (C) Bordas detectadas; e (D) Resultado final mais expressivo.**

Após a geração do esboço, o resultado pode ser melhorado com a adição de estilos. Hertzmann (2003) elencou vários tipos de estilos de traços, classificando-os como *Stroke-Based Rendering (SBR)*, onde pontos e riscos podem ser renderizados de forma parametrizada, i.e., diferentes parâmetros produzem diferentes resultados. Vários algoritmos e estilos de *SBR* foram propostos e classificados pelo autor, e.g., mosaicos, nanquim e pontilhismo.

Tang e Wang (2003) propuseram um sistema de indexação e reconhecimento de faces que utiliza uma abordagem baseada em esboços de fotos, focado na ajuda a agentes da lei para o reconhecimento de retratos falados. O esboço gerado por tal sistema aproxima-se da foto real, não retrata os cabelos na composição e usa elementos de forma e textura da imagem original. Foi realizada uma validação do sistema por meio de experimentos de cunho subjetivo usando um processo de votação, bem

como experimentos numéricos, ambos realizados em um conjunto grande de imagens (>300).

Mignotte (2003) propôs uma técnica, baseada em inferência Bayesiana e distribuição do gradiente da imagem, para a geração de esboços, adicionando parâmetros randômicos ou perturbações controladas por parâmetros para gerar uma ampla gama de estilos, tendo nas bordas um elemento importante para a composição do resultado final, como pode ser observado na Figura 2.7.



**Figura 2.7 – Esboços gerados a partir de argumentos estatísticos: (A) Entrada; e (B) Resultado.**

Tresset e Fol Leymarie (2005) dissertaram sobre a motivação para criação de um sistema capaz de gerar esboços que um artista humano geraria em segundos. O propósito mais relevante é o de estudar o processo de percepção da beleza, assim como estudar o meio pelo qual a beleza retratada pelas mãos de um artista se revela, aprofundando-se no conhecimento do processo de criação de esboços, e, em última análise, o processo de estudo de criação de obras de arte.

No tocante ao processo de criação de um desenho, Tresset e Fol Leymarie (2005) desenvolveram um estudo sobre o fluxo adotado por um artista ao iniciar o processo de desenho, que começa na observação, passa pela identificação de saliências, desenho das linhas estruturantes e superfícies e vai até a composição da imagem final. No estudo do fluxo, há destaque para as linhas do esboço inicial, que constituem o elemento básico para a composição da arte final. Com base em tal estudo, foi desenvolvido um sistema computacional para a geração automática de esboços, focado

em faces humanas. No mesmo trabalho, detalhes relevantes foram destacados, como a aplicação do espaço de cor mais adequado a confecção de esboços, resultando em menos falsos positivos para a localização de faces e pele, bem como o processamento da imagem de entrada pela aplicação de um algoritmo de *constância cromática* (Barnard, Cardei e Funt, 2002), e.g., expansão histogrâmica ou *white patch*<sup>8</sup>. Um exemplo de imagem de esboço sobreposta em uma imagem com variação tonal é mostrado na Figura 2.8.



**Figura 2.8 – Exemplo de esboço gerado usando a técnica proposta por Tresset e Fol Leymarie.**

Kang, Lee e Chui (2007) propuseram o algoritmo FDoG (***Flow-based Difference of Gaussians***), o qual visa à geração automática de representações não foto-realísticas de desenhos compostos por linhas. Os autores ressaltaram que o desenho de linhas é a forma mais simples e mais antiga de comunicação visual. Adicionalmente, os autores afirmaram que várias técnicas de NPR usam linhas como base para a criação de outros estilos de NPR, mas que há uma escassez de pesquisa centrada fundamentalmente em desenhos compostos apenas por linhas. O algoritmo dos autores fundamenta-se na computação de um “fluxo” de características salientes da imagem, o que produz um melhoramento no efeito produzido pelo detector de bordas DoG (***Difference of Gaussians***), como mostrado na Figura 2.9.

---

<sup>8</sup> Método que assume a presença de pixels brancos brilhantes representativos da luz ambiente.



**Figura 2.9 – Filtro DoG aprimorado: (A) Entrada; e (B) FDoG**

No que concerne aos esboços feitos a partir de fotos com faces humanas, algumas aplicações se destacam: (i) a composição para a indexação de retratos falados; (ii) a estrutura-base para animações; (iii) o uso em entretenimento e publicidade; e (iv) o uso em serviço de impressão e personalização de cartões de felicitações, dentre outras (Wen et al., 2006; Rajendran e Chang, 2000).

Os trabalhos apresentados nesta seção mostraram-se os mais relevantes durante a pesquisa realizada, uma vez que focaram na geração de esboços usando técnicas de NPR. As técnicas usadas estão sumarizadas na Tabela 2.1. Aspectos relevantes da bibliografia estudada deram motivação para a proposição dos módulos da abordagem proposta, apresentada no Capítulo 3.

**Tabela 2.1 – Sumário das técnicas de geração de esboços.**

DeCarlo e Santella (2002)	Usaram atenção visual, obtida a partir de um aparato específico, para inferir informações sobre zonas onde as bordas do esboço deveriam ser melhoradas. Usam uma variante do detector de Canny como base para o estilo das bordas.
Tang e Wang (2003)	Usaram pontos fiduciais na extração de informações de forma e textura da imagem, recompondo faces usando <i>templates</i> pré-desenhados.
Mignotte (2003)	Usou abordagem estatística, com base em inferência Bayesiana, para prover estilos às linhas do esboço.
Tresset e Fol Leymarie (2005)	Fizeram um estudo do fluxo adotado pelo artista ao desenhar. Adicionalmente, usaram algoritmos de constância cromática e mudança de espaço cromático para melhorar os resultados, impressos em <i>pen-plotter</i> .
Kang et al. (2005)	Propuseram um arcabouço <i>semi-automático</i> de edição de linhas, usando curvas deformáveis ( <i>spline</i> ) para geração do esboço, construídas a partir da interação com o usuário.
Kang, Lee, Chui (2007)	Propuseram melhorias ao filtro DoG para geração de bordas, fundamentadas no de fluxo de características salientes da imagem.

No contexto do processamento digital de imagens, processos de segmentação buscam a extração de atributos de imagens por meio da

subdivisão da imagem de entrada em suas regiões constituintes (Gonzalez e Woods, 2008). Segundo Gonzalez e Woods (2008), a segmentação de imagens não triviais é uma das tarefas mais árduas na área de processamento de imagens. A maioria dos algoritmos de segmentação revisados por Gonzalez e Woods (2008) divide-se em duas categorias: descontinuidades e similaridade. A categoria de algoritmos de segmentação baseada em descontinuidades procura repartir a imagem baseado em mudanças bruscas nas intensidades dos valores de cinza. Por sua vez, a categoria de algoritmos de segmentação baseados em similaridade procura o agrupamento de regiões similares por meio de critérios pré-definidos.

De acordo com Gonzalez e Woods (2008), no contexto da segmentação baseada em descontinuidades, existem três tipos de características de interesse: pontos isolados, linhas e bordas. Adicionalmente, segundo os autores, **bordas** são caracterizadas por transições abruptas nos valores de intensidade dos pixels de borda. **Imagens de bordas** são, por sua vez, imagens formadas por conjuntos de pixels de borda conectados. **Detectores de bordas** são métodos para processamento local que visam à detecção de pixels de bordas.

Conforme ressaltado, um aspecto marcante da aplicação de algoritmos de detecção de bordas é a capacidade de extração de limites entre objetos relevantes, seja em imagens estáticas ou em cenas em movimento. Tal capacidade constituiu a "mola motora" da pesquisa desenvolvida nesta dissertação, tendo como cerne a capacidade de extração de características faciais marcantes, como olhos, boca e nariz, de forma a obter resultados em tempo computacional hábil, de forma coerente com a realidade, com traços bem localizados em relação à imagem original e melhor resposta a bordas relevantes, com o propósito de gerar representações não foto-realísticas de forma automática.

Conforme anteriormente discutida, a geração de esboços constitui uma parte importante da área conhecida com NPR e a detecção de bordas uma parte importante na geração de esboços. Para a tarefa de detecção de bordas, vários detectores foram propostos ao longo dos anos. Ziou e

Tabbone (1998) fizeram uma revisão do estado da arte de técnicas de detecção de bordas, na qual são apresentadas as técnicas clássicas de detecção de bordas, retratando propriedades, aspectos desejáveis de um detector, efeitos indesejáveis e ainda uma discussão sobre implementação dos mesmos. Dentre os vários detectores de borda disponíveis, neste trabalho ressaltam-se o detector de Canny e o DoG. Uma breve explanação destes faz-se necessária para melhor entendimento da justificativa do uso destes detectores no presente trabalho.

Kang, Lee e Chui (2007) afirmaram que o filtro para detecção de bordas de Canny é considerado o detector de bordas mais usado, e que este serve como base para vários métodos de desenho de linhas no estilo NPR. Segundo Gonzalez e Woods (2008), o filtro de Canny foi concebido com três objetivos principais: (i) as bordas detectadas deveriam ser tanto fiéis quanto possível às bordas verdadeiras; (ii) pontos de borda deveriam ser bem localizados em relação às bordas originais; e (iii) o detector não deveria indicar múltiplas respostas para presença de borda onde houvesse apenas um pixel de borda. O diferencial do trabalho de Canny é que este expressou os critérios desejáveis de forma matemática, buscando a solução ótima para cada um dos critérios. A melhor resposta do filtro de Canny traz consigo maior complexidade de implementação, maior custo computacional e maior tempo de operação do mesmo, que devem ser levados em consideração, a depender da aplicação em questão.

Por sua vez, o detector de bordas DoG é um método derivado do algoritmo de Marr-Hildreth (1980). O algoritmo de Marr-Hildreth (1980) é a convolução da imagem com o Laplaciano do Gaussiano, sendo o filtro DoG uma aproximação do algoritmo de Marr-Hildreth (1980). O DoG pode ser expresso por:

$$DoG(x,y) = \frac{1}{2\pi\sigma_1^2} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{2\pi\sigma_2^2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \quad (4.4)$$

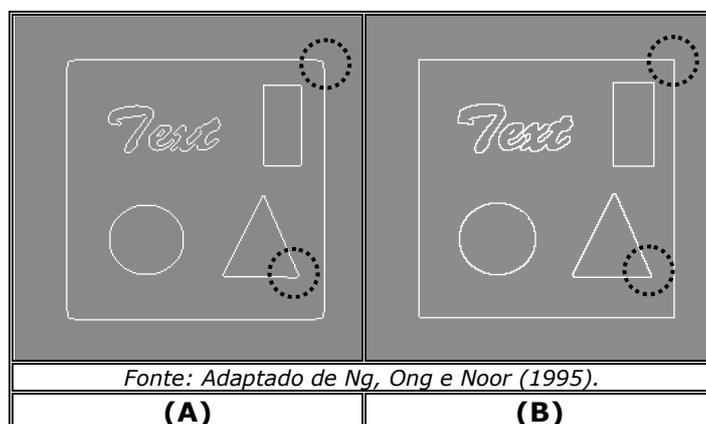
em que  $\sigma_1$  e  $\sigma_1$  são os desvios padrão das máscaras da função gaussiana usada na convolução.

Gooch et al. (2004) apresentaram um sistema voltado a ilustrações

faciais baseado no algoritmo diferença de gaussianas, usado posteriormente em conjunção com uma binarização para produção de ilustrações em preto-e-branco.

Apesar dos vários algoritmos de detecção de bordas disponíveis, desejou-se, para efeitos desta pesquisa, investigar o uso de outra abordagem para detecção de bordas, i.e., o uso da aprendizagem de máquina para detecção de bordas. A motivação para essa investigação veio da percepção da **escassez** de técnicas que tratem especificamente da geração de esboços usando aprendizagem de máquina, com avaliação centrada em faces humanas. A seguir, algumas dessas técnicas são apresentadas.

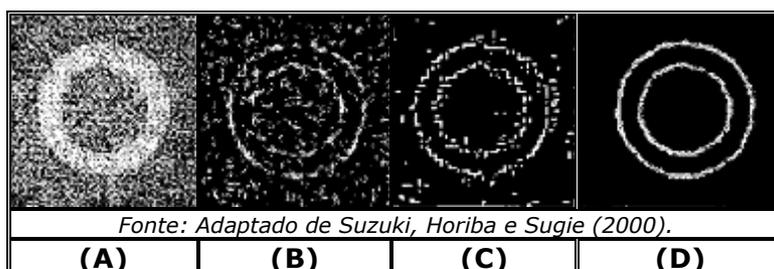
Ng, Ong e Noor (1995) destacaram problemas com detectores de borda clássicos, tais como o arredondamento de arestas (Figura 2.10), tendo proposto um detector de bordas neural, com abordagem híbrida, i.e., parcialmente supervisionada e parcialmente não-supervisionada, usando redes neurais do tipo *perceptron* multicamadas, com entrada de dados com base em máscaras de tamanho 3x3, mas apenas *cinco* (5) imagens de treinamento.



**Figura 2.10 – Arredondamento de arestas na detecção de bordas: (A) Canny (B) Neural.**

Lee et al. (2000) melhoraram o desempenho computacional de algoritmos clássicos, como Sobel e Canny, combinando-os com aprendizagem de máquina, via algoritmos genéticos. O algoritmo genético é treinado com exemplos de uma imagem real e uma imagem de borda, em que a função-objetivo é usada para aperfeiçoar os pesos de parâmetros de algoritmos clássicos.

Suzuki, Horiba e Sugie (2000) propuseram um filtro de detecção de bordas usando redes neurais supervisionadas para imagens com muito ruído. Experimentos objetivos e comparação visual mostraram que o detector neural gerou linhas bem conectadas e houve boa supressão de ruído, como mostrado na Figura 2.11. Os autores relataram que o treinamento foi realizado com imagens com e sem ruído. No entanto, detalhes quanto à arquitetura da rede usada não foram expostos.



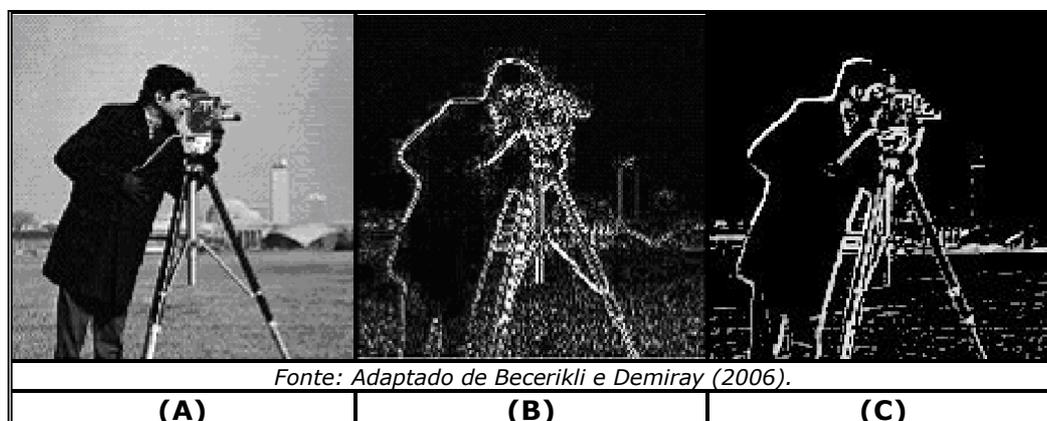
**Figura 2.11 – Bordas em imagem com ruído: (A) Entrada; (B) Sobel; (C) Canny; e (D) Neural.**

Castro e Silva (2004) fizeram um estudo comparativo entre a detecção de bordas usando o algoritmo de Canny e uma abordagem neural, visando à implementação e análise de vários modelos de redes neurais artificiais que extraíssem informação precisa de um ambiente ruidoso. Testes objetivos e comparação visual indicam que o resultado da detecção de bordas usando redes neurais se aproxima de resultados usando o operador clássico de Canny. Destaca-se que, para o processo de treinamento, foram usadas máscaras ou janelas deslizantes de dimensões 3x3. Há também uma estimativa de variância de tons da janela que foi submetida a julgamento da rede neural.

Rajab, Woolfson e Morgan (2004) propuseram uma abordagem para segmentação de regiões lesionadas de pele humana. Tal método se baseia na detecção de bordas usando redes neurais para reconhecimento de padrões de borda que, em última instância, podem indicar câncer de pele em humanos. Os padrões usados no treinamento são constituídos exclusivamente de máscaras sintéticas de padrões 3x3.

Becerikli e Demiray (2006) mostraram uma forma de detecção de bordas usando redes neurais artificiais. Em tal abordagem, um detector de bordas Laplaciano foi usado para treinamento supervisionado, onde o

conjunto de treinamento continha imagens de borda e imagens corrompidas com ruído. Os autores relatam que qualquer método clássico de detecção de bordas pode ser adaptado para servir como base para o treinamento do classificador. Na Figura 2.12, é ilustrado o resultado do estudo.



**Figura 2.12 – Bordas usando abordagem neural: (A) Entrada; (B) Laplaciano; e (C) Neural.**

Vários outros autores (e.g. Toivanen et. al, 2003; Chang, 2004; Gupta e Sukhendu, 2006, Wang et. al, 2007, Pian et. al, 2007) usaram aprendizagem de máquina para a detecção de bordas. Como ressaltado anteriormente, percebeu-se a **escassez** de técnicas especificamente destinadas à geração de representações não foto-realísticas a partir de aprendizagem de máquina.

Ressalta-se que, além dos aspectos levantados pela revisão de artigos de geração de esboços usando técnicas de NPR, aspectos encontrados nos trabalhos que usam aprendizagem de máquina para detecção de bordas também surgiram como inspiradores para a abordagem proposta nesta dissertação. Alguns exemplos desses aspectos são: (i) o uso de padrões de treinamento com ruído; (ii) o uso de padrões de treinamento com tamanho fixo (3x3 ou 5x5); (iii) a estimativa de variância/contraste de uma janela de classificação usando redes neurais (borda ou não borda); (iv) o uso de padrões sintéticos para treinamento; e (v) o uso de paradigmas de redes neurais diversos (*perceptron* multicamadas, mapas auto-organizáveis, redes de Hopfield), dentre outros.

## **2.3. Considerações Finais do Capítulo**

Diante do exposto, constata-se que a detecção de bordas e a geração de

esboços são etapas importantes no contexto da NPR. Constata-se, igualmente, que a dificuldade na geração de um *esboço robusto* reside no destaque de características-chave e sua reprodução a partir de poucos elementos estruturantes.

Todavia, registra-se uma carência de abordagens que utilizem aprendizagem de máquina voltada para a geração de esboços. Tal carência estende-se à aplicação de aprendizagem de máquina ao aprimoramento ou redução de artefatos indesejáveis nos resultados finais dos esboços gerados.

O trabalho ora desenvolvido focaliza-se em uma abordagem automática para a geração de esboços não foto-realísticos bidimensionais a partir de fotografias digitais, adotando técnicas de aprendizagem de máquina, em especial o uso de redes neurais supervisionadas. Adicionalmente, foca-se a validação da técnica proposta considerando-se um estudo de caso realizado em imagens frontais de faces humanas.

No próximo capítulo, é apresentada a abordagem proposta para a geração de esboços, imagens dos passos intermediários, organização do sistema implementado para validar a abordagem e detalhes de implementação.

# Capítulo 3

## Abordagem Proposta

Neste capítulo, é apresentada a abordagem proposta para a geração de representações não foto-realísticas a partir de imagens digitais. A abordagem proposta foi inspirada em características de abordagens apresentadas no Capítulo 2. Como exemplos, podem-se citar a abordagem adotada por Tresset e Fol Leymarie (2005), no tocante à mudança do espaço cromático para a melhoria do esboço gerado, assim como a abordagem proposta por Rajab, Woolfson e Morgan (2004), no tocante ao uso de padrões de treinamento e exemplos de imagens ruidosas para a detecção de bordas com o uso de redes neurais. Os módulos que compõem a abordagem são apresentados na Figura 3.1.

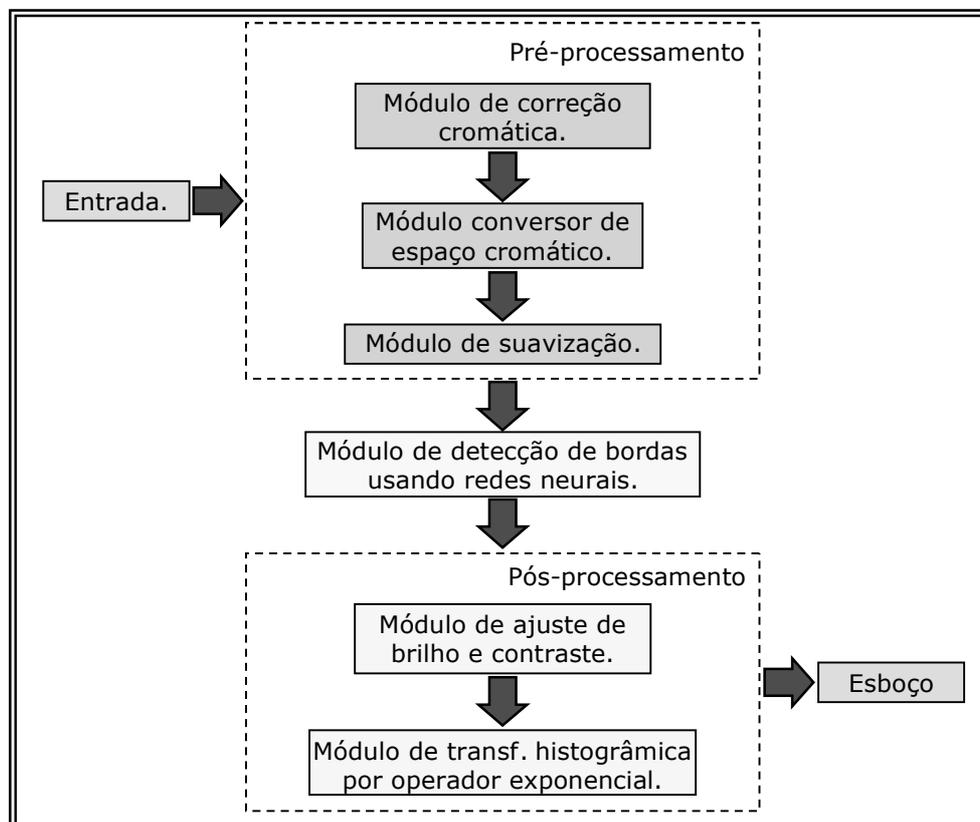


Figura 3.1 – Macro-arquitetura da técnica proposta.

A abordagem proposta é sucintamente descrita como segue:

- a. A imagem de entrada é submetida a três etapas de *pré-processamento*, a saber: (i) *correção cromática*; (ii) *conversão de espaço cromático*; e (iii) *suavização*.
- b. Após a etapa de pré-processamento, a imagem resultante é submetida a uma *filtragem espacial* usando uma máscara (*kernel*) de tamanho 5x5. Tal filtragem é realizada por um classificador neural tipo *Perceptron* Multicamadas, treinado para a detecção de bordas. O resultado é a produção de um mapa de bordas.
- c. Após a filtragem, o mapa de bordas é submetido a duas etapas de *pós-processamento*: (i) *ajuste de brilho e contraste*; e (ii) *transformação histogrâmica* por operador exponencial. A saída desta etapa é a representação não foto-realística desejada.

Com inspiração em artigos relevantes na área de geração de representações não foto-realísticas, algumas técnicas da área de Processamento Digital de Imagens (PDI) foram estudadas e implementadas, com o intuito de serem usadas nos módulos de pré-processamento e pós-processamento da abordagem proposta. Cada técnica estudada é brevemente apresentada no Apêndice A. A apresentação das técnicas contém: (i) o fundamento e princípio de funcionamento da técnica; (ii) o objetivo esperado com a aplicação da operação; e (iii) pares de imagens de entrada e saída do processo.

Dentre os vários algoritmos disponíveis a serem usados na etapa de pré-processamento ou de pós-processamento, apenas alguns foram escolhidos para serem avaliados. No entanto, o propósito do presente trabalho é a definição de uma abordagem para a geração de representações não foto-realísticas, usando, dentre um conjunto finito de possibilidades, os algoritmos que atingiram melhores resultados. Portanto, algumas técnicas de PDI estudadas podem ser substituídas por técnicas similares, e.g., um tipo de suavização diferente da suavização gaussiana ou da suavização pelo filtro da mediana. O cerne da abordagem proposta está, portanto, na aprendizagem de máquina usada na detecção de bordas para a geração de

representações não foto-realísticas, bem como na proposição do fluxo (*pipeline*) de etapas a serem seguidas para geração de esboços.

Os resultados apresentados neste capítulo são resultantes de análises de técnicas em cujos experimentos foram verificados os melhores resultados, considerando-se cada uma das etapas da abordagem proposta. Tais experimentos são apresentados e discutidos no Capítulo 4. A seguir, cada etapa do fluxo de execução mostrado na Figura 3.1 é detalhada e um par de imagens de entrada e saída, após a aplicação de cada módulo, é apresentado. A imagem usada para exemplificar o método proposto é apresentada na Figura 3.2A. Outros resultados gerados usando a abordagem proposta são apresentados no Apêndice B.

### **3.1. Pré-Processamento**

A etapa de *pré-processamento* engloba três sub-etapas, a saber: (i) *correção cromática*; (ii) *conversão de espaço cromático*; e (iii) *suavização*, conforme já explicitado anteriormente (vide Figura 3.1).

#### **3.1.1. Correção Cromática**

Variações/alterações na iluminação de uma cena podem implicar o surgimento de problemas no processo de imageamento da cena, tais como falsas bordas geradas por sombras. *Bordas de iluminação* (*illumination edges*) são geradas quando há diferentes fontes de iluminação incidindo sobre uma mesma superfície, produzindo reflexos de cores diferentes. Por sua vez, *bordas de reflectância* (*reflectance edges*) surgem quando há mudanças na reflectância de uma superfície. Bordas de iluminação e bordas de reflectância acarretam uma resposta indesejada do detector de bordas. Gilchrist et al. (1999) notaram que seres humanos podem distinguir facilmente os dois tipos de borda, enquanto vários algoritmos de detecção de bordas não são capazes de diferenciá-las.

Tal fato inspira o uso de algoritmos de compensação de iluminação para melhorar a qualidade das bordas geradas, característica também ressaltada por Tresset e Fol Leymarie (2005), o que inspirou o uso de tais técnicas no presente trabalho.

Para a operação de correção cromática, foram realizados experimentos com três algoritmos, dois dos quais visando à constância cromática (*Retinex* e *GrayWorld*) e o terceiro correspondendo à expansão linear do histograma da imagem (ou expansão de contraste). Detalhes de cada um destes algoritmos podem ser encontrados no Apêndice A. Na abordagem proposta, é aplicada apenas uma operação deste tipo. A operação que mostrou melhores resultados foi a aplicação do algoritmo *GrayWorld*, como pode ser verificado nos experimentos contidos no Capítulo 4.

Conforme indicado no Apêndice A, o método *GrayWorld* se mostra limitado quando a cena possui mais de um iluminante, sendo também mais adequado quando a cena possui grande variação de cores. Na Figura 3.2, é mostrada a imagem de teste após a aplicação do algoritmo.



Figura 3.2 – Etapa 1 – Correção cromática: (A) Imagem de entrada; e (B) *GrayWorld*.

### 3.1.2. Conversão de Espaço Cromático

Após a aplicação da operação de correção cromática, mencionada na seção anterior, a imagem de entrada é submetida a uma conversão de espaço cromático. Nesta pesquisa, dois espaços cromáticos foram estudados e

avaliados: (i) *HSV*; e (ii) *YCbCr*. Para a geração do resultado usando a abordagem proposta, apenas um canal é usado: canal V no modelo *HSV* e canal Y no modelo *YCbCr*. O canal cromático Y do espaço cromático *YCbCr* foi o que mostrou melhores resultados, conforme apresentado no Capítulo 5. Detalhes de cada espaço cromático estudado podem ser vistos no Apêndice A.

Na Figura 3.3, é mostrada a decomposição da imagem colorida nos canais que compõem o modelo cromático *YCbCr*. Para efeitos de exibição, a escala dos tons de cinza das imagens foi normalizada para o intervalo [0; 255]. O resultado da segunda etapa da abordagem proposta é ilustrado nas Figura 3.3A (após aplicação do *GrayWorld*) e B (canal Y).

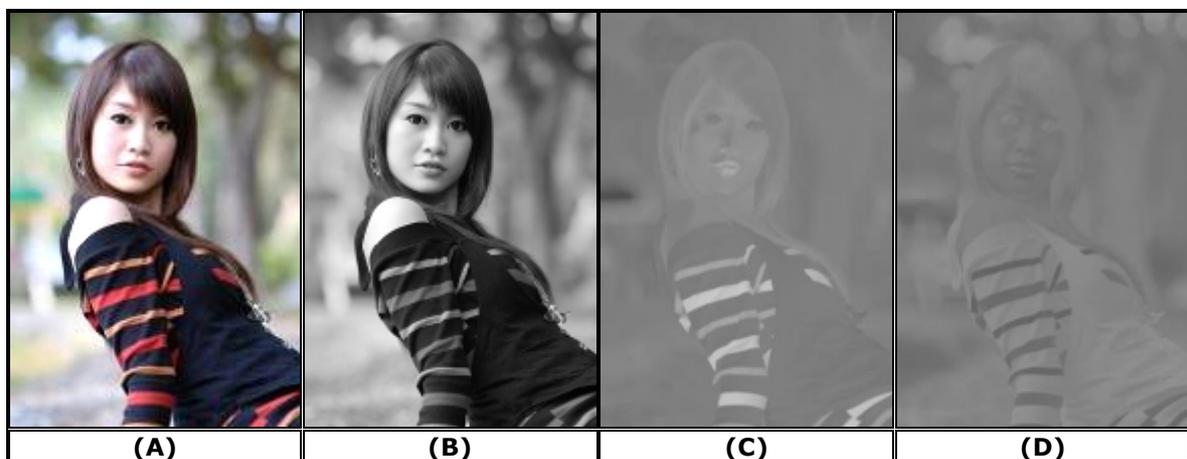


Figura 3.3 – Etapa 2 – Conversão para o espaço *YCbCr*: (A) Original; (B) Canal Y; (C) Canal Cb; e (D) Canal Cr

### 3.1.3. Suavização

A utilização de um filtro de suavização nesta pesquisa teve como propósito a redução de detalhes da imagem de entrada, bem como a eliminação de ruídos. A redução de detalhes faz parte da abordagem de abstração da imagem para a geração da representação não foto-realística desejada.

Duas técnicas de suavização foram pesquisadas e testadas: (i) a *suavização gaussiana*; e (ii) a *suavização pelo filtro da mediana*. Detalhes de cada técnica de suavização podem ser encontrados no Apêndice A. Apesar do filtro da mediana apresentar a característica de preservação das bordas em função do tamanho da máscara usada, a suavização que apresentou melhores resultados foi a suavização gaussiana, com  $\sigma=2$ . A justificativa para a escolha

deste valor na versão final da abordagem proposta é apresentada no Capítulo 4. Na Figura 3.4, é ilustrado o resultado da terceira etapa da abordagem proposta.



Figura 3.4 – Etapa 3 – Suavização: (A) Canal Y; e (B) Imagem Suavizada.

## 3.2. Detecção de Bordas

Máquinas de aprendizagem são modelos computacionais que aplicam o princípio da aproximação de uma função a partir de exemplos, o que significa que elas podem “aprender” uma função por meio da observação de exemplos de tal função. Para que uma máquina de aprendizagem seja capaz de solucionar um problema, é necessário que este problema seja passível de representação em forma de uma função para a qual um conjunto de entradas e saídas é conhecido (Nissen, 2005).

Neste contexto, um exemplo de treinamento é um par de vetores que contém uma entrada definida e uma saída desejada para a função a ser aproximada. No processo de treinamento, os exemplos de entrada são apresentados e os pesos associados às conexões vão sendo ajustados, de

forma a produzir na saída da rede o resultado desejado para cada exemplo (Haykin, 1998).

Após a etapa de pré-processamento, descrita na Seção 3.1, a etapa de detecção de bordas é realizada usando uma máquina de aprendizagem do tipo rede neural, treinada com exemplos de borda e não borda. No contexto deste trabalho, a detecção de bordas usando redes neurais apresenta como resultado um valor que contém a estimativa de contraste para uma região da imagem, sendo tal estimativa uma indicação da evidência de borda. As entradas para o treinamento foram compostas de exemplos de borda e não borda em uma vizinhança de tamanho fixo (matriz 5x5) e a saída desejada é o valor de contraste da região.

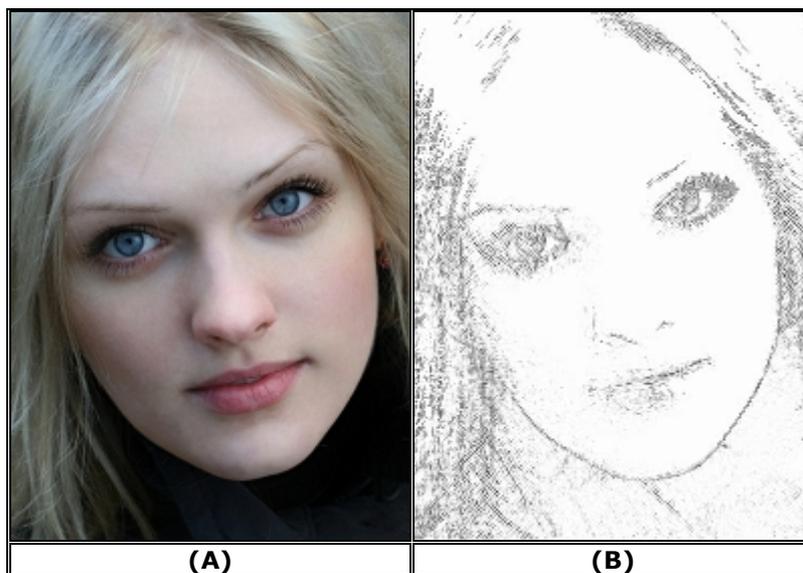
Ao solucionar problemas de reconhecimento de padrões em imagens é comum a representação dos valores de entrada e valores de saída por meio de vetores numéricos. Os vetores de entrada são formados pelos valores dos pixels que compõem a imagem e os vetores de saída são formados pelos valores desejados. Para a formação dos vetores de entrada e também no processo de utilização da rede neural é necessária uma normalização, de modo que a intensidade de um pixel varie no intervalo [-1; 1]. Esta normalização é expressa, de acordo com (Milani e Cazella, 2005), como:

$$aNew = MinMax(a) = \frac{a - \min\_a}{\max\_a - \min\_a} \times (newMax - newMin) + newMin \quad (3.1)$$

sendo  $minA$  e  $maxA$  os valores mínimos e máximos de um atributo. A normalização *MinMax* procura mapear um valor  $a$  para um valor  $aNew$  no intervalo  $[newMin; newMax]$ . Conforme Han e Kamber (2001), esta normalização é essencial para manter a consistência de escalas entre todos os valores. Os valores usados foram:  $\min\_a=0$ ,  $\max\_a=255$ ,  $newMax=1$ ,  $newMin=-1$ .

O módulo de detecção de bordas utiliza uma rede neural do tipo *Perceptron* Multicamadas (*Multi Layer Perceptron - MLP*), treinada com o algoritmo *backpropagation*. Tal modelo de rede neural foi escolhido após

testes realizados com outros dois modelos não supervisionados: redes recorrentes de Hopfield (1982) e mapas auto-organizáveis de Kohonen (**Self Organizing Maps – SOM**) (Kohonen, 1988). Os testes realizados com mapas de Kohonen não resultaram em respostas satisfatórias à classe de estímulos ou quanto ao agrupamento e classificação de padrões de borda o de padrões de não borda. Por sua vez, os testes com redes recorrentes de Hopfield geraram resultados com muitos segmentos de borda não contíguos, como pode ser observado na Figura 3.5.



**Figura 3.5 – Resultado com rede de Hopfield: (A) Original; e (B) Bordas.**

Adicionalmente, as redes neurais MLP são o modelo de aprendizado supervisionado mais proeminente na tarefa de reconhecimento de padrões (Alsmadi, Omar e Noah, 2009). O processo de treinamento foi realizado utilizando o software MATLAB, enquanto o processo de utilização da rede neural foi desenvolvido na linguagem C++.

Ao longo desta pesquisa, *dois* métodos de aquisição de conhecimento para treinamento da rede neural foram estudados: (i) o *uso de conjunto de padrões de borda* com base em *ground-truth*, obtidos pela anotação de imagens de borda feita por um humano; e (ii) o *uso de conjunto de padrões de borda*, criados sinteticamente.

A rede neural utilizada na detecção de bordas apresenta uma arquitetura em três camadas. Para reduzir a magnitude dos vetores de entrada e selecionar características discriminativas, foi usada a Análise de

Componentes Principais (*Principal Component Analysis – PCA*), aplicados individualmente a cada conjunto de bordas e não-bordas. Para o uso do critério de parada do treinamento da rede neural por validação, o conjunto total de padrões de borda e não borda foi dividido em três subconjuntos: treinamento, teste e validação. A aplicação da técnica PCA explicitada nesta seção é aplicada, separadamente, a cada um dos subconjuntos. Para possibilitar a aplicação da técnica PCA, cada padrão de treinamento, seja borda ou não-borda, deve ser linearizado em um vetor  $\mathbf{X}_T$ :

$$\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} \\ x_{5,1} & x_{5,2} & x_{5,3} & x_{5,4} & x_{5,5} \end{bmatrix}_{5 \times 5} \rightarrow [x_{1,1} \ x_{1,2} \ \dots \ x_{1,5} \ x_{2,1} \ \dots \ x_{5,4} \ x_{5,5}]_{1 \times 25} = \mathbf{X}_T \quad (3.2)$$

em que  $x_{i,j}$  é o valor cinza do pixel na posição  $i,j$  do padrão de treinamento de dimensão  $5 \times 5$ . Após a linearização, há a criação da matriz  $\mathbf{X}^{classe}$ :

$$\mathbf{X}^{classe} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_T \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,25} \\ x_{2,1} & x_{2,2} & \dots & x_{2,25} \\ \vdots & \vdots & & \vdots \\ x_{T,1} & x_{T,2} & \dots & x_{T,25} \end{bmatrix}_{T \times 25} \quad (3.3)$$

em que *classe* define a categoria do padrão (borda ou não borda),  $T$  define o tamanho do conjunto de exemplos de padrões de bordas ou não bordas originais (sem dimensionalidade reduzida) e cada vetor  $\mathbf{X}_T$  foi obtido após a linearização do padrão de treinamento.

No contexto da PCA, a construção da matriz de redução de dimensionalidade necessita da determinação de um vetor médio  $\bar{\mathbf{X}}$  por coluna, para cada conjunto de vetores obtidos no passo anterior:

$$\bar{\mathbf{X}} = \frac{1}{N} \left[ \sum_{i=1}^T x_{i,1}, \sum_{i=1}^T x_{i,2}, \dots, \sum_{i=1}^T x_{i,25} \right]_{1 \times 25} \quad (3.4)$$

em que  $N$  é o número total de padrões. Pelo fato de haver duas abordagens de aquisição de conhecimento para treinamento da rede, o valor  $N$  da Equação 3.4 é variável (mostrado na Tabela 3.1). A determinação do vetor médio também foi realizada para os padrões de não-borda.

**Tabela 3.1 – Número de padrões usados na determinação do vetor médio para PCA.**

Conjunto	Número de padrões ( $N$ )	
	Treinamento baseado em <i>ground-truth</i>	Treinamento baseado em bordas sintéticas
Treinamento	11942	766
Teste	2984	191
Validação	4976	319

Após a determinação do vetor médio  $\bar{\mathbf{X}}$ , o conjunto de vetores é ajustado a partir da subtração de cada vetor de treinamento (linhas da matriz  $\mathbf{X}^{\text{classe}}$ ) pelo vetor médio  $\bar{\mathbf{X}}$ , gerando um novo conjunto de vetores  $\mathbf{A}$ , ajustando os valores para uma média igual a 0 (zero):

$$\mathbf{A} = \mathbf{X}_i - \bar{\mathbf{X}} \quad (3.5)$$

Em seguida a tal ajuste, uma matriz quadrada de covariância  $\mathbf{C}$  de tamanho  $wh \times wh$  é criada a partir da Equação 3.6.

$$\mathbf{C} = \frac{1}{N-1} \mathbf{A} \mathbf{A}^T \quad (3.6)$$

na qual  $w=5$  e  $h=5$ , pela natureza do padrão de treinamento (padrão de dimensão 5x5). O conjunto de autovalores e autovetores normalizados da matriz de covariância  $\mathbf{C}$  são calculados:

$$\mathbf{C} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T = \begin{bmatrix} e_1 & \dots & e_D \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_D \end{bmatrix} \begin{bmatrix} e_1^T \\ \vdots \\ e_D^T \end{bmatrix} \quad (3.7)$$

em que  $e$  representa os autovetores e  $\lambda$  representa os autovalores associados. Como resultado do cálculo anterior, os autovalores encontram-se ordenados ascendentemente por magnitude.

Após a criação da matriz  $\mathbf{C}$ , há o descarte do autovetor associado ao maior autovalor, pois este não contém características discriminativas, visto que tal autovetor aproxima-se do vetor médio. Em seguida, há a redução da dimensionalidade dos vetores de borda e não-borda, a partir da escolha de um pivô de autovalor na Equação 3.7: são descartados os autovetores cujos índices são maiores do que o pivô. Para as duas abordagens de aquisição de conhecimento estudadas (com base em *ground-truth* ou geradas sinteticamente), o pivô de descarte para a redução da dimensionalidade foi obtido a partir de um fator de retenção

de 99% da variabilidade dos dados, definido após experimentos.

O processo descrito anteriormente é seguido da construção de uma matriz híbrida  $\mathbf{H}$ . Tal matriz é criada a partir da concatenação das matrizes de autovetores resultantes para os conjuntos de borda e não borda, na forma:

$$\mathbf{H} = [\mathbf{U}_{(25 \times \beta_{borda})}^{borda} \quad \mathbf{U}_{(25 \times \beta_{-borda})}^{-borda}]_{25 \times (\beta_{borda} + \beta_{-borda})} \quad (3.8)$$

em que  $\beta_{borda} + \beta_{-borda}$  define o número de colunas da matriz híbrida  $\mathbf{H}$ , que varia acordo com a abordagem para aquisição do conhecimento.

Para o treinamento com o uso de conjunto de padrões de borda com base no *ground-truth*, a matriz  $\mathbf{H}$  possui tamanho  $25 \times 19$  ( $\beta_{borda} = 10$  e  $\beta_{-borda} = 9$ ). Similarmente, a matriz  $\mathbf{H}$  possui tamanho  $25 \times 11$  para treinamento com o uso de conjunto de padrões de borda criados sinteticamente ( $\beta_{borda} = 7$  e  $\beta_{-borda} = 4$ ). O valor 25 corresponde ao tamanho dos padrões de borda usados ( $5 \times 5$ ) e o valor 19 (ou 11) corresponde à quantidade *total* de autovetores selecionados (autovetores de borda *somados* a autovetores de não-borda).

Finalmente, a redução da dimensionalidade e extração de características usando a técnica PCA ocorre com a projeção da matriz híbrida  $\mathbf{H}$  com relação ao vetor de entrada  $\mathbf{X}$ , conforme:

$$projecção_{(1 \times \beta_{total})} = \mathbf{X}_{(1 \times 25)} \times \mathbf{H}_{(25 \times \beta_{total})} \quad (3.9)$$

em que  $\beta_{total}$  é o número de colunas da matriz  $\mathbf{H}$ , igual a 19 ou 11, a depender da abordagem de aquisição de conhecimento considerada.

Portanto, a quantidade de neurônios na camada de entrada corresponde ao tamanho da saída da etapa de extração de componentes principais, que depende, por sua vez, da técnica de aquisição de conhecimento usada.

Adicionalmente, a quantidade de neurônios na camada escondida também depende da abordagem de extração de padrões de conhecimento. Tal quantidade de neurônios foi obtida experimentalmente. A rede neural

utilizada possui dois neurônios na camada de saída (um representando a classe borda e a outra representando a classe não-borda). A saída desejada representa a estimativa de contraste do conjunto de entrada que, segundo (Gomes, 2002), é dada por:

$$c = \frac{|L_{max} - L_{min}|}{L_{max} + L_{min}} \quad (3.10)$$

na qual  $L_{max}$  e  $L_{min}$  são as intensidades máxima e mínima associadas a um recorte de imagem.

A arquitetura com duas saídas foi obtida experimentalmente, já que a rede neural com apenas uma saída não apresentou resultados satisfatórios. Um delineamento da arquitetura da rede neural usada nesta pesquisa é mostrado na Figura 3.6.

A estimativa de contraste é calculada apenas para os exemplos de borda, não sendo calculada para os exemplos de não-borda. Observando o conjunto de padrões de borda, a saída da rede para um padrão de borda apresenta-se no intervalo  $[0,2; 1]$  para o neurônio 1 e valor 0 para neurônio 2. Tal intervalo inspirou-se no resultado obtido por Gomes (2002), após experimentos que levaram em conta a resposta visual aos níveis de cinza de uma imagem. Por sua vez, a saída desejada para um padrão de não-borda é 0 para o neurônio 1 e 1 para o neurônio 2.

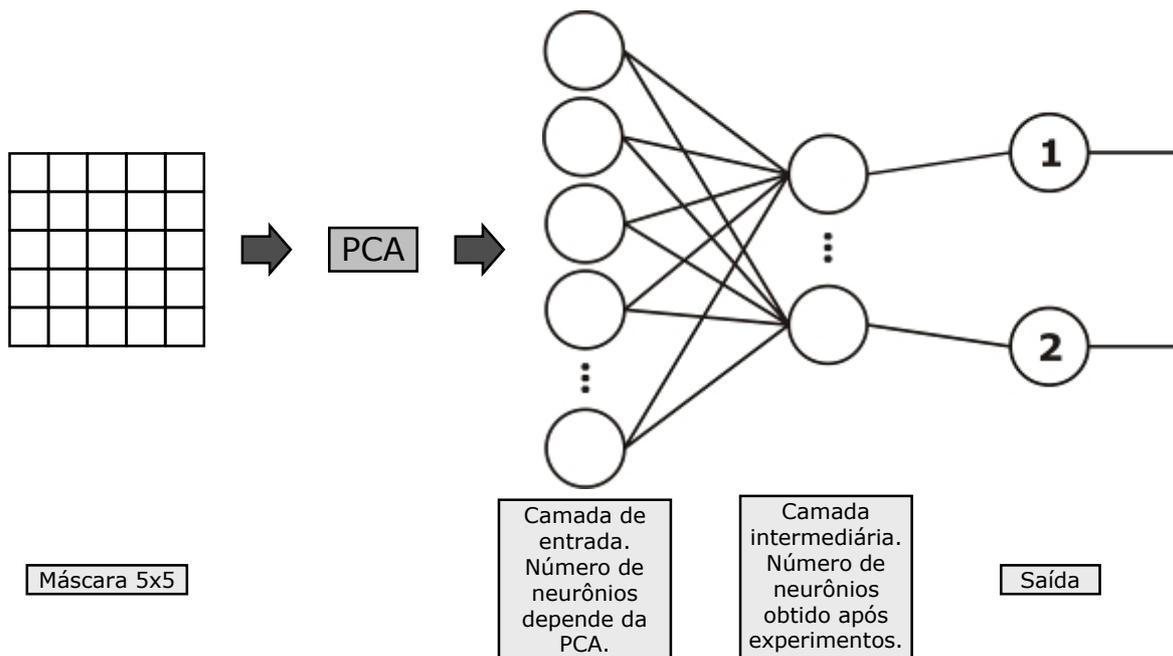


Figura 3.6 – Topologia da Rede Neural.

Vários experimentos foram realizados com os dois métodos de obtenção de conhecimento. Nas próximas seções, serão apresentados detalhes do processo de treinamento das duas abordagens estudadas, bem como os resultados obtidos. A análise de desempenho de cada um dos classificadores treinados foi realizada usando curvas ROC (**R**eciever **O**perating **C**haracteristics), sendo tal análise apresentada no Capítulo 4, no qual se justifica o uso de uma abordagem em detrimento da outra.

### 3.2.1. Treinamento Usando *Ground-truth*

Para este método de treinamento, os padrões de borda e não-borda usados no treinamento foram obtidos a partir de padrões retirados com base em uma anotação de imagens feitas por um ser humano. A motivação para tal treinamento é a aquisição de conhecimento pela imitação do comportamento humano, ao desenhar uma representação não foto-realística de uma imagem de face.

Ao todo, 14 imagens de face foram utilizadas no processo de aquisição do conhecimento. Na Figura 3.7, é mostrado um exemplo de imagem e o respectivo *ground-truth*. As demais imagens são apresentadas no Apêndice C. Os padrões de borda e não borda foram obtidos pela seleção aleatória de recortes de tamanho 5x5. Os padrões a serem usados no treinamento foram selecionados com base no *ground-truth*: se o pixel central do recorte do recorte for branco, então o padrão será classificado como borda. Se o pixel central do recorte for preto, o padrão será classificado como não-borda.

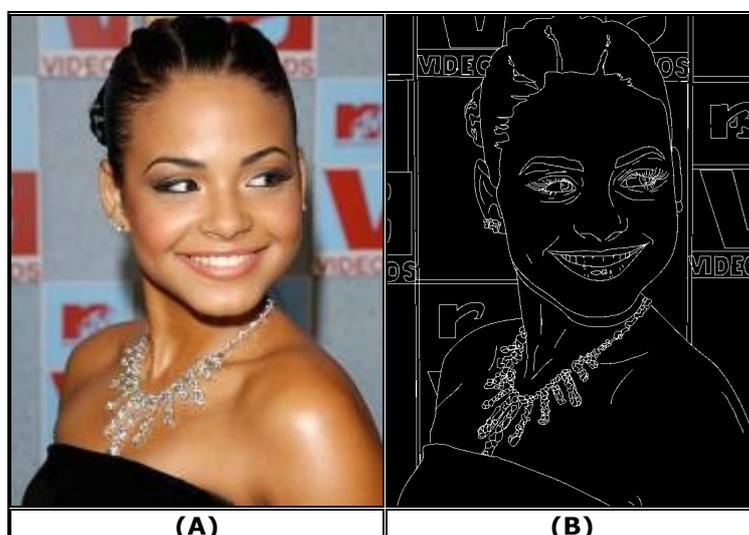
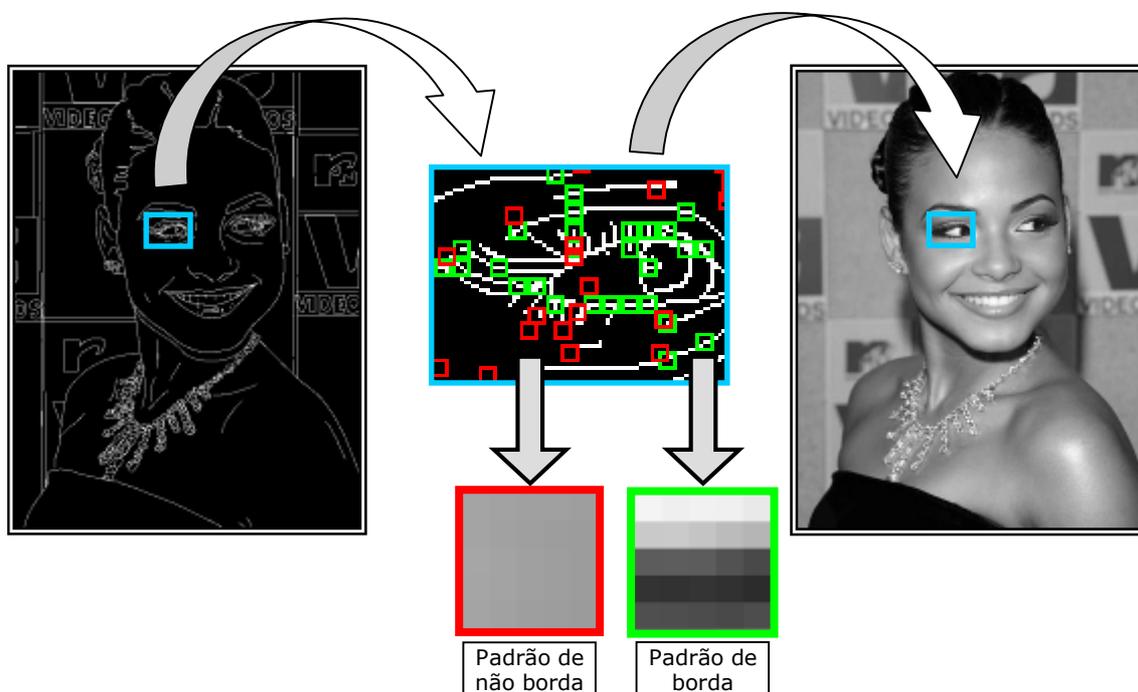


Figura 3.7 – Imagem de treinamento: (A) Original; e (B) *Ground-truth*.

As entradas da rede neural, na etapa de treinamento, são compostas por padrões de borda e não borda, retirados da imagem original em escala de cinza. A saída desejada é a estimativa a partir do contraste do recorte. Os padrões de borda e não-borda não se repetem dentro do escopo de uma mesma imagem. Na Figura 3.7, é ilustrado o processo de aquisição dos padrões.

Para aumentar a variabilidade do conjunto de padrões de borda, outros padrões de treinamento foram obtidos, sendo então adicionados aos padrões da imagem original em escala de cinza. Tais padrões foram retirados da imagem degradada com ruído gaussiano e também foram considerados padrões obtidos da imagem suavizada com suavização gaussiana. O processo de aquisição dos padrões para o aumento de variabilidade é semelhante àquele ilustrado na Figura 3.7, havendo a mudança da imagem em tons de cinza, fonte dos recortes, i.e., ao invés de usar a imagem original em escala de cinza, usa-se a imagem suavizada em escala de cinza.



**Figura 3.8 – Seleção dos padrões de borda e não-borda.**

Para maximizar o desempenho de um classificador é importante que o conjunto de valores escolhidos como entrada seja representativo. Adicionalmente, é desejável que os padrões de aprendizagem possuam uma

baixa variação entre padrões de uma mesma classe (intra-classe) e alta variação entre padrões de classes diferentes (inter-classes).

Com o intuito de aumentar a variabilidade inter-classe, **todos** os padrões de borda e não borda foram submetidos a um processo de comparação após a extração, seguindo o seguinte procedimento: (i) os níveis de cinza do padrão de borda são comparados com os níveis de cinza do padrão de não borda, usando distância Euclidiana; (ii) se a distância for menor que um limiar, os dois padrões devem ser descartados. O limiar que apresentou melhores resultados após a fase de treinamento foi o limiar 90, obtido experimentalmente.

Na etapa de treinamento, foi variado o número de neurônios na camada escondida no intervalo [1; 50], tendo sido treinadas, para cada iteração desta etapa, dez redes neurais, com inicialização aleatória de pesos, a partir do algoritmo *backpropagation*, com critério de parada por validação. O conjunto de amostras utilizado no treinamento foi dividido de acordo com a Tabela 3.2. Foi usado um número equivalente de padrões de borda e não-borda.

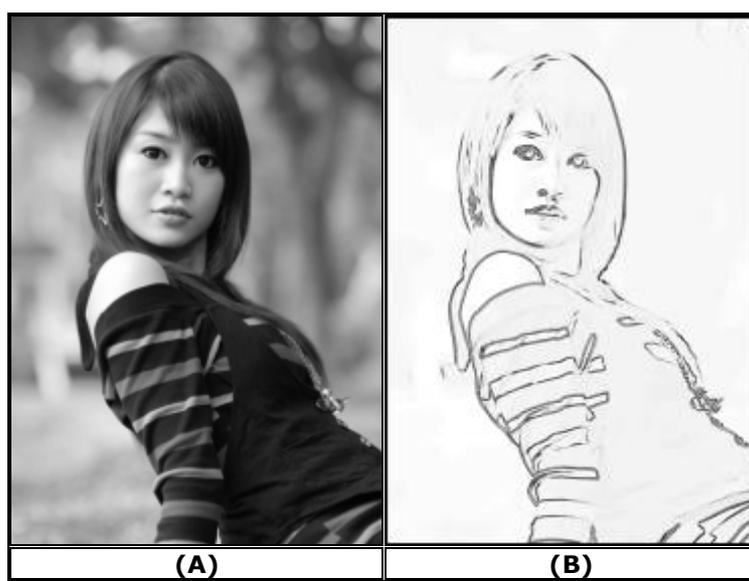
**Tabela 3.2 – Distribuição de padrões.**

<b>Etapa</b>	<b>Padrões de borda + padrões de não borda</b>	<b>Percentual</b>
Treinamento	11942	60%
Teste	2984	15%
Validação	4976	25%
<b>Total</b>	<b>19902</b>	<b>100%</b>

Após o treinamento, foi selecionada a rede que apresentou o melhor resultado visual, por meio da avaliação subjetiva da qualidade, já que uma simples inspeção no menor valor da soma dos erros médios quadráticos entre todos os ciclos do treinamento foi insuficiente para determinar a melhor rede. Ao final, a rede que apresentou melhores resultados visuais apresentou taxa de 96,75% de acerto em relação ao conjunto de testes.

A arquitetura da rede que apresentou melhores resultados, com a

abordagem usando *ground-truth*, possui uma topologia de 19 neurônios na camada de entrada, quantidade obtida após a etapa de extração de componentes principais, 5 neurônios na camada escondida e 2 neurônios na camada de saída. Conforme anteriormente explicitado, a saída desejada é a estimativa de contraste do padrão submetido à entrada. Um exemplo de resultado desta abordagem é ilustrado na Figura 3.9.



**Figura 3.9 – Etapa 4 – Bordas usando treinamento com *Ground-truth*: (A) Suavizada; e (B) Bordas.**

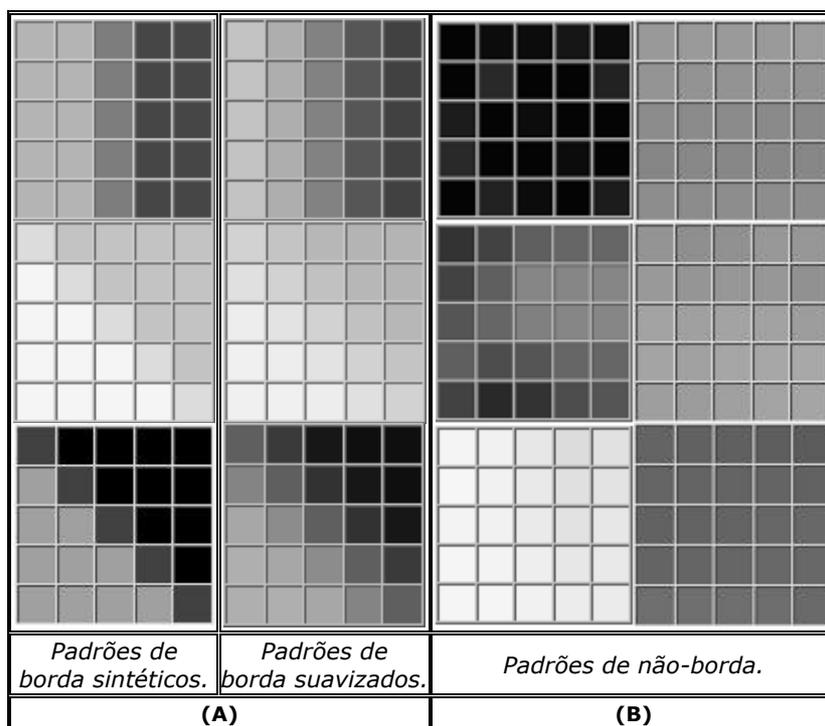
Apesar dos resultados promissores obtidos com o treinamento usando padrões adquiridos conforme descrição apresentada nesta seção, alguns problemas foram identificados, a saber: (i) problemas com a falta de controle do tipo de amostra de borda e não-borda obtido, o que dificultou o treinamento; (ii) problemas na imagem de bordas gerada, e.g. insuficiência de bordas.

### **3.2.2. Treinamento Usando Padrões Sintéticos**

Além dos problemas citados anteriormente com respeito à abordagem de treinamento utilizando amostras manualmente rotuladas, foram identificados outros problemas, a saber: (i) falta de amostras representativas para um grande universo de bordas; (ii) etapa laboriosa para anotar manualmente o *ground-truth*, razão da indisponibilidade de imagens anotadas para uso em testes desta natureza; (iii) falta de

detalhamento de alguns elementos da face após etapa de pós-processamento. Adicionalmente, uma análise usando curvas ROC (vide Capítulo 4) evidenciou um melhor desempenho da rede neural treinada com padrões sintéticos.

Considerando-se os problemas apresentados, uma nova abordagem de treinamento foi experimentada. A nova abordagem fundamenta o treinamento em amostras de padrões gerados sinteticamente. Exemplos de tais padrões sintéticos são apresentados na Figura 3.10.



**Figura 3.10 – Exemplos de padrões sintéticos: (A) Padrões de borda; e (B) Padrões de não-borda.**

Os padrões sintéticos de borda foram obtidos por meio de um algoritmo que, a partir de um par de tons de cinza distintos no intervalo  $[0; 255]$ , gerou padrões que correspondiam a quatro orientações de borda:  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ . Visualmente, o padrão de orientação  $0^\circ$  dá origem a uma linha vertical que corresponde à coluna central do padrão de borda, como pode ser visto no topo da Figura 3.10. A linha vertical corresponde a uma borda de espessura unitária (um *pixel*). Para o padrão de  $0^\circ$ , o menor valor de tom de cinza equivale ao valor da coluna central menos 5. Similarmente, o maior valor do tom de cinza equivale ao valor da coluna

central mais 5. Para aumentar a variabilidade dos padrões de borda sintéticos, os padrões gerados foram suavizados usando uma suavização gaussiana. O mesmo processo aplicado à geração dos padrões com orientação 0° foi repetido para as outras orientações.

Os padrões de não-borda foram obtidos por meio de variação sistemática dos tons de cinza no intervalo [0; 255]. A variação restringiu que um padrão deve conter apenas tons que não variassem acima ou abaixo de 20% da intensidade do pixel central. Além dessa restrição, 18 ou mais *pixels*, no conjunto de 25 *pixels* de cada padrão de não borda, deveriam ter pouca variação nos valores dos tons de cinza, originando um padrão homogêneo, sem transição abrupta de valores.

Seguindo o roteiro de treinamento usado para os conjuntos obtidos a partir de anotações do conjunto de *ground-truth*, a etapa de treinamento consistiu na variação do número de neurônios na camada escondida (intervalo [1; 50]) e treinamento de dez redes neurais (inicialização aleatória de pesos, parada por validação e *backpropagation*). O conjunto de padrões de borda utilizado antes da técnica PCA foi dividido de acordo com a Tabela 3.3.

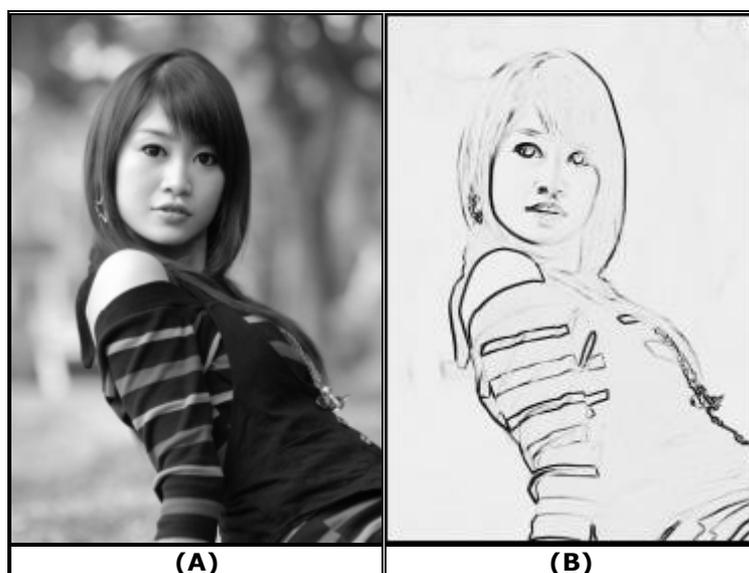
**Tabela 3.3 – Distribuição de padrões de borda sintéticos.**

<b>Etapa</b>	<b>Padrões de borda + padrões de não borda</b>	<b>Percentual</b>
Treinamento	766	60%
Teste	191	15%
Validação	319	25%
<b>Total</b>	<b>1276</b>	<b>100%</b>

De forma equivalente à abordagem anterior, foi selecionada a rede que apresentou o melhor resultado visual, por meio da avaliação subjetiva da qualidade, pois uma inspeção no menor valor da soma dos erros médios quadráticos nos ciclos de treinamento foi insuficiente para determinar a melhor rede. Levando este fato em consideração, a rede que apresentou melhores resultados visuais apresentou taxa de 98,5% de

acerto em relação ao conjunto de testes.

A arquitetura da rede que apresentou melhores resultados com a abordagem usando padrões de borda sintéticos possui 11 neurônios na camada de entrada – quantidade obtida após a etapa de extração de componentes principais; 2 neurônios na camada escondida; e 2 neurônios na camada de saída. De acordo com o que já foi explicitado, a saída desejada é a estimativa de contraste do padrão submetido à entrada. Um exemplo do resultado desta abordagem é mostrado na Figura 3.11.



**Figura 3.11 – Etapa 4 – Detecção de bordas usando treinamento com padrões sintéticos: (A) Imagem Suavizada; e (B) Bordas.**

### **3.3. Pós-Processamento**

Após a geração do mapa de bordas, a imagem é submetida a duas etapas de pós-processamento, a saber: (i) correção de brilho e contraste; e (ii) ajuste de curvas. O brilho e contraste da imagem resultante da etapa anterior são modificados segundo valores fixos, sendo o valor do ajuste do brilho igual a -80 e o valor do ajuste para o contraste igual a 80. Tais valores foram obtidos após votação com usuários para calibração de parâmetros, conforme apresentado no Capítulo 4, na Seção 4.3.1. Os efeitos da aplicação da correção de brilho e contraste, usando os valores fixos mencionados, podem ser visualizados na Figura 3.12.

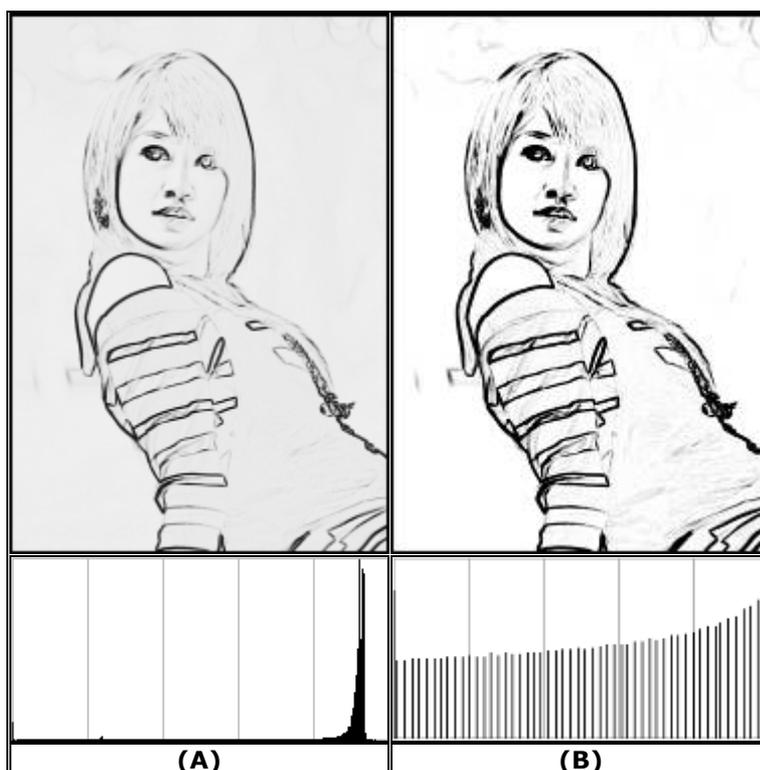


Figura 3.12 – Etapa 5 - Ajuste de brilho e Contraste: (A) Mapa de Bordas; e (B) Imagem modificada.

Após a modificação do brilho e contraste, ocorre a última etapa da abordagem proposta, uma operação histogrâmica por operador exponencial, que será discutida na próxima seção.

### 3.3.1. Transformação de Histograma

O operador exponencial pode ser aplicado para modificar a faixa dinâmica do mapa de bordas, enfatizando os valores de pixels de alta intensidade. No caso do operador “elevado a potência de”, o valor dos pixels da imagem de saída é igual ao valor da imagem de entrada (base) elevado a uma potência fixa. Após votação com usuários para calibração de parâmetros, conforme apresentado no Capítulo 4, na Seção 4.3.1, a potência que se revelou mais adequada à transformação histogrâmica foi a potência 4.

Tal operação, no contexto desta pesquisa, visou ressaltar as linhas de bordas mais importantes, fazendo com que mais detalhes sejam percebidos na representação não foto-realística gerada, a fim de melhorar o aspecto visual final. Um exemplo da aplicação desta operação na imagem de bordas após a modificação de brilho e contraste é mostrado na Figura 3.13.

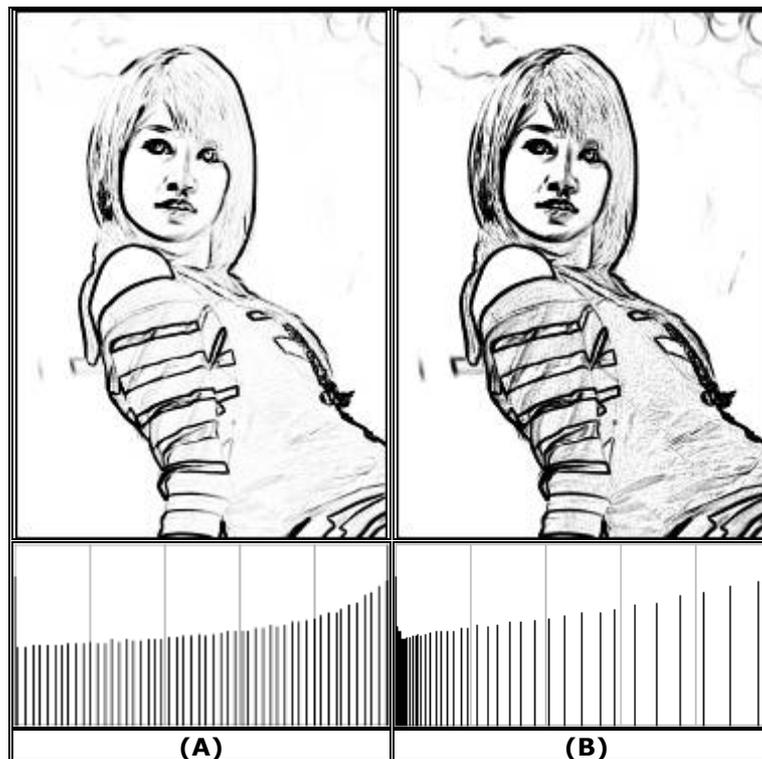


Figura 3.13 – Etapa 6 – Transformação histogrâmica: (A) Bordas após ajuste de brilho e Contraste; e (B) Imagem final.

Para efeitos de comparação, exemplos de imagens geradas usando a abordagem de treinamento com padrões obtidos de um conjunto de *ground-truths* e com treinamento usando padrões de bordas sintéticas podem ser vistos na Figura 3.14. Subjetivamente, a abordagem com o uso de padrões sintéticos para treinamento apresenta um resultado de melhor qualidade, uma vez que atributos faciais importantes, como o nariz e a boca, podem ser melhor visualizados.



Figura 3.14 – Comparação: (A) Treinamento com GT; e (B) Treinamento com bordas sintéticas.

### 3.4. Detalhes de Projeto e Implementação

Para a validação da abordagem proposta apresentada nas seções anteriores, foi necessário desenvolver um sistema para a geração de representações não foto-realísticas, seguindo as etapas e o fluxo de execução apresentado e discutido nas seções anteriores. Nesta seção, são apresentados detalhes de projeto e implementação do *software* desenvolvido para a validação da abordagem descrita, assim como para a realização dos testes descritos no Capítulo 5.

A implementação foi realizada utilizando a linguagem C++, a partir do ambiente de desenvolvimento integrado (Integrated Development Environment - IDE) Microsoft Visual Studio 2005. A linguagem C++ foi escolhida, dentre vários aspectos, por: (i) apresentar um desempenho superior para a tarefa, em relação a outras linguagens; (ii) ser muito utilizada na área de Visão Computacional; (iii) se dispor de bibliotecas tais como a Intel OpenCV, dotada de facilidades destinadas ao suporte no desenvolvimento do trabalho; (iv) possibilitar o uso de redes neurais mediante ligação com bibliotecas do Matlab.

Há, no entanto, algumas desvantagens no uso da linguagem e do ambiente, tais como a dependência em relação ao sistema operacional Microsoft Windows, problemas de gerenciamento de memória e apontadores e peculiaridades do ambiente de desenvolvimento. O ambiente de desenvolvimento Microsoft Visual Studio 2005 foi escolhido pela disponibilidade de utilização a partir do programa MSDN Academic Alliance e pela familiaridade do desenvolvedor com suas funcionalidades. O projeto do *software*, em termos de classes, é apresentado na Figura 3.15.

A classe base *Classificador* tem como classe derivada a representação de um classificador baseado em redes neurais. Tal arquitetura facilita a inclusão de novos classificadores baseados em outro paradigma diferente das redes neurais. A classe *ClassificadorBordasNeural* faz uso das classes *PCA* e da representação de uma rede neural como matriz (classe *RedeNeuralComoMatriz*), utilizando resultados de treinamento exportados a partir do MATLAB.

Como elemento centralizador das operações, aparece a classe *NPR\_SketchApp* que agrega operações de leitura e escrita de arquivos e centraliza operações referentes ao fluxo de execução da abordagem proposta, fazendo uso dos vários módulos implementados, como *PosProcessador* e *CorretorCromatico*.

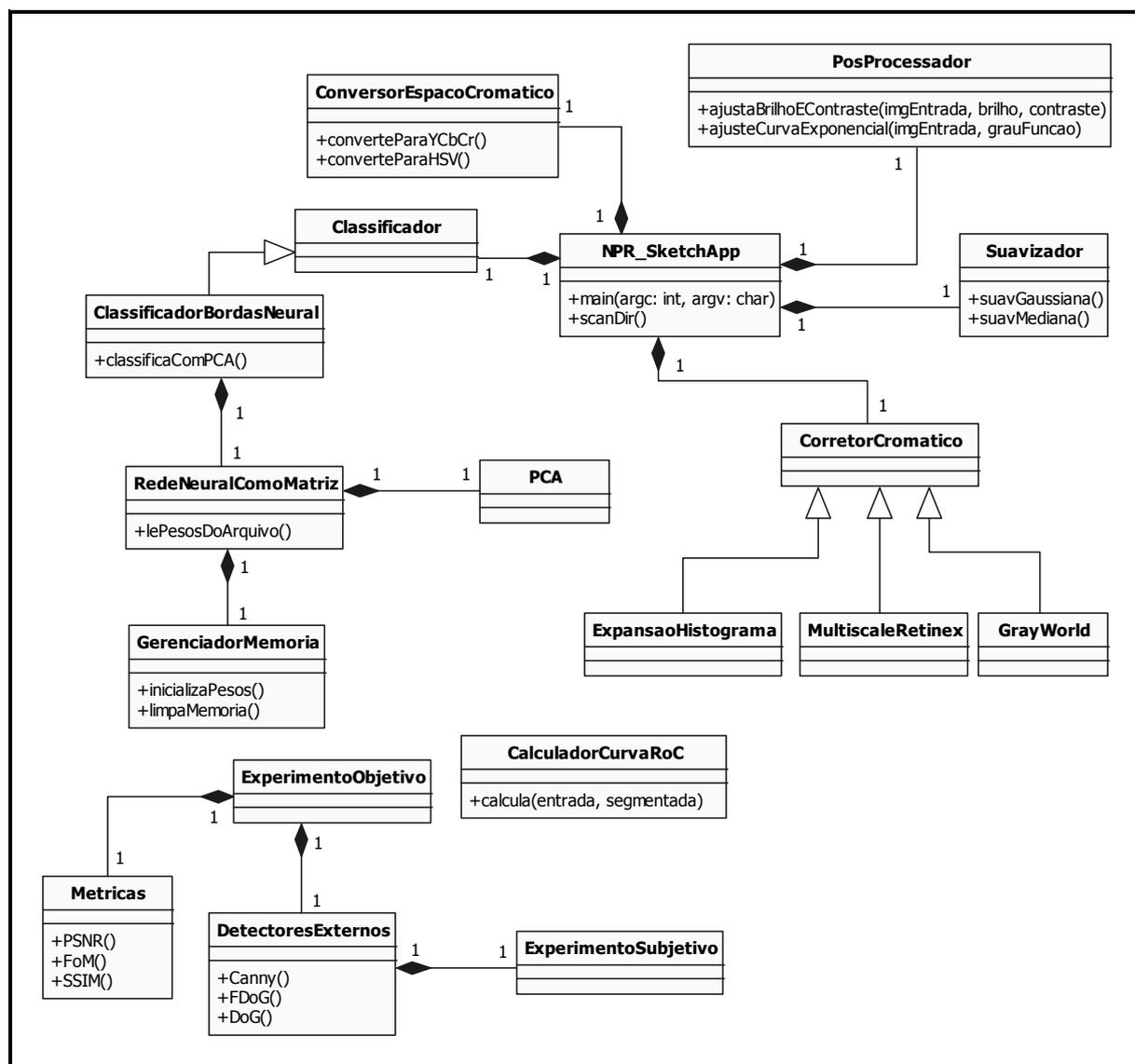


Figura 3.15 – Diagrama de classes do software desenvolvido.

### 3.5. Considerações Finais do Capítulo

Neste capítulo, foi proposta uma abordagem para a geração de representações não foto-realísticas, a qual é dividida nas seguintes sub-etapas: (i) *correção cromática*; (ii) *conversão de espaço cromático*; (iii) *suavização*; (iv) *detecção de bordas*; (v) *ajuste de brilho e contraste*; e (vi) *transformação histogrâmica* por operador exponencial.

Na etapa inicial, há um processamento para melhoria da iluminação da cena em questão, usando um algoritmo de constância cromática (*GrayWorld*). Após a melhoria das cores da cena, há uma conversão do espaço cromático RGB para um espaço cromático mais adequado a geração de representações não foto-realísticas (espaço *YCbCr*). Em seguida, ocorre uma suavização da imagem para supressão de detalhes, para melhoria do resultado gerado e para remoção de ruído (suavização gaussiana). Há, então, a detecção de bordas usando uma rede neural do tipo *Perceptron* Multicamadas, treinada com o algoritmo *backpropagation* usando padrões de bordas sintéticas. Por fim, o brilho e contraste da imagem de bordas são modificados. Adicionalmente, ocorre uma modificação histogrâmica com o uso de um operador exponencial.

Algumas vantagens do método proposto incluem: (i) a adição de um algoritmo de constância cromática para a atenuação de sombras e efeitos indesejáveis quando a imagem de entrada possui condições de iluminação desfavoráveis; (ii) uso do canal Y, possibilitando sub-amostragem de valores de cinza sem perda de detalhe perceptível; e (iii) uso de redes neurais na confecção de representações não foto-realísticas, fato escasso na literatura pesquisada.

O próximo capítulo contém a descrição dos experimentos realizados e a discussão dos resultados obtidos. A apresentação dos experimentos contempla os testes realizados nos vários caminhos disponíveis até que fosse atingida a abordagem descrita nesse capítulo. O próximo capítulo contém ainda gráficos de avaliação dos classificadores treinados com uso da análise da curva ROC, bem como uma análise comparativa, objetiva e subjetiva, com outras técnicas de geração de representações não foto-realísticas.

# Capítulo 4

## Apresentação e Discussão dos Experimentos e Resultados

Neste capítulo, são descritos e comentados os experimentos realizados para validação da abordagem proposta para geração de representações não foto-realísticas, apresentada no Capítulo 3, assim como os resultados associados. Vários experimentos foram conduzidos em etapas distintas do estudo, sendo cada uma dessas etapas descrita em seções separadas, a saber:

- Na Seção 4.1, é apresentada uma comparação entre dois classificadores de borda. A diferença entre cada classificador reside no método de treinamento, discutido no Capítulo 3. A análise de desempenho e comparações entre classificadores é apresentada por meio de análise de curvas ROC (*Receiver Operating Characteristics*).

- Na Seção 4.2, são apresentados resultados de comparação objetiva, usando métricas numéricas tradicionais na literatura estudada. A Seção 4.2 se divide em duas partes: (i) *comparação numérica* entre todos os caminhos estudados para a montagem da abordagem proposta, i.e., análise da melhor combinação de algoritmos de pré-processamento para a geração de representações não foto-realísticas; e (ii) *comparação objetiva* com outros detectores de borda. Para a obtenção dos resultados explicitados na Seção 4.2 foi usado o classificador com melhor desempenho, conforme descrição contida na Seção 4.1.

- Na Seção 4.3, relata-se a realização de um experimento subjetivo, o qual fundamentou-se em um esquema de votação envolvendo usuários de teste. Tal validação subjetiva dividiu-se em duas partes: (i) *calibração de parâmetros* para quatro detectores de bordas (Canny, DoG, FDoG e a Abordagem Proposta); e (ii) *votação de imagens*.

## 4.1. Curva ROC

De acordo com Fawcett (2006), uma curva ROC é um gráfico usado para visualizar, organizar e selecionar classificadores com base no desempenho. Trata-se de um gráfico de valores de sensibilidade *versus* especificidade de um classificador binário, quando há variação de seu limiar de classificação. Gráficos de curvas ROC são, há muito, utilizados na teoria de sinais, pois permitem observar as compensações entre alarmes falsos (*false alarms*) e taxa de acerto (*hit rate*). A análise de curvas ROC foi estendida para a análise e visualização de diagnósticos médicos, sendo largamente usadas na Medicina (Zou, 2002).

Os classificadores de bordas avaliados neste trabalho possuem um conjunto de parâmetros de treinamento que influenciam os resultados obtidos. Objetivando a investigação do melhor método de treinamento, foi feita uma análise quantitativa baseada no conceito de curvas ROC. Aos classificadores foram apresentados exemplos de borda e não-borda, assim, o problema de classificação consiste em enquadrar a entrada em uma das classes disponíveis do classificador binário (borda ou não-borda), tendo como resultado uma estimativa de contraste, i.e., qual a “força” da borda para um determinado pixel. Considerando um problema de classificação binária para bordas, quatro saídas são possíveis. Tais saídas estão sumariadas na Tabela 4.1.

Tabela 4.1 – Métricas usadas na análise ROC.

Métrica	Descrição
TP	Positivos verdadeiros. Valor que representa a condição: é borda na imagem <b>e</b> é borda no <i>ground-truth</i> .
TN	Negativos verdadeiros. Valor que representa a condição: não é borda na imagem <b>e</b> não é borda no <i>ground-truth</i> .
FN	Negativos falsos. Valor que representa a condição: não é borda na imagem <b>mas</b> é borda no <i>ground-truth</i> .
FP	Positivos falsos. Valor que representa a condição: é borda na imagem <b>mas</b> não é borda no <i>ground-truth</i> .
Taxa de TP	$TPR = TP / (TP + FN)$
Taxa de FP	$FPR = FP / (FP + TN)$

Dois classificadores de borda foram treinados: (i) um usando aprendizagem com extração de bordas de imagens com base em um *ground-truth*; e (ii) outro usando aprendizagem com extração de bordas

sintéticas. A curva ROC para os dois classificadores é apresentada na Figura 4.1. Cada ponto da curva representa a média de TPR *versus* FPR em um conjunto de 14 imagens de *ground-truths*, anotadas manualmente e apresentadas no Apêndice C.

Para a obtenção das médias de TPR e FPR de cada classificador, que representam, respectivamente, as coordenadas Y e X de cada ponto apresentado na Figura 4.1, foi variado o intervalo da estimativa de contraste da saída desejada. A análise visual da curva permite constatar que o classificador treinado com bordas sintéticas tem melhor desempenho do que o classificador treinado com bordas obtidas com o uso de *ground-truths*.

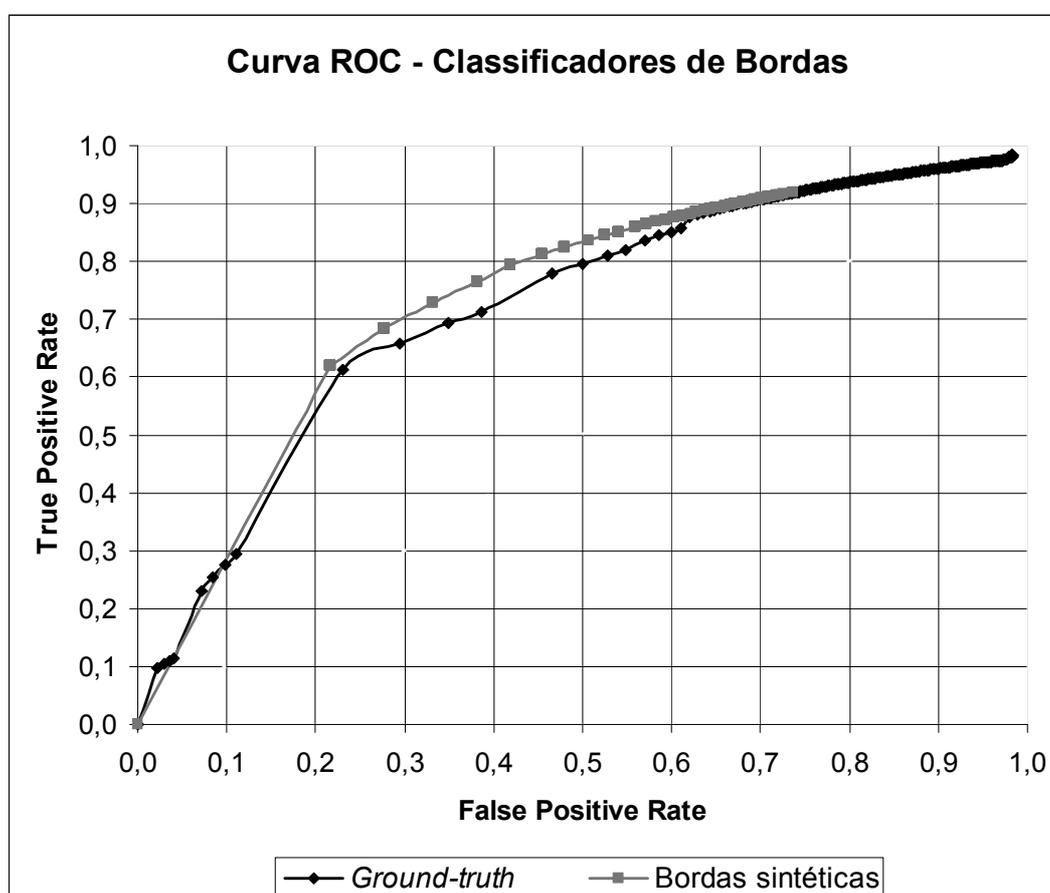


Figura 4.1 – Curva ROC para os classificadores estudados.

Portanto, sendo uma curva ROC uma descrição pictórica de desempenho de classificadores, a comparação entre classificadores deve ser feita com base na inspeção visual. Adicionalmente, a comparação entre múltiplos classificadores é realizada pela comparação de um único valor escalar, que representa o desempenho esperado de cada classificador. De acordo com Fawcett (2006), um valor escalar comum na literatura é o valor

da área sob a curva (AUC – **A**rea **U**nder the **C**urve). Valores para a AUC são apresentados no intervalo [0, 1]. No entanto, uma área menor que 0,5 é indesejável, pois um classificador com tal área tem como saída uma resposta aleatória de classificação (*random guess*).

É possível que um classificador seja melhor do que outro em uma determinada área da curva ROC, e.g., na região próxima ao ponto (0,05; 0,1) da Figura 4.1. No entanto, a AUC, enquanto medida geral de desempenho, prevê com certa acurácia o melhor classificador. A AUC pode ser calculada usando trapézios sucessivos, usando os pontos que dão origem à curva ROC. Para os classificadores estudados, tem-se que o classificador treinado com bordas sintéticas resulta em uma área maior da curva ROC, conforme Tabela 4.2:

**Tabela 4.2 – AUC para curvas ROC.**

<b>Treinamento</b>	<b>AUC</b>
Usando <i>ground-truth</i>	0,718583
Usando bordas sintéticas	0,738617

Diante das evidências apresentadas, o classificador que obteve melhor desempenho foi o classificador treinado com bordas sintéticas. Após essa escolha, procedeu-se a avaliação objetiva e subjetiva da abordagem proposta, apresentadas nas próximas seções.

## **4.2. Experimentos Objetivos**

Cada métrica objetiva usada nesta seção é brevemente detalhada no Apêndice E. Nesta seção são apresentados os resultados comparativos usando cada uma das métricas. Ressalta-se que, como apresentado no Apêndice E, mesmo com os problemas apresentados nas métricas clássicas como a FoM e sendo a métrica PCM de difícil implementação, uma comparação objetiva fez-se necessária para avaliar o desempenho da abordagem proposta. Tal comparação objetiva serviu, primeiramente, para a comparação de diferentes decisões a serem tomadas, e.g., o tipo de algoritmo de pré-processamento mais adequado ou o canal cromático mais adequado a geração de representações não foto-realísticas. Em segundo

lugar, a avaliação objetiva foi usada para comparar a abordagem proposta com outros geradores de representações não foto-realísticas.

Três métricas de similaridade foram usadas: (i) Relação *Sinal/Ruído* Máxima (**Peak Signal to Noise Ratio** – PSNR); (ii) *Figura de Mérito* de Pratt (**Pratt's Figure of Merit** - FoM) (Pratt, 1978); e (iii) *Similaridade Estrutural* (**Structural SIMilarity** - SSIM) (Wang, Lu e Bovik, 2002; Wang e Simoncelli, 2005). Tais métricas procuram a compatibilidade entre duas imagens, a fim de estimar a similaridade entre elas, seja por meio de comparação pixel a pixel ou baseada em vizinhança.

Apesar da existência de uma base de imagens com os respectivos *ground-truths* (Martin et al., 2001), anotados após um experimento com usuários, uma inspeção da referida base evidenciou que sua composição envolve majoritariamente cenas naturais e paisagens, havendo apenas 2 imagens com faces humanas. Ademais, a avaliação de imagens de faces não era o foco de estudo dos autores. No entanto, tal fato não invalidou a busca por um denominador comum com respeito a avaliação de detectores de borda.

Em virtude da lacuna apresentada anteriormente, os experimentos objetivos apresentados nesta seção usaram 14 imagens com *ground-truths* anotados manualmente. As imagens e os respectivos *ground-truths* podem ser visualizadas no Apêndice B. Os parâmetros da abordagem proposta foram calibrados para uso em imagens de face, sendo os valores de cada parâmetro apresentados no Apêndice A.

É importante salientar que a avaliação objetiva por meio das métricas usadas foi possível apenas em **imagens binárias**, pelo fato de o *ground-truth* ter sido concebido apenas em preto-e-branco. No entanto, o aspecto visual de algumas renderizações não foto-realísticas possuem como componente importante o efeito gerado pela variação em tons de cinza da imagem final. Todas as avaliações objetivas foram realizadas após a limiarização com o valor de corte igual a 127.

Como afirmado por Jain (1991), a comparação simples das médias dos resultados podem levar a conclusões não confiáveis. Devido a este fato,

após a medição dos valores numéricos de cada métrica, foi realizada uma breve análise estatística dos dados. O propósito da análise estatística foi investigar evidências de variações estatisticamente significativas nos grupos considerados, tendo em vista tratar-se de um experimento ao longo do qual várias categorias de algoritmos podiam ser consideradas.

O procedimento estatístico mais adequado ao propósito almejado foi o teste F ANOVA fator único (Levine, Berenson e Stephan, 2000), pois há apenas um único fator a ser testado: o valor de similaridade da imagem com o *ground-truth*. Uma vez que este teste evidencia apenas diferenças entre médias, não possibilitando comparações entre pares de grupos, utilizou-se, quando a análise estatística evidenciava a possibilidade, o teste de Tukey (Levine, Berenson e Stephan, 2000).

Assim, a hipótese nula de nenhuma diferença nas médias aritméticas

$$H_0: \mu_{G1} = \mu_{G2} = \dots = \mu_{G12}$$

será testada, em relação à hipótese alternativa de haver diferenças entre as referidas médias ( $H_a$ : nem todas as médias são iguais). A média  $\mu_{Gi}$  significa a média dos valores obtidos para as 14 imagens, sendo  $Gi$  o  $i$ -ésimo grupo de combinações possíveis para os algoritmos de pré-processamento. Deve-se ressaltar, no entanto, que o teste F ANOVA fator único permitiu verificar se a interpretação dos casos pode ser feita, ou seja, se as médias das amostras coletadas são significativamente diferentes com um grau de confiança igual a 95%.

Heath (1996) afirma que uma das limitações de métodos de comparação de borda que usam *ground-truth* é a dependência da marcação manual, bem como o fato de que, historicamente, alguns testes dessas métricas terem sido realizados com imagens sintéticas ou imagens reais pouco complexas, o que facilita a geração de *ground-truth*, mas não avalia o desempenho dos detectores em condições reais de uso. Para mitigar o viés da avaliação objetiva e avaliar o desempenho de imagens geradas em escala de cinza, uma avaliação subjetiva foi conduzida, a qual é apresentada na Seção 4.3.

### 4.2.1. Relação Sinal/Ruído Máxima (PSNR)

Na Tabela 4.3, é mostrada a comparação de diferentes estratégias da abordagem proposta usando a métrica PSNR, com o maior valor obtido destacado **em negrito**.

Tabela 4.3 – Avaliação objetiva usando PSNR.

Imagem	GrayWorld				Retinex				Expansão de histograma			
	Canal Y		Canal V		Canal Y		Canal V		Canal Y		Canal V	
	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
1	6,80456	6,87699	7,02440	7,06384	5,64986	5,82649	5,92577	6,09668	6,78579	6,85966	6,99568	7,04717
2	7,58620	7,73303	7,78580	7,92708	6,83610	7,00270	7,05089	7,22176	7,49509	7,64293	7,26112	7,40033
3	7,04561	7,14232	7,17203	7,22940	6,30448	6,51565	6,06571	6,29144	6,84686	6,94080	6,70060	6,78803
4	8,29647	8,33607	9,15878	9,12455	7,93990	8,02404	8,74810	8,79097	8,28841	8,33502	9,10043	9,11642
5	7,63174	7,67486	8,07531	8,07456	7,04557	7,19464	7,74346	7,82401	7,78195	7,81525	8,12191	8,13144
6	8,73936	8,77798	8,37392	8,42178	8,40715	8,46547	8,20221	8,30656	8,63009	8,67674	8,68067	8,72841
7	8,38595	8,41399	8,15871	8,22752	6,70153	6,86392	6,74048	6,92094	8,37616	8,39663	7,80541	7,82330
8	6,15994	6,22599	5,99034	6,06114	5,62263	5,72832	5,72227	5,81640	6,20041	6,26046	6,09672	6,17023
9	8,75645	8,84147	8,33108	8,42449	8,15817	8,30883	8,33255	8,46014	8,79200	8,88429	8,72497	8,82625
10	9,02967	9,14071	8,97188	9,08480	7,12782	7,37935	7,05966	7,32814	8,93646	9,05049	8,92451	9,03372
11	9,00363	9,06039	8,91766	8,98532	7,92770	8,14074	7,88290	8,07351	8,83489	8,90836	8,84778	8,90991
12	8,61059	8,61467	8,04164	8,03945	8,01331	8,05382	7,83312	7,85189	8,49999	8,51160	8,32135	8,31165
13	5,68331	5,70133	5,23549	5,29231	5,60980	5,65721	5,66974	5,70123	5,56203	5,58744	6,49175	6,51350
14	11,10500	11,04290	10,89760	10,88900	7,93360	8,17169	8,09441	8,39600	10,84610	10,79210	9,78569	9,75352
$\bar{X}$	8,05989	<b>8,11305</b>	8,00962	8,06037	7,09126	7,23806	7,21938	7,36283	7,99116	8,04727	7,98990	8,03956
s	1,38082	1,36117	1,39148	1,37815	1,00453	1,00510	1,04990	1,04802	1,34338	1,32708	1,11827	1,10415

Após a obtenção dos resultados numéricos para a métrica PSNR, o teste F ANOVA fator único foi realizado (Tabela 4.4), com  $\alpha = 0,05$ , com a intenção de verificar a variação significativa ao longo dos grupos considerados.

Tabela 4.4 – Teste F ANOVA para valores de PSNR.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	25,23999	11	2,294544	<b>1,54252</b>	0,121439	<b>1,850485</b>
Dentro dos grupos	232,0546	156	1,487529			
Total	257,2946	167				

Pela comparação simples dos resultados obtidos após a avaliação com PSNR, a melhor estratégia a adotar é o uso do algoritmo GrayWorld no canal Y do espaço YCbCr e do filtro de suavização da mediana. No entanto, conforme elucidado pelo teste F ANOVA fator único explicitado na Tabela 4.4, o valor F é *menor* do que  $F_{crit}$ , significando que **não há diferenças** estatisticamente significativas entre as médias das várias estratégias contempladas, quando usada a avaliação com métrica PSNR.

## 4.2.2. Similaridade Estrutural (SSIM)

Na Tabela 4.5, pode ser visualizada a comparação entre diferentes estratégias contempladas pela abordagem proposta usando a métrica SSIM, sendo **destacado** o maior valor obtido.

Tabela 4.5 – Avaliação objetiva usando SSIM.

Imagem	GrayWorld				Retinex				Expansão de histograma			
	Canal Y		Canal V		Canal Y		Canal V		Canal Y		Canal V	
	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
1	0,13099	0,13051	0,12960	0,12868	0,12700	0,12740	0,12798	0,12806	0,13079	0,13028	0,13059	0,13010
2	0,19100	0,19042	0,19256	0,19209	0,18743	0,18673	0,19124	0,19156	0,19051	0,18983	0,18936	0,18884
3	0,12974	0,12936	0,12829	0,12749	0,12460	0,12588	0,12337	0,12488	0,12909	0,12861	0,12763	0,12693
4	0,18630	0,18496	0,19368	0,19187	0,18372	0,18345	0,18960	0,18863	0,18601	0,18469	0,19313	0,19226
5	0,17066	0,17065	0,18184	0,18124	0,16296	0,16530	0,17599	0,17671	0,17355	0,17345	0,18192	0,18217
6	0,17201	0,17195	0,16736	0,16662	0,16721	0,16742	0,16340	0,16399	0,17141	0,17110	0,17092	0,17102
7	0,17258	0,17215	0,17019	0,16962	0,16028	0,16094	0,16131	0,16191	0,17272	0,17211	0,16843	0,16744
8	0,13577	0,13552	0,13645	0,13621	0,12803	0,12901	0,13367	0,13453	0,13644	0,13603	0,13606	0,13640
9	0,20402	0,20439	0,20082	0,20139	0,19815	0,19944	0,19885	0,19952	0,20459	0,20484	0,20045	0,20050
10	0,21575	0,21607	0,21518	0,21578	0,18718	0,18897	0,18520	0,18719	0,21469	0,21511	0,21425	0,21465
11	0,20289	0,20335	0,20171	0,20193	0,19482	0,19610	0,19497	0,19603	0,20175	0,20193	0,20182	0,20159
12	0,17696	0,17651	0,17094	0,17010	0,17389	0,17336	0,17055	0,16923	0,17616	0,17575	0,17440	0,17370
13	0,12783	0,12360	0,10785	0,10712	0,12724	0,12533	0,12468	0,12120	0,12488	0,12058	0,14427	0,14159
14	0,25170	0,24960	0,24565	0,24513	0,21232	0,21202	0,21337	0,21477	0,24819	0,24606	0,23775	0,23588
$\bar{X}$	0,17630	0,17565	0,17444	0,17395	0,16677	0,16724	0,16816	0,16844	0,17577	0,17503	<b>0,17650</b>	0,17593
s	0,03649	0,03674	0,03825	0,03847	0,02975	0,02985	0,03016	0,03064	0,03612	0,03642	0,03297	0,03302

Após a verificação dos resultados numéricos obtidos pela métrica SSIM, o teste F ANOVA fator único foi conduzido, e é mostrado na Tabela 4.6. O teste F ANOVA foi realizado com a intenção de verificar se havia variação significativa ao longo dos grupos considerados.

Tabela 4.6 – Teste F ANOVA para valores de SSIM.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,002371	11	0,000216	<b>0,184008</b>	0,99824	<b>1,850485</b>
Dentro dos grupos	0,182764	156	0,001172			
Total	0,185136	167				

Por meio da comparação de resultados obtidos após avaliação com SSIM, a melhor estratégia a adotar é o uso da expansão histogrâmica no canal V do espaço HSV e da suavização gaussiana. No entanto, conforme elucidado pelo teste F ANOVA fator único presente na Tabela 4.6, o valor F é *menor* do que  $F_{crit}$ , o que implica que **não há diferenças** estatisticamente aceitáveis entre as médias dos vários caminhos propostos quando usada a avaliação com métrica SSIM.

### 4.2.3. Figura de Mérito (FoM)

Na Tabela 4.7, é apresentada a comparação das diferentes estratégias contempladas pela abordagem proposta usando a métrica FoM, sendo **destacado** o maior valor obtido.

Tabela 4.7 – Avaliação objetiva usando FoM.

Imagem	GrayWorld				Retinex				Expansão de histograma			
	Canal Y		Canal V		Canal Y		Canal V		Canal Y		Canal V	
	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
1	0,55920	0,52002	0,48631	0,48625	0,50478	0,50366	0,49942	0,49860	0,51104	0,54987	0,48387	0,48285
2	0,43747	0,39878	0,37239	0,37381	0,36082	0,35945	0,36727	0,36819	0,38964	0,42923	0,37195	0,37095
3	0,67064	0,62907	0,59144	0,59089	0,59322	0,59398	0,57593	0,57990	0,62110	0,65803	0,59594	0,59237
4	0,41807	0,37650	0,34367	0,34128	0,32574	0,32503	0,30509	0,30490	0,36765	0,40489	0,33463	0,33368
5	0,41328	0,37849	0,35663	0,36017	0,29832	0,30734	0,32046	0,32759	0,37237	0,41579	0,35277	0,35436
6	0,51003	0,46645	0,41199	0,40669	0,41153	0,40609	0,36641	0,36953	0,46539	0,49765	0,42195	0,41616
7	0,58043	0,53477	0,50193	0,50130	0,46871	0,46786	0,44784	0,45093	0,53185	0,56449	0,50403	0,49854
8	0,61327	0,57322	0,54860	0,54973	0,53513	0,53607	0,55471	0,55761	0,56616	0,60415	0,54080	0,54152
9	0,54110	0,50207	0,45213	0,45514	0,42572	0,42720	0,41595	0,41936	0,49300	0,53202	0,42435	0,41621
10	0,48561	0,45191	0,42147	0,42819	0,33903	0,34867	0,32938	0,34049	0,43717	0,48180	0,41450	0,41913
11	0,52690	0,49143	0,45630	0,45872	0,45092	0,44570	0,44510	0,44338	0,47907	0,52027	0,45643	0,45892
12	0,54250	0,50493	0,45177	0,45203	0,50195	0,50048	0,48359	0,47959	0,49580	0,53588	0,47135	0,46972
13	0,50913	0,46371	0,36971	0,36808	0,44321	0,43607	0,44167	0,43359	0,46001	0,49322	0,49046	0,48375
14	0,47189	0,43060	0,37713	0,38074	0,25080	0,26250	0,23790	0,25221	0,42095	0,45741	0,36087	0,35866
$\bar{x}$	<b>0,51997</b>	0,48014	0,43868	0,43950	0,42213	0,42286	0,41362	0,41613	0,47223	0,51034	0,44456	0,44263
s	0,07332	0,07235	0,07445	0,07435	0,09745	0,09439	0,09777	0,09495	0,07278	0,07181	0,07587	0,07516

Foi conduzido o teste F ANOVA fator único para os valores obtidos com a métrica FoM, com  $\alpha = 0,05$ , com a intenção de verificar a variação significativa ao longo dos grupos considerados (Tabela 4.8).

Tabela 4.8 – Teste F ANOVA para valores de FoM.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,198802	11	0,0180729	<b>2,693397</b>	0,003363	<b>1,850485</b>
Dentro dos grupos	1,046774	156	0,0067101			
Total	1,245576	167				

Conforme elucidado pelo teste F ANOVA fator único presente na Tabela 4.8, o valor F é maior do que  $F_{crit}$ . Portanto, **há diferenças** estatisticamente significativas entre as médias dos vários caminhos propostos quando usada a avaliação com métrica FoM.

O teste F ANOVA fator único evidencia a presença de diferenças entre médias, mas não possibilita comparações entre pares de grupos. Para melhor inferir o conjunto de algoritmos mais adequados a serem usados na

abordagem proposta, foi conduzido um teste de Tukey para múltiplas comparações de médias. Os resultados do teste de Tukey com um intervalo de confiança de 90% para os valores apresentados na Tabela 4.7 podem ser vistos no Apêndice D.

Diferentemente do intervalo de confiança de 95% utilizado para o teste F ANOVA fator único, foi utilizado no teste de Tukey um intervalo de confiança de 90% ( $\alpha=0,1$ ). A razão para tal diferença residiu no fato de a adoção de um intervalo de confiança de 95% para o teste de Tukey não ter permitido bom delineamento do melhor conjunto de algoritmos a ser utilizado na abordagem proposta. Em outras palavras, diferenças significativas na comparação dos pares de grupos foram melhor ressaltadas quando o intervalo de confiança de 90% foi utilizado.

Levando em consideração os grupos G1 a G12, pode-se inferir, a partir da análise do teste de Tukey, que o agrupamento que contém o Retinex como algoritmo de constância cromática como etapa inicial do pré-processamento (grupos G5 a G8) difere significativamente do grupo G1 (*GrayWorld* → canal Y → suavização gaussiana). Adicionalmente, pode ser inferido que os grupos que usam expansão de histograma (G9 a G12) não diferem significativamente do grupo G1.

Pelo fato do G1 ter obtido maior média quando usada a métrica FoM, pela diferença significativa entre as médias computadas para a referida métrica, evidenciada pelo teste F ANOVA fator único e pelas análises de múltiplas comparações, por meio do teste de Tukey, há evidências de que o **melhor conjunto** de algoritmos de pré-processamento é a composição usada no **G1**: uso do algoritmo *GrayWorld* no canal Y do espaço YCbCr, seguido de suavização gaussiana.

No entanto, algumas ressalvas devem ser feitas quanto à métrica FoM. Chabrier et al. (2008) revisaram várias métricas disponíveis para avaliação supervisionada (com uso de *ground-truth*) de técnicas de detecção de contornos. Os resultados, após avaliação de imagens sintéticas e reais,

acusaram que a métrica FoM exibe resultados mais discriminatórios do que as outras técnicas avaliadas. Os autores afirmaram ainda que a métrica FoM não possui prova teórica sendo, no entanto, uma das mais utilizadas. No entanto, tal métrica não é simétrica nem expressa corretamente supersegmentação ou subsegmentação, além de que é sensível a problemas de localização, i.e., um número idêntico de pixels de não-borda, localizados em partes diferentes da imagem, podem resultar em um mesmo valor para a métrica. Todavia, tal métrica mostrou-se a de uso em maior escala na literatura revisada, o que valida, com ressalvas, as comparações apresentadas nesta seção. Na próxima seção, as métricas apresentadas foram usadas para comparação objetiva entre os melhores resultados obtidos com a abordagem proposta e outros algoritmos de geração de representações não foto-realísticas.

#### **4.2.4. Comparação Objetiva entre Detectores**

Após a avaliação dos algoritmos de pré-processamento a serem usados na abordagem proposta para geração de representações não foto-realísticas, foi realizada uma avaliação visando a comparação dos resultados objetivos do método proposto nesta dissertação com dois outros algoritmos que geram representações não foto-realísticas: DoG – *Difference of Gaussians* (Marr e Hildreth, 1980) e FDoG – *Flow Based Difference of Gaussians* (Kang, Lee e Chui, 2007). Portanto, as 14 imagens com *ground-truths* anotados foram submetidas a uma avaliação objetiva usando as três métricas descritas anteriormente (PSRN, SSIM e FoM).

Conforme ressaltado no Capítulo 2, o algoritmo DoG é um detector de bordas, criado a partir do detector de Marr-Hildreth (1980). Para prover uma comparação justa entre os vários métodos comparados quando usadas métricas objetivas, foi acrescentado ao DoG uma etapa de pós-processamento igual à etapa de pós-processamento usada na abordagem proposta: (i) ajuste de brilho e contraste; e (ii) transformação histogrâmica. Desta forma, o resultado do algoritmo DoG representa um mapa de bordas

para uma representação não foto-realística, com diferentes gradações de cinza e diferentes espessuras de linhas. No entanto, quando considerado o FDoG, o acréscimo de uma etapa de pós-processamento não se justifica, pois a abordagem possui uma etapa própria de pós-processamento, destinada à limiarização dos níveis de cinza (*gray thresholding*).

Os parâmetros do filtro FDoG usados foram os valores-padrão usados na implementação disponibilizada pelo autor. As dimensões das máscaras gaussianas usadas no filtro DoG foram 15 e 7, tendo como parâmetros de pós-processamento os valores -85 para brilho, 60 para contraste, potência 4 para a transformação histogrâmica e limiar de binarização igual a 170. Já os parâmetros usados na abordagem proposta foram os mesmos descritos no Capítulo 3. Os resultados desta comparação são apresentados na Tabela 4.9, com os maiores valores de médias para cada métrica destacados em **negrito**.

**Tabela 4.9 – Comparação objetiva entre a abordagem proposta, DoG e FDoG.**

Imagem	Abordagem Proposta			DoG			FDoG		
	PSNR	FoM	SSIM	PSNR	FoM	SSIM	PSNR	FoM	SSIM
1	6,80456	0,55920	0,13099	5,88657	0,48473	0,05268	6,89506	0,55504	0,12346
2	7,58620	0,43747	0,19100	7,28366	0,36588	0,08521	8,77689	0,44807	0,19502
3	7,04561	0,67064	0,12974	6,37078	0,60401	0,09232	6,89338	0,62521	0,12372
4	8,29647	0,41807	0,18630	7,36675	0,31648	0,12890	9,14672	0,37622	0,18253
5	7,63174	0,41328	0,17066	6,56296	0,27251	0,07653	9,00917	0,34951	0,15691
6	8,73936	0,51003	0,17201	7,23620	0,43379	0,11399	9,47370	0,53852	0,16169
7	8,38595	0,58043	0,17258	7,17260	0,49141	0,11737	8,39269	0,56737	0,17053
8	6,15994	0,61327	0,13577	5,91636	0,55331	0,06518	6,88630	0,64293	0,13768
9	8,75645	0,54110	0,20402	8,15403	0,41381	0,15025	9,96172	0,48834	0,20802
10	9,02967	0,48561	0,21575	8,54648	0,36812	0,15850	9,37622	0,40923	0,19329
11	9,00363	0,52690	0,20289	8,35591	0,48046	0,14881	9,29177	0,53914	0,20465
12	8,61059	0,54250	0,17696	7,72499	0,52881	0,13335	9,89073	0,69292	0,19278
13	5,68331	0,50913	0,12783	5,30347	0,42346	0,03773	7,34480	0,58692	0,12904
14	11,10500	0,47189	0,25170	9,34934	0,28045	0,14630	10,80090	0,32798	0,23230
$\bar{x}$	8,05989	<b>0,51997</b>	<b>0,17630</b>	7,23072	0,42980	0,10765	<b>8,72429</b>	0,51053	0,17226
s	1,38082	0,07332	0,03649	1,14493	0,10110	0,03947	1,26687	0,11386	0,03461

Mais uma vez, a fim de elucidar a presença de diferenças estatisticamente significativas entre os valores de cada métrica quando considerados os três algoritmos comparados, foram conduzidos três testes F ANOVA fator único, um para cada métrica, tendo sido comparadas as médias obtidas por cada um dos três detectores.

Os testes F ANOVA fator único são apresentados na ordem que segue: (i) Tabela 4.10 para o teste F ANOVA com valores de PSNR; (ii) Tabela 4.11 para o teste F ANOVA com valores de SSIM; e (iii) Tabela 4.12 para o teste F ANOVA com valores de FoM.

**Tabela 4.10 – Teste ANOVA para valores de PSNR.**

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	15,67856	2	7,839282	<b>4,876696</b>	0,012873	<b>3,238096</b>
Dentro dos grupos	62,69245	39	1,607499			
Total	78,37101	41				

**Tabela 4.11 – Teste ANOVA para valores de SSIM.**

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,041548	2	0,020774	<b>15,24669</b>	0,000128	<b>3,238096</b>
Dentro dos grupos	0,053138	39	0,001363			
Total	0,094685	41				

**Tabela 4.12 – Teste ANOVA para valores de FoM.**

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	0,068765	2	0,034382	<b>3,61137</b>	0,036398	<b>3,238096</b>
Dentro dos grupos	0,371303	39	0,009521			
Total	0,440067	41				

Conforme a análise dos três testes F ANOVA fator único para cada uma das métricas, em relação à comparação entre a abordagem proposta, o FDoG e o DoG, constata-se, usando  $\alpha = 0,05$ , que todos os valores F são maiores do que  $F_{crit}$ , o que significa que **há diferenças** estatisticamente significativas entre as médias dos resultados dos três algoritmos comparados, em relação às três métricas usadas.

De acordo com o que foi afirmado na seção anterior, o teste F ANOVA fator único não possibilita comparações entre médias de pares de grupos. Portanto, foi conduzido um teste de Tukey para múltiplas comparações de médias, para cada métrica avaliada. Os resultados do teste com um intervalo de confiança de 95% para os valores apresentados na Tabela 4.9 são apresentados na ordem que segue: (i) Tabela 4.13 para o teste de

Tukey para valores de PSNR; (ii) Tabela 4.14 para o teste de Tukey para valores de SSIM; e (iii) Tabela 4.15 para o teste de Tukey para valores de FoM. Tais tabelas fundamentam a comparação entre múltiplas médias, possibilitando a confrontação da abordagem proposta com os outros algoritmos considerados.

**Tabela 4.13 – Teste de Tukey (*Honestly Significant Differences*) para valores de PSNR.**

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
FDoG vs DoG	1,494	3,117	2,437	0,009	<b>Sim</b>
FDoG vs Abordagem Proposta	0,664	1,386	2,437	0,358	Não
Abordagem Proposta vs DoG	0,829	1,730	2,437	0,207	Não
Valor crítico d de Tukey:			3,446		

**Tabela 4.14 – Teste de Tukey (*Honestly Significant Differences*) para valores de SSIM.**

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Abordagem Proposta vs DoG	0,069	4,921	2,437	< 0,001	<b>Sim</b>
Abordagem Proposta vs FDoG	0,004	0,290	2,437	0,955	Não
FDoG vs DoG	0,065	4,631	2,437	0,000	Sim
Valor crítico d de Tukey:			3,446		

**Tabela 4.15 – Teste de Tukey (*Honestly Significant Differences*) para valores de FoM.**

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Abordagem Proposta vs DoG	0,090	2,445	2,437	0,049	<b>Sim</b>
Abordagem Proposta vs FDoG	0,009	0,256	2,437	0,965	Não
FDoG vs DoG	0,081	2,189	2,437	0,086	Não
Valor crítico d de Tukey:			3,446		

A partir da análise dos testes de Tukey apresentados nas Tabelas 4.13 a 4.15, pode ser verificado que **não há**, estatisticamente, diferença significativa entre a abordagem proposta e o algoritmo FDoG. No entanto, após a realização dos testes de Tukey, considerando-se as três métricas (PSNR, SSIM e FoM), foram constatadas diferenças entre: (i) o DoG e o FDoG (teste com PSNR); (ii) a abordagem proposta e o DoG (teste com SSIM); e (iii) a abordagem proposta e o DoG (teste com FoM).

A comparação numérica dos vários algoritmos destacados anteriormente foi confrontada com a comparação visual, levando-se em conta os maiores e menores resultados obtidos por cada métrica. Tal comparação é apresentada como segue: (i) Tabela 4.13 para a comparação visual da PSNR; (ii) Tabela 4.14 para a comparação visual da SSIM; e (iii) Tabela 4.15 para comparação visual da FoM.

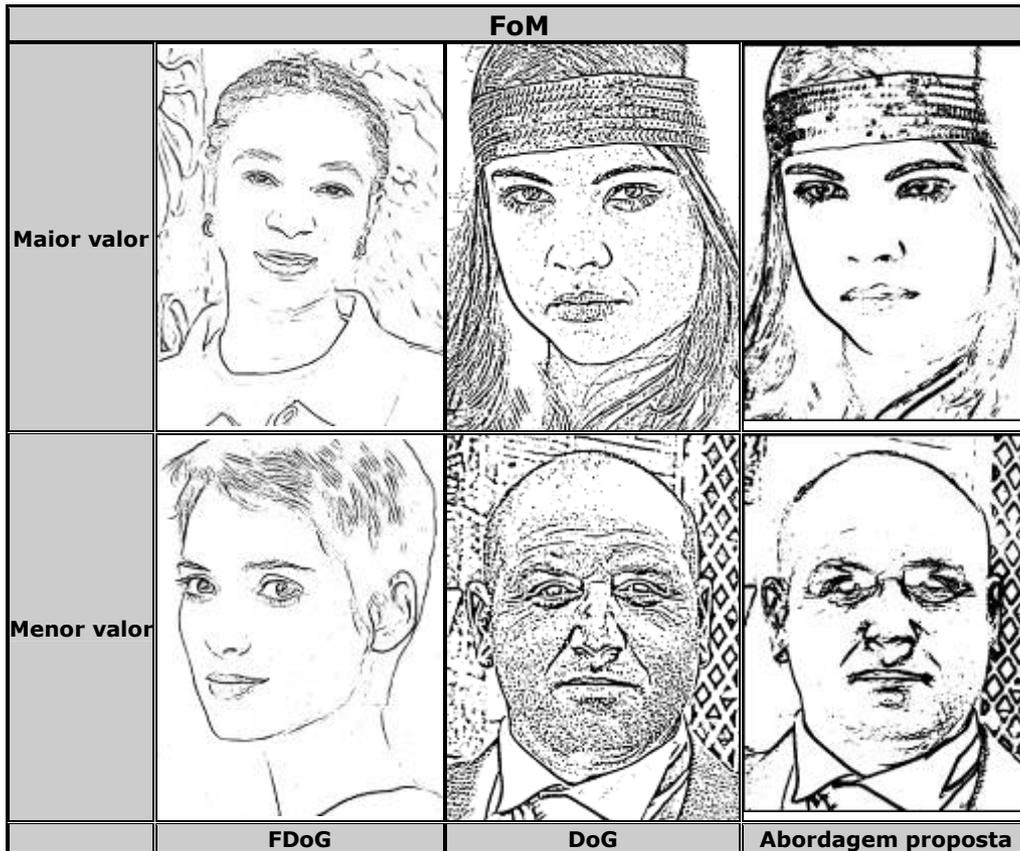
Tabela 4.16 – Comparação visual – PSNR.

PSNR			
Maior valor			
Menor valor			
	FdoG	DoG	Abordagem proposta

Tabela 4.17 – Comparação visual – SSIM.

SSIM			
Maior valor			
Menor valor			
	FDoG	DoG	Abordagem proposta

**Tabela 4.18 – Comparação visual – FoM.**



Observa-se, pela comparação visual, que a presença de ruído na imagem interfere, em maior grau, no baixo valor para cada métrica, mesmo que a resposta visual contenha elementos bem definidos de borda para os aspectos relevantes da face.

No entanto, a avaliação objetiva, no contexto da geração de representações não foto-realísticas, pode não esgotar a variabilidade de nuances de resultados de técnicas que visam o apelo visual, uma vez que, como tal, é inerentemente ambíguo e subjetivo. Tal aspecto não invalida a discussão objetiva realizada anteriormente, mas suscita a busca por um método ou ferramental para a avaliação da técnica proposta no âmbito da geração de representações não foto-realísticas. A avaliação subjetiva surge, portanto, como um segundo meio de avaliação adotado no intuito de sanar tais problemas. Tal avaliação foi conduzida mediante a estruturação de um experimento subjetivo dividido em duas partes, o qual será apresentado na próxima seção.

### 4.3. Experimento Subjetivo

Heath et al. (1997) desenvolveram um método para avaliação relativa de algoritmos de detecção de borda. Os algoritmos testados foram Canny, Nalwa-Binford, Iverson-Zucker, Bergholm e Rothwell. A avaliação foi baseada na medida de desempenho visual, ou seja, indicou a qualidade percebida por um ser humano ao identificar um objeto por meio das bordas geradas por cada detector. O processo de avaliação, após a coleta de um conjunto fixo de imagens, consistiu de duas etapas: (i) a otimização dos parâmetros de cada algoritmo de detecção de bordas; e (ii) a avaliação visual dos resultados. Segundo os autores, o uso de análise visual difere da avaliação objetiva pelo uso de votação com humanos.

No contexto deste trabalho, a abordagem para avaliação subjetiva foi inspirada no trabalho realizado em Heath et al. (1997) e Heath (1996), tendo como principais modificações: (i) o método de calibração dos parâmetros de cada detector avaliado; (ii) a escolha do tipo de imagem a utilizar; (iii) a votação em meio eletrônico (e não impresso); e o uso de algoritmos diferentes para a avaliação (o FDoG, o DoG, o Canny e a abordagem proposta).

O detector de bordas de Canny foi usado de forma bruta (sem pré ou pós-processamento) na comparação subjetiva entre detectores, pelo fato de ser o detector mais usado (DeCarlo e Santella, 2002; Gonzalez e Woods, 2008; Kang, Lee e Chui, 2007), em comparações entre diferentes abordagens.

As imagens usadas no experimento subjetivo estão sob a licença *Creative Commons*<sup>9</sup>. Tal licença, usada pelos autores das fotos no site Flickr<sup>10</sup>, permite que qualquer indivíduo, em qualquer país, saiba claramente que possui o direito de utilizar a obra, de acordo com o subtipo de licença escolhida. As imagens escolhidas e as respectivas *URLs* são apresentadas no Apêndice E.

No experimento conduzido por Heath et al. (1997), as imagens de

---

<sup>9</sup> Disponível em [www.creativecommons.org.br](http://www.creativecommons.org.br)

<sup>10</sup> Site hospedeiro [www.flickr.com](http://www.flickr.com)

avaliação foram divididas em duas categorias, a saber: (i) imagens contendo objetos produzidos por processo de manufatura industrial; e (ii) imagens contendo objetos encontrados na natureza em estado bruto. Tais categorias foram subdivididas em: (i) objetos que possuem superfície rugosa ou com textura; e (ii) objetos que possuem superfície lisa. No total, o número de imagens classificadas foi 28. Inspirado nesta divisão e tendo como foco a avaliação de representações não foto-realísticas de faces humanas, a categorização se deu na forma descrita na Tabela 4.16, totalizando 20 imagens.

**Tabela 4.19 – Categorias de fotos para experimento subjetivo.**

<b>Categoria</b>	<b>Foto <i>close-up</i> 1 face</b>	<b>Foto em grupo Máximo 3 faces</b>
Crianças	5 fotos	5 fotos
Adultos	5 fotos	5 fotos
<b>Total</b>	<b>20 Fotos</b>	

O uso de imagens de crianças e imagens de adultos deveu-se à uniformidade da textura da pele em indivíduos mais jovens. Tal divisão foi orientada pela avaliação do desempenho dos detectores ao gerar uma representação não foto-realística robusta de faces de crianças, bem como pela avaliação do excesso de linhas em representações não foto-realísticas de faces de indivíduos mais idosos.

### **4.3.1. Primeira Parte do Experimento**

Em Heath et al. (1997), a primeira parte do experimento subjetivo consistiu em calibrar com os melhores parâmetros cada detector de borda a ser avaliado. Tal calibração foi feita por um único indivíduo, o pré-selecionador (*pre-screener*), o qual gerou um conjunto de 64 imagens para cada detector avaliado, a partir da variação dos parâmetros de acordo com: (i) recomendações dos autores de cada algoritmo; e (ii) variação linear dos parâmetros. Para cada algoritmo, o pré-selecionador reduziu o conjunto de 64 para 12 imagens, levando em conta, para a supressão de uma imagem do conjunto, o excesso ou a falta de bordas. Após a pré-seleção dos parâmetros, ocorreu uma votação com 9 usuários de teste, a fim de avaliar a conformidade do conjunto de 12 imagens com a capacidade de avaliação

de bordas e reconhecimento de objetos. É conveniente ressaltar que o foco do trabalho de Heath (1996) era a proposição de uma metodologia de avaliação subjetiva de algoritmos de geração de bordas, teoricamente imune a um viés na avaliação de cada detector.

No contexto do presente trabalho, foi proposta uma abordagem para geração de representações não foto-realísticas que se parecem com esboços. Por temer um viés na calibração dos parâmetros para cada um dos métodos comparados, a etapa de pré-calibração foi conduzida sem a presença de um pré-selecionador. Vinte usuários de teste foram convocados e, para cada método avaliado, foram exibidas as 20 imagens de faces. Para cada método, uma variação aleatória de parâmetros foi gerada, resultando em um conjunto de 8 imagens de borda para cada uma das 20 imagens de face. Na primeira etapa do experimento, cada usuário avaliou 640 imagens (8 imagens de borda  $\times$  20 imagens de face  $\times$  4 detectores). Um excerto da interface gráfica utilizada na primeira etapa do experimento subjetivo é mostrada na Figura 4.2.



Figura 4.2 – Etapa 1 do experimento subjetivo: pré-seleção dos parâmetros.

Na primeira etapa do experimento, o usuário deveria escolher no máximo 3 imagens para cada imagem de face(s). As 2 imagens mais votadas, e, conseqüentemente, os 2 melhores conjuntos de parâmetros para cada detector, formaram o conjunto de parâmetros que foram usados na segunda parte do experimento. Os dois melhores parâmetros para cada detector avaliado são apresentados na Tabela 4.17.

Tabela 4.20 – Melhores parâmetros por detector.

<b>Canny</b>		
<b>Parâmetros</b>	<b>Melhores imagens</b>	
	<i>1</i>	<i>2</i>
<i>Limiar máximo</i>	101	102
<i>Limiar mínimo</i>	24	20
<b>DoG</b>		
<b>Parâmetros</b>	<b>Melhores imagens</b>	
	<i>1</i>	<i>2</i>
<i>Kernel 1</i>	15	13
<i>Kernel 2</i>	9	7
<b>FDoG</b>		
<b>Parâmetros</b>	<b>Melhores imagens</b>	
	<i>1</i>	<i>2</i>
<i>Half_w</i>	9	7
<i>M</i>	1	2
$\rho$	3,9937902	3,9180493
$\sigma_1$	1,3616940	1,2509143
$\sigma_2$	4,6370154	4,2813101
<i>Gray Threshold</i>	0,99	0,99
<b>Abordagem Proposta</b>		
<b>Parâmetros</b>	<b>Melhores imagens</b>	
	<i>1</i>	<i>2</i>
Algoritmo de constância cromática	<i>GrayWorld</i>	Expansão
Canal	Y	Y
Tipo de suavização	Gaussiana (5x5)	Mediana (5x5, 1 iteração)
Valor brilho	-80	-78
Valor contraste	80	81
Ordem do operador exponencial	4	3

Observa-se que, com o uso da pré-calibração por meio de votação, os algoritmos de pré-processamento que obtiveram maior número de votos são equivalentes àqueles validados na avaliação objetiva, ou seja, **a análise objetiva não diferiu da análise subjetiva quanto à melhor combinação de algoritmos de pré-processamento**. Adicionalmente,

para efeito de comparação visual, são apresentadas na Figura 4.3 as imagens que obtiveram maior e menor número de **votos não nulos**.

<b>Canny</b>		
<b>DoG</b>		
<b>FDoG</b>		
<b>Abordagem proposta</b>		
	<b>Maior número de votos não nulos</b>	<b>Menor número de votos não nulos</b>

Figura 4.3 – Imagens mais e menos votadas – experimento subjetivo - parte 1.

### 4.3.2. Segunda Parte do Experimento

Em Heath et al. (1997), a segunda parte do experimento subjetivo consistiu na votação das imagens geradas pela calibração de parâmetros efetuada na primeira etapa. Nesta fase, dezesseis usuários de teste avaliaram 280 imagens de borda (2 imagens de borda × 28 imagens × 5 detectores).

A comparação de cada um dos detectores se deu de forma indireta, levando em conta uma pontuação no intervalo [0; 7]. No contexto do trabalho ora proposto, o intervalo utilizado diferiu do intervalo proposto por Heath et al. (1997), tendo sido usado o intervalo [0; 10], sob a forma de uma escala de Likert com semântica diferencial nos extremos do intervalo, no qual 0 (zero) correspondia a MUITO DIFÍCIL e 10 (dez) a MUITO FÁCIL, estando ambas as âncoras semânticas relacionadas à identificação das características faciais das faces utilizadas no experimento.

Similarmente, a segunda parte do experimento subjetivo de comparação representações não foto-realísticas de faces conduzido no âmbito desta dissertação teve a participação de 25 usuários de teste, que avaliaram 10 imagens de face, geradas por 4 métodos, sendo 1 imagem por método (totalizando 40 imagens por avaliador). Durante a realização do teste-piloto, verificou-se que o uso da mesma quantidade de imagens (20) usada na etapa 1 do experimento subjetivo consumiria um tempo proibitivo (cerca de 1 hora e 10 minutos por votante). Com a redução do conjunto de imagens para 10 e após reformulação da interface do experimento, o tempo médio foi reduzido para 25 minutos. Um fragmento da interface usada na segunda parte do experimento é apresentado na Figura 4.4.





Figura 4.4 – Etapa 2 do experimento subjetivo – pontuação.

O código da imagem ou a URL presente na interface do experimento não identificava o tipo de método avaliado. Contrapondo a estratégia de votação adotada por Heath (1996), a avaliação não foi realizada em meio impresso, mas em meio eletrônico. Na Tabela 4.21, estão dispostos os valores de média e desvio padrão do desempenho relativo por cada detector avaliado.

Tabela 4.21 – Desempenho relativo por detector.

Detector	$\bar{X}$	s
FDoG	7,0917	2,0880
Abordagem proposta	5,2375	1,8449
DoG	4,9917	1,9210
Canny	3,8375	2,2301

Novamente, a fim de elucidar a presença de diferenças estatisticamente significativas entre os valores de média do desempenho relativo por método, foi conduzido um teste F ANOVA fator único, tendo sido comparadas as médias obtidas por cada método avaliado. O teste F ANOVA fator único é apresentado na Tabela 4.22.

Tabela 4.22 – Teste ANOVA para o experimento subjetivo – parte 2.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	1307,404	3	435,8014	<b>106,1173</b>	2,65E-59	<b>2,614214</b>
Dentro dos grupos	3926,092	956	4,10679			
Total	5233,496	959				

Após análise do teste F ANOVA fator único, constata-se, usando  $\alpha = 0,05$ , que todos o valor F é maior do que  $F_{crit}$ , o que significa que **há diferenças** estatisticamente significativas entre as médias dos resultados obtidos. De posse de tal constatação, procedeu-se a comparação entre médias de pares de grupos, por meio do teste de Tukey. Os resultados do teste com um intervalo de confiança de 95% para os valores apresentados na Tabela 4.9 são apresentados na Tabela 4.23.

Tabela 4.23 – Teste de Tukey para o experimento subjetivo – parte 2.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
FDoG vs Canny	3,254	17,591	2,569	< 0,0001	<b>Sim</b>
FDoG vs DoG	2,100	11,352	2,569	< 0,0001	<b>Sim</b>
FDoG vs Proposta	1,854	10,023	2,569	< 0,0001	<b>Sim</b>
Proposta vs Canny	1,400	7,568	2,569	< 0,0001	<b>Sim</b>
Proposta vs DoG	0,246	1,329	2,569	0,544	Não
DoG vs Canny	1,154	6,239	2,569	< 0,0001	<b>Sim</b>
<b>Valor crítico d de Tukey:</b>			3,633		

A partir da análise do teste de Tukey apresentado na Tabela 4.23, verifica-se que **há**, estatisticamente, diferença significativa entre o FDoG e as outras abordagens, quando considerado o intervalo de confiança de 95%. Observa-se, portanto, que o FDoG apresenta a maior média e esta é significativamente diferente de todas as outras. Observa-se, adicionalmente, que **não há** diferenças estatisticamente significativas entre a abordagem proposta e o método DoG.

O fato de o FDoG ter apresentado uma média maior em relação à abordagem proposta (7,09 versus 5,23) reflete a possibilidade de realização de seu aprimoramento, tanto no sentido da redução de artefatos espúrios quanto no refinamento das bordas da imagem de saída, o que ocasionará uma resposta visual mais satisfatória. Adicionalmente, as imagens que obtiveram o melhor e pior desempenho na avaliação subjetiva são

apresentadas na Figura 4.5.

<b>Canny</b>		
<b>DoG</b>		
<b>FDoG</b>		
<b>Abordagem Proposta</b>		
	<b>Imagem com maior média</b>	<b>Imagem com menor média</b>

Figura 4.5 – Comparação visual do melhor desempenho por detector.

## 4.4. Considerações Finais do Capítulo

Neste capítulo, foi descrita a parte experimental do trabalho desenvolvido no âmbito desta dissertação. Tal parte experimental teve como propósito a complementação do processo de validação da abordagem proposta e

descrita no Capítulo 3. Após vários experimentos conduzidos em etapas distintas do estudo, constatou-se que:

- A aquisição de conhecimento para treinamento com base em uma anotação manual por um humano não obteve melhores resultados do que a aquisição por meio de exemplos sintéticos. Tal fato pode ser explicado pela falta de controle na separação dos dois conjuntos de classificação (borda e não-borda). Tal constatação foi apresentada por meio de uma análise de curvas ROC.
- A disponibilidade de várias métricas para a avaliação de bordas na literatura especializada não exclui o fato de que ressalvas sejam necessárias quando de seu uso. A métrica mais popular, a FoM, indicou, após uma análise estatística, fatos relevantes para a presente proposta: (i) qual o melhor conjunto de algoritmos de pré-processamento a ser utilizado; e (ii) se há diferenças significativas entre a abordagem proposta e outras abordagens tidas como estado da arte na área de geração de representações não foto-realísticas. Segundo os resultados, o melhor conjunto de algoritmos engloba *GrayWorld*, conversão para o canal Y do espaço YCbCr e suavização gaussiana. Não foi verificada, estatisticamente, diferença significativa entre a abordagem proposta e o algoritmo FDoG.
- A realização de um experimento subjetivo objetivou a complementação do processo de validação da abordagem proposta, assim como a coleta de dados que possibilitaram uma análise do processo, do ponto de vista do usuário de teste, a qual se mostrou complementar da análise objetiva previamente conduzida. A calibração de parâmetros para os quatro detectores de bordas considerados (o Canny, o DoG, o FDoG e a abordagem proposta) fortaleceu as conclusões obtidas a partir da avaliação objetiva. A abordagem proposta apresentou, na análise subjetiva, uma média menor em relação ao FDoG, tendo sido superior, em média, ao DoG e ao algoritmo Canny. Tal fato remete a possibilidade de realização de melhorias na abordagem proposta para atingir uma resposta visual aprimorada.
- O algoritmo de Canny mostrou-se pouco adequado à geração de representações não foto-realísticas com avaliação centrada em faces

humanas, tendo em vista o baixo valor de média obtido com a avaliação subjetiva (média 3,84).

- Após a realização de um teste de carga que visou a comparação de desempenho entre a abordagem proposta e o FDoG, foi constatado, para 20 imagens de dimensões variadas, o que segue: o tempo médio da abordagem proposta, quando comparado ao FDoG, foi *equivalente*. O desempenho computacional foi, em média, 3,5 segundos de processamento. O teste foi realizado em um computador com processador Intel Core 2 Duo, 2.0 GHz, 2GB RAM. As implementações algorítmicas avaliadas não contemplaram o uso de dois núcleos para processamento paralelo.

# Capítulo 5

## Considerações Finais

Neste capítulo, são apresentados: (i) um sumário dos principais pontos abordados nesta dissertação; (ii) as contribuições resultantes da pesquisa desenvolvida; e (iii) a formulação de propostas para trabalhos futuros, a qual leva em conta as dificuldades e os problemas encontrados durante o desenrolar da pesquisa.

### 5.1. Sumário da Dissertação

A motivação para o desenvolvimento deste trabalho surgiu da lacuna revelada na literatura relacionada à viabilidade de aprendizagem de máquina para a geração de representações não foto-realísticas. A motivação fundamentou-se na carência de abordagens bem documentadas, na literatura da área, que utilizem aprendizagem de máquina voltada para a geração de representações não foto-realísticas, bem como na aplicação de aprendizagem de máquina para o aprimoramento ou redução de artefatos indesejáveis nos resultados finais das representações não foto-realísticas geradas.

Dentre as aplicações de NPR, foram destacadas nesta pesquisa a transformação automática de fotos (ou quadros de vídeo) em representações que imitam desenhos ou pinturas, com fins ao apoio a desenhistas de histórias em quadrinhos, à geração de caricaturas com fins de entretenimento e apelo publicitário, à transferência e automatização de estilos de desenho, a técnicas de visualização alternativa, a aplicações na geração de desenhos animados (*cartoons*), pinturas artísticas, desenhos técnicos, comunicação publicitária, visualização de desenhos científicos.

Adicionalmente, a geração de representações não foto-realísticas que

lembram esboços está intimamente ligada à detecção de bordas, método importante para a extração de características, segmentação e reconhecimento, análise de imagens e amplamente usada em vários algoritmos de visão computacional.

A abordagem proposta para a geração de representações não foto-realísticas é constituída de módulos para a compensação de iluminação por meio de uso de algoritmos de constância cromática, a conversão de espaço cromático, a suavização, a detecção de bordas com o uso de redes neurais supervisionadas, a adequação de brilho e contraste e a transformação da faixa dinâmica a partir de um operador exponencial.

O sistema computacional desenvolvido com base na abordagem proposta foi submetido a experimentos para a avaliação de sua viabilidade. Além disto, o resultado do sistema proposto foi comparado com o desempenho de outros algoritmos destinados à geração de representações não foto-realísticas que se assemelham a esboços. Considerações a respeito da realização de tais experimentos são apresentadas na próxima seção.

## **5.2. Contribuições do Trabalho Realizado**

A partir dos resultados obtidos nos experimentos realizados no Capítulo 4, é possível concluir que objetivo de desenvolver uma abordagem automática para geração de representações não foto-realísticas bidimensionais, a partir de fotografias digitais, foi alcançado. Para validar a técnica, um estudo de caso foi conduzido em fotografias de faces humanas frontais. A técnica proposta tem como diferencial a agregação de módulos para atenuar os efeitos advindos da falta de controle das condições de iluminação ao capturar a imagem e o uso de redes neurais, no contexto da aprendizagem de máquina, destinadas à detecção de bordas. A implementação do sistema proposto atingiu o critério de sucesso esperado. Assim, as principais contribuições da pesquisa podem ser enumeradas como segue:

1. A geração de um compêndio mínimo das principais técnicas usadas na avaliação de detectores de borda;
2. A constatação experimental de benefícios obtidos com a utilização de

- técnicas de constância cromática e algoritmos de suavização para a abstração de representações não foto-realísticas;
3. Uma comparação objetiva, baseada em resultados experimentais, que possibilitou a confrontação da abordagem proposta com o FDoG, revelando a inexistência de diferenças estatisticamente relevantes entre ambas;
  4. A constatação experimental de algumas limitações inerentes às métricas usadas para a avaliação objetiva da detecção de bordas e a compensação de tais limitações, a partir da proposição de experimentos de cunho subjetivo;
  5. O teste dos módulos a partir de imagens reais, evidenciando o potencial da abordagem proposta na área de entretenimento, dentre outras; e
  6. A proposição de um fluxo de execução que usa aprendizagem de máquina, por meio de redes neurais, para a tarefa específica de geração de representações não foto-realísticas.

A seguir, são apresentadas sugestões para a melhoria e/ou para a complementação dos resultados obtidos nesta pesquisa.

### **5.3. Trabalhos Futuros**

Durante a realização dos experimentos para a definição do melhor conjunto de algoritmos a serem usados na abordagem proposta, foi perceptível a degradação da imagem ao aplicar o filtro Retinex. Tal fato é merecedor de uma investigação mais detalhada, pois, segundo bibliografia pesquisada, o algoritmo Retinex possui grande poder para agregar constância cromática à cena. Todavia, devido à sua complexidade, os valores atribuídos aos parâmetros envolvidos merecem um estudo mais aprofundado para melhor calibração, e.g., a partir do uso de algoritmos genéticos.

Adicionalmente, outra linha de investigação poderá explorar a inclusão de um filtro homomórfico (Parker, 1997), adequado à homogeneização de iluminação de uma cena, ao conjunto de algoritmos disponíveis para compensação de iluminação. O mesmo vale para a inclusão de outras

técnicas de suavização de imagens, e.g., do filtro bilateral e/ou de outras de suas variantes.

Quanto ao paradigma de aprendizagem de máquina adotado, as redes neurais tipo *perceptron* multicamadas, constatou-se boa velocidade na fase de uso da rede. No entanto, tal paradigma não é o único disponível. Alguns testes foram realizados no que concerne ao uso de redes do tipo Hopfield, mas não foram obtidos resultados satisfatórios. Outra abordagem para uso de aprendizagem de máquina para a geração de representações não foto-realísticas é a adoção de um paradigma diferente daquele adotado nesta pesquisa para a aquisição do conhecimento como, por exemplo, o uso de uma abordagem não supervisionada.

No que concerne ao treinamento de redes neurais, os classificadores treinados ao longo do trabalho usaram apenas máscaras de tamanho fixo 5x5. De acordo com a revisão bibliográfica realizada, alguns autores usam máscaras de tamanho 3x3, o que inspira outra linha de investigação, com fins ao refinamento da técnica proposta. Adicionalmente, algumas bordas aparecem excessivamente espessas no resultado final gerado, suscitando uma investigação que poderá envolver o afinamento das bordas ou o aumento das dimensões da máscara de filtragem.

Ainda no que concerne o treinamento de redes neurais, a abordagem que tenta usar a transferência de conhecimento da geração de representações não foto-realísticas por um humano para uma representação tipo rede neural falhou. No entanto, o acréscimo vertiginoso de exemplos desse conhecimento certamente traria mais generalidade ao classificador treinado, concomitantemente à separação bem definida de grupos de borda e não-borda para a produção de melhores resultados pelo classificador.

## Referências Bibliográficas

- AGRAWALA, M. e STOLTE, C. Rendering effective route maps: improving usability through generalization. In *SIGGRAPH '01: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM press, p. 241-249, 2001.
- ALSMADI, M. K. S., OMAR, H. S. e NOAH, S. A. Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm. In: *IJCSNS - International Journal of Computer Science and Network Security*, (9)4, Abril, 2009.
- BARLA, P., BRESLAV. S., MARKOSIAN, L., e THOLLOT, J. Interactive hatching and stippling by example. *Rapport de Recherche*, INRIA, 2006.
- BARNARD, K. e FUNT, B. Investigations into multi-scale retinex. In *Colour Imaging: Vision and Technology*. John Wiley and Sons, p. 9-17, 1999.
- BARNARD, K., CARDEI, V., e FUNT B. A Comparison of Computational Color Constancy Algorithms. In *IEEE Transactions on Image Processing*, 11(9), p. 972-983, 2002.
- BECERIKLI, Y. e DEMIRAY, H. E. Alternative Neural Network Based Edge Detection. In *Neural Information Processing - Letters and Reviews*, Vol. 10, n. 8-9, 2006.
- BOUSSEAU, A., KAPLAN, M., THOLLOT, J., e SILLION, F. X. Interactive watercolor rendering with temporal coherence and abstraction. In *NPAC '06: Proceedings of the 4th International Symposium on Non-Photorealistic Animation and Rendering*, New York, NY, USA. ACM Press, p. 141-149, 2006
- BOVIK, A. *Handbook of image and video processing*. Elsevier, 2005.
- BUSCHSBAUM, G. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310 (1), 337-350, 1980.
- CARPENDALE, M. S. T. *A Framework for Elastic Presentation Space*. Ph.D. thesis, School of Computer Science, Simon Fraser University, 1999.
- CASTRO, A. P. A., e SILVA, J. D. S. Redes Neurais Supervisionadas para Detecção de Bordas. In *Anais do 4º Workshop dos Cursos de Computação Aplicada do INPE*. São José dos Campos, São Paulo, 2004.
- CHABRIER, S., LAURENT, H., ROSENBERGER, C., e EMILE, B. Comparative study of contour detection evaluation criteria based on dissimilarity measures. *Journal of Image and Video Processing*, p. 1-13, 2008.
- CHANG, C-Y. A contextual-based Hopfield neural network for medical image edge detection. In *IEEE International Conference on Multimedia and Expo*, p. 1011-1014, 2004.
- CHEN, H., LIANG, L., XU, Y., SHUM, H., ZHU, S., e ZHENG, N. Example-

- based Automatic Portraiture. In *ACCV2002: Proceedings of the 5<sup>th</sup> Asian Conference on Computer Vision*, Melbourne, Australia, 2002a.
- CHEN, H., LIU, Z., ROSE, C., XU, Y., SHUM, H., e SALESIN, D. Example-Based Composite Sketching of Human Portraits. In *NPAR'04: Proceedings of the 3<sup>rd</sup> International Symposium on Non-Photorealistic Animation and Rendering*, Annecy, France, p. 95-153, 2004.
- CHEN, H., XU, Y., SHUM, H., ZHU, S., e ZHENG, N. Example-based facial sketch generation with non-parametric sampling. In *Proceedings of the 8<sup>th</sup> IEEE International Conference on Computer Vision*, Vancouver, Canada, p. 433-438, 2001.
- CHEN, H., ZHENG, N., LIANG, L., LI, Y., XU, Y., e SHUM, H. PicToon: a personalized image-based cartoon system. In *Proceedings of the ACM Multimedia*, Juan Les Pins, France, p. 171-178, 2002b.
- COOTES, T. F., TAYLOR, C. J., COOPER, D. H., e GRAHAM, J. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1), p. 38-59, 1995.
- CURTIS, C. J., ANDERSON, S. E., SEIMS, J. E., FLEISCHER, K. W., e SALESIN, D. H. Computer-generated watercolor. In *SIGGRAPH '97: Proceedings of the 24<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., p. 421-430, 1997.
- DECARLO, D., e SANTELLA, A. Stylization and abstraction of photographs. In *Proceedings of the 29<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 769-776, 2002.
- DEUSSEN, O., HILLER, S., VAN OVERVELD, C. W. A. M., e STROTHOTTE, T. Computer-Generated Stipple Drawings. In *Proceedings of VMV'99, Workshop on Vision, Modelling and Visualization*, Erlangen, Germany. 1999.
- DI BLASI, G., e GALLO, G. Artificial mosaics. *The Visual Computer*, 21, p. 373-383, 2005.
- DOBASHI, Y., HAGA, T., JOHAN, H., e NISHITA, T. A method for creating mosaic images using Voronoi diagrams. In *Proceedings of Eurographics Short Presentations*, p. 341-348, 2002.
- DOOLEY, D. L., e COHEN, M. F. Automatic Illustration of 3D Geometric Models: Lines. In *Proceedings of 1990 Symposium on Interactive 3D Graphics*, Snowbird, UT. New York. ACM Press, p. 77-82, 1990.
- EBNER, M. *Color Constancy*. John Wiley & Sons, 416 páginas. ISBN-13: 978-0470058299. Junho, 2007.
- FAUSTINO, G. M., e DE FIGUEIREDO, L. H. Simple adaptive mosaic effects. In *SIBGRAPI '05: Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing*, 2005.
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*,

- 27, p. 861-874, 2006.
- FEKETE, J.-D., BIZOUARN, E., COURNARIE, E., GALAS, T., e TAILLEFER, F. TicTacToon: A paperless system for professional 2D animation. In *SIGGRAPH '95: Proceedings of the 22<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, CA, USA, p. 79–90, 1995.
- FINLAYSON, G. D. Color in perspective. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 18(10), 134-138, 1996.
- FOLEY, J. D., DAM, A. V., FEINER, S. K., e HUGHES, J. F. *Computer Graphics: Principle and Practice*, Addison-Wesley, 1990.
- GILCHRIST, A. L., KOSSYFIDIS, C., BONATO, F., AGOSTINI, T., CATALIOTTI, J., LI, X., SPEHAR, B., ANNAN, V., e ECONOMOU, E. An anchoring theory of lightness perception. *Psychological Review*, 106, p. 795–834, 1999.
- GOMES, H. M. *Model Learning in Iconic Vision*. PhD thesis, University of Edinburgh, Edinburgh, Scotland, 2002.
- GONZALEZ, R. C. e WOODS, R. E. *Digital Image Processing*. Prentice Hall, 976 páginas. ISBN: 013168728X. 2008.
- GOOCH, B. *Artistic vision: automatic digital painting using computer vision algorithms*. 2001. 48 p. Dissertação (Mestrado em Ciência da Computação) – School of Computing, The University of Utah.
- GOOCH, B., e GOOCH, A. *Non-Photorealistic Rendering*. A. K. Peters, 2001.
- GOOCH, B., REINHARD, E., e GOOCH, A. Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics*, 23(1):27-44, 2004.
- GUPTA, L., e SUKHENDU, D. Texture edge detection using multi-resolution features and SOM. In *Proceedings of the 18<sup>th</sup> International Conference on Pattern Recognition*, Washington, DC, USA, IEEE Computer Society Press, p. 199-202, 2006.
- HAEBERLI, P. Paint by numbers: abstract image representations. In *SIGGRAPH '90: Proceedings of the 17<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 207–214, 1990.
- HAN, J. e KAMBER, M. *Data mining: concepts and techniques*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2001.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2<sup>nd</sup> Edition. Prentice Hall, 1998.
- HEATH, M. D. *A robust visual method for assessing the relative performance of edge detection algorithms*. Master's Thesis. Department of Computer Science and Engineering, College of Engineering, University of South Florida. 1996.
- HEATH, M. D., SARKAR, S., SANOCKI, T., e BOWYER, K. W. A Robust Visual

- Method for Assessing the Relative Performance of Edge-Detection Algorithms. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 19, 12, 1338-1359, 1997.
- HERTZMANN, A. Introduction to 3D non-photorealistic rendering: silhouettes and outlines. In *SIGGRAPH '99: Course Notes of the 26<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, 1999.
- HOOVER, G. JEAN-BAPTISTE, D. GOLDFOG, e K. BOWYER. A methodology for evaluating range image segmentation techniques. *Technical Report*. 95-01, Department of Computer Science. University of South Florida, 1995.
- HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. In: *Proceedings of the National Academy of the Sciences of the U.S.A.* 79, pp. 2554-2558, 1982.
- HWANG, W.-I., LEE, P.-J, CHUN, B.-K., RYU, D.-S. e CHO, H.-G., Cinema Comics: cartoon generation from video stream. *Computer Graphics Theory and Applications*, 2006.
- JAIN, R. *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*. Wiley & Sons, 1991.
- JIANG, X., MARTI, C., IRNIGER, C., e BUNKE, H. Distance measures for image segmentation evaluation. *EURASIP Journal on Advances in Signal Processing*, p. 209, 2006.
- JUDD, D. B. Appraisal of Land's work on two-primary color projections. *Journal of the optical Society of America*, 50(3), 254-268, 1960.
- KALNINS, R. D., MARKOSIAN, L., MEIER, B. J., KOWALSKI, M. A., LEE, J. C., DAVIDSON, P. L., WEBB, M., HUGHES, J. F., e FINKELSTEIN, A. WYSIWYG NPR: Drawing strokes directly on 3D models. In *Proceedings of the 29<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 755-762, 2002.
- KANG, H. W., HE, W., CHUI, C. K., e CHAKRABORTY, U. K. Interactive sketch generation. *The Visual Computer*, 21(8-10):821-830, 2005.
- KANG, H., LEE, S., e CHUI, C. K. Coherent line drawing. In *Proceedings of the 5<sup>th</sup> International Symposium on Non-Photorealistic Animation and Rendering*. San Diego, California, 2007.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. In: *Neurocomputing: Foundations of Research*, J. A. Anderson and E. Rosenfeld, Eds. MIT Press, Cambridge, MA, 509-521, 1988.
- KOLLIPOULOS, A., WANG, J. M., e HERTZMANN, A. Segmentation-based 3D artistic rendering. In *17<sup>th</sup> Eurographics Workshop on Rendering Techniques*, p. 361-370, 2006.
- KYPRIANIDIS, J. E., e DÖLLNER, J. Image Abstraction by Structure Adaptive Filtering. In *Proceedings EG UK Theory and Practice of Computer Graphics*, p. 51-58, 2008.

- LAKE, A., MARSHALL, C., HARRIS, M., e BLACKSTEIN, M. Stylized rendering techniques for scalable real-time 3d animation. In *NPAR '00: Proceedings of the 1<sup>st</sup> International Symposium on Non-Photorealistic Animation and Rendering*, New York, NY, USA. ACM Press, p. 13-20, 2000.
- LAND, E., e MCCANN, J. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61(1), 1-11, 1971.
- LANSDOWN, J., e SCHOFIELD, S. Expressive rendering: A review of non-photorealistic techniques. *IEEE Computer Graphics and Applications*, 15(3):29-37, 1995.
- LEE, H., KWON, S., e LEE, S. Real-time pencil rendering. In *Proceedings of the 4<sup>th</sup> International Symposium on Non-photorealistic Animation and Rendering*, New York, NY, USA. ACM Press, p. 37-45, 2006.
- LEE, M. K., LEUNG, S. W., PUN, T. L., CHEUNG, H. L., e LEE, A. M. K. Edge detection by genetic algorithm. In *Proceedings of the 2000 International Conference on Image Processing*, 1:478-480, 10-13, Set 2000.
- LEE, T.-Y., YAN, C.-R., e CHI, M.-T. Stylized rendering for anatomic visualization. *Journal on Computing in Science and Engineering*, 9(1):13-19, 2007.
- LEVINE, D. M., BERENSON, M. L., e STEPHAN, D. *Estatística: Teoria e aplicações*. LTC Editora. Rio de Janeiro. ISBN: 85-216-1211-7.
- LEWIS, J. P., FONG, N., XUEXIANG, X., SOON, S. H., e FENG, T. More optimal strokes for NPR sketching. In *Proceedings of the 3<sup>rd</sup> Annual Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*. ACM Press, p. 47-50, 2005.
- LITWINOWICZ, P. Processing images and video for an impressionist effect. In *SIGGRAPH '97: Proceedings of the 24<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co., p. 407-414, 1997.
- LUFT, T., e DEUSSEN, O. Non-photorealistic real-time rendering of characteristic faces. In *Proceedings of the 12<sup>th</sup> Pacific Conference on Computer Graphics and Applications*, Washington, DC, USA. IEEE Computer Society, p. 339-347, 2004.
- MARKOSIAN, L. *Art-based Modeling and Rendering for Computer Graphics*. PhD thesis, Brown University, 2000.
- MARKOSIAN, L., KOWALSKI, M. A., TRYCHIN, S. J., BOURDEV, L. D., GOLDSTEIN, D., e HUGHES, J. F. Real-time non-photorealistic rendering. In *SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, California, p. 415-420, 1997.
- MARR, D., e HILDRETH, E. Theory of Edge Detection. In: *Proceedings of the Royal Society London*, 207, 187-217, 1980.
- MARTIN, D., FOWLKES, C., TAL, D., e MALIK, J. A database of human

- segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8<sup>th</sup> IEEE International Conference on Computer Vision*, 2(2), p. 416-423, 2001.
- MEER, P. e GEORGESCU, B. Edge detection with embedded confidence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1315-1365, 2001.
- MEIER, B. Painterly rendering for animation. In *SIGGRAPH '96: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, Louisiana, USA. ACM Press, p. 477-484, 1996.
- MIGNOTTE, M. Unsupervised statistical sketching for non-photorealistic rendering models. In *ICIP 2003: Proceedings of the International Conference on Image Processing*, 3:III-573-6 vol.2, Set 2003.
- MILANI, F. e CAZELLA, S. C. Um Modelo para Determinar a Autoridade de Usuários em Sistemas de Recomendação. In *III Fórum de Inteligência Artificial*, Canoas, 2005.
- MOULD, D. A stained glass image filter. In *EGRW '03: Proceedings of the 14<sup>th</sup> Eurographics Workshop on Rendering*, Aire-la-Ville, Switzerland. Eurographics Association, p. 20-25, 2003.
- NG, G.S., ONG, F. L., e NOOR, N. M. Neural Edge Detector. *ASEAN Journal on Science and Technology for Development*. Standards and Industrial Research Institute of Malaysia, 12(1), p. 35-42, 1995.
- NISSEN, C. Neural networks made simple. *Software 2.0*, 2:14-19, 2005.
- NORTHRUP, J. D., e MARKOSIAN, L. Artistic silhouettes: a hybrid approach. In *Proceedings of the 1<sup>st</sup> International Symposium on Non-photorealistic Animation and Rendering*, New York, NY, USA. ACM Press, p. 31-37, 2000.
- OPENCV. Intel® Open Computer Vision Library. Disponível em: <<http://sourceforge.net/projects/opencvlibrary/>>, 2006. Acesso em 01/01/2008.
- PALMER, P. L., DABIS, H., e KITTLER, J. A performance measure for boundary detection algorithms. *Computer Vision and Image Understanding*, 63(3), p.476-494, 1996.
- PARADELLA, W. R. Introdução ao processamento digital de imagens de sensores remotos aplicados a geologia. *Relatório Técnico*. São José dos Campos: INPE, 45p. (INPE-5023-RPE/616), 1990.
- PARKER, J. R. *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons, 1997.
- PIAN, Z., GAO, L., WANG, K., GUO, L., e WU, J. Edge Enhancement Post-processing Using Hopfield Neural Net. In *Proceedings of the 4th International Symposium on Neural Networks: Advances in Neural Networks, Part III*, Nanjing, China. Springer-Verlag, p. 846-852, 2007.
- POYTON, C. *Digital Video and HDTV Algorithms and Interfaces*. Morgan

- Kaufmann Publishers, San Francisco, California. 2003.
- PRATT, W. K. *Digital Image Processing*. John Wiley & Sons, New York, 1978.
- PRIETO, M. S., e ALLEN, A. R. A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, n. 25, p. 1265-1273, 2003.
- RAHMAN, Z., JOBSON D. J. E WOODSELL, G. A. Method of improving a digital image. *United States Patent N. 5.991.456*, 1999.
- RAJAB, M. I., WOOLFSON, M.S., e MORGAN, S. P. Application of region-based segmentation and neural network edge detection to skin lesions. *Computerized medical imaging and graphics*. 28(1-2), 2004.
- RAJENDRAN, R. K., CHANG, S.-F. Image retrieval with sketches and compositions. In *IEEE International Conference on Multimedia and Expo*, 2:717-720 vol.2, 2000.
- RIZZI, A., GATTA, C., e MARINI, D. From Retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization. *Journal of Electronic Imaging*, 13(1), 75-8, 2004.
- SAITO, T., e TAKAHASHI, T. Comprehensible rendering of 3D shapes. In *SIGGRAPH '90: Proceedings of the 17<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, Dallas, TX, USA. ACM Press, p. 197-206, 1990.
- SALISBURY, M. P., ANDERSON, S. E., BARZEL, R., e SALESIN, D. H. Interactive pen-and-ink illustration. In *SIGGRAPH '94: Proceedings of the 21<sup>st</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 101-108, 1994.
- SANTELLA, A. *The Art of Seeing: Visual Perception in Design and Evaluation of Non-Photorealistic Rendering*. PhD thesis, Rutgers University, New Brunswick, New Jersey, USA, 2005.
- SASADA, T. T. Drawing Natural Scenery by Computer Graphics. *Computer-Aided Design*, 19(4):212-218, 1987.
- SAYEED, R., e HOWARD, T. State-of-the-art of non-photorealistic rendering for visualisation. In *Theory and Practice of Computer Graphics*, M. McDerby, L. Lever, p. 1-10, 2006.
- SHARMA, G. *Digital Color Imaging Handbook*. CRC Press, 2002.
- SHET, R., LAI, K., EDIRISINGHE, E., e CHUNG, P. Use of neural networks in automatic caricature generation: An approach based on drawing style capture. In *Iberian Conference on Pattern Recognition and Image Analysis*, p. 343, 2005.
- SOUSA, M. C., e BUCHANAN, J. W. Computer-generated graphite pencil rendering of 3D polygonal models. *Computer Graphics Forum*, 18, 3, p. 195-208, 1999.
- SOUSA, M. C., e BUCHANAN, J. W. Observational models of graphite pencil materials. *Computer Graphics Forum*, 19, 1, p. 27-49, 2000.

- STRASSMANN, S. Hairy brushes. In *SIGGRAPH '86: Proceedings of the 13<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 225–232, 1986.
- STROTHOTTE, T., e SCHLECHTWEG, S. *Non-photorealistic computer graphics: modeling, rendering and animation*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
- SUZUKI, K., HORIBA, I., e SUGIE N. Edge detection from noisy images using a neural edge detector. In *Proceedings of the 2000 IEEE Signal Processing Society Workshop, X Neural Networks for Signal Processing*. 2:487-496, 2000.
- TANG, X., e WANG, X. Face sketch synthesis and recognition. In *Proceedings of the 9<sup>th</sup> IEEE International Conference on Computer Vision*, Washington, DC, USA. IEEE Computer Society, p. 687, 2003.
- TATEOSIAN, L. G., e HEALEY, C. G. NPR: Art enhancing computer graphics. *Technical Report*, Knowledge Discovery Lab, Department of Computer Science, North Carolina State University, 2003.
- TOIVANEN, P. J., ANSAMAKI, J., PARKKINEN, J. P. S., e MIELIKAINEN, J. Edge detection in multispectral images using the Self-Organizing Map. *Pattern Recognition Letters*, 24(16):2987-2994, 2003.
- TRESSET, P., e FOL LEYMARIE, F. Generative portrait sketching. In *Proceedings of the 11<sup>th</sup> International Conference on Virtual Systems and MultiMedia*, Digital Studios, Goldsmiths College, University of London, U.K. Hal Twaites, p. 739-748, 2005.
- WANG, J., XU, Y., SHUM, H.-Y., e COHEN, M. F. Video tooning. In *SIGGRAPH '04: Proceedings of the 13<sup>th</sup> Annual Conference on Computer Graphics and Interactive Techniques*, p. 574-583, 2004.
- WANG, K., GAO, L., PIAN, Z., GUO, L., e WU, J. Edge detection combined entropy threshold and self-organizing map (SOM). In *ISNN '07: Proceedings of the 4<sup>th</sup> International Symposium on Neural Networks*, Berlin, Heidelberg. Springer-Verlag, p. 931-937, 2007.
- WANG, S., GE, F., e LIU, T. Evaluating edge detection through boundary detection. *EURASIP Journal on Advances in Signal Processing*, p. 213-213, 2006.
- WANG, Z. e SIMONCELLI, E. P. An adaptive linear system framework for image distortion analysis. *IEEE International Conference on Image Processing*, Genoa, Italy, 2005.
- WANG, Z., LU, L. e BOVIK, A. C. A universal image quality index. *IEEE Signal Processing Letters*, 9(3), p. 81-84, 2002.
- WEN, F., LUAN, Q., LIANG, L., XU, Y.-Q., e SHUM, H.-Y. Color sketch generation. In *Proceedings of the 4<sup>th</sup> International Symposium on Non-photorealistic Animation and Rendering*, New York, NY, USA, ACM Press, p. 47-54, 2006.
- WINKENBACH, G., e SALESIN, D. H. Computer-generated pen-and-ink

illustration. In *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA. ACM Press, p. 91–100, 1994.

WINKENBACH, G., e SALESIN, D. H. Rendering parametric surfaces in pen and ink. In *SIGGRAPH '96: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, New York, NY, USA, ACM Press, p. 469–476, 1996.

WINNEMÖLLER, H. *Perceptually-motivated non-photorealistic graphics*. PhD thesis, Evanston, IL, USA, 2006.

ZIOU, D., e TABBONE, S. Edge detection techniques - an overview. *International Journal of Pattern Recognition and Image Analysis*, 8:537-559, 1998.

ZOU, K.H.. Receiver operating characteristic (ROC) literature research, 2002. Disponível em: <<http://splweb.bwh.harvard.edu:8000/pages/ppl/zou/roc.html>>. Acesso em: 05 agosto 2008.

# Apêndice A

## Algumas Técnicas de PDI Adotadas Nesta Pesquisa

Este apêndice contém fundamentos das técnicas de processamento digital de imagens (PDI) que foram adotadas em duas das etapas da abordagem proposta para geração de representações não foto-realísticas. As técnicas apresentadas estão divididas em:

- a. Técnicas usadas no pré-processamento: (i) *correção cromática*; (ii) *conversão de espaço cromático* e (iii) *suavização*. Tais operações visam, respectivamente: (i) a atenuação de efeitos de iluminação indesejados; (ii) o uso de um espaço cromático mais adequado à geração de representações não foto-realísticas; e (iii) a supressão de detalhes e ruídos.
- b. Técnicas usadas no pós-processamento: (i) *ajuste de brilho e contraste*; e (ii) *transformação histogrâmica* por operador exponencial. As operações de pós-processamento visam melhorar o aspecto visual do esboço final gerado.

A seguir, são detalhadas as técnicas citadas. A apresentação de cada técnica envolve: (i) a fundamentação e o princípio de funcionamento; (ii) o objetivo esperado com a aplicação da operação; e (iii) pares de imagens de entrada e saída.

### A.1. Técnicas para Pré-Processamento

As técnicas usadas no pré-processamento se dividem em três sub-etapas: (i) *correção cromática*; (ii) *conversão de espaço cromático*; e (iii) *suavização*. A primeira visa à melhoria das condições de iluminação da imagem, a fim de que falsas bordas não sejam geradas, assim como à não omissão de bordas necessárias. A segunda etapa consiste na conversão do

espaço cromático RGB para um espaço cromático mais adequado à geração de representações não foto-realísticas. A terceira etapa consiste na suavização da imagem para a omissão de detalhes, a abstração da imagem e a melhoria da qualidade final do esboço.

### **A.1.1. Correção Cromática**

O sistema visual humano é capaz de processar descritores de cenas que permanecem constantes, mesmo quando ocorrem alterações na iluminação da cena. Um observador humano é capaz de reconhecer a cor dos objetos, a despeito da luz usada para iluminá-los. Tal habilidade é mais conhecida como *constância cromática (color constancy)* e envolve o processamento conjunto pelo cérebro e pela retina (Ebner, 2007).

O processamento de cores é importante na visão computacional, pois simplifica tarefas quando a cor correta de um objeto é conhecida. No âmbito da visão computacional, o desenvolvimento de algoritmos que imitam a constância cromática pode considerar uma de duas abordagens possíveis: (i) a determinação da reflectância dos objetos da cena, ou seja, da quantidade de luz refletida pelos objetos da cena; ou (ii) a correção cromática da cena, no sentido de imitar o desempenho do sistema visual humano.

A primeira abordagem é importante para o desenvolvimento de sistemas de visão artificial capazes de reconhecer objetos baseados no processamento das cores. A segunda é importante, por exemplo, para melhorar ajustes tais como o balanço de branco automático em câmeras digitais de uso amador. Devido à falta de várias informações usadas pelo sistema visual humano, como o tipo e cor do iluminante, assim como a geometria do objeto retratado, algumas simplificações do mundo real devem ser feitas, a fim de se poder desenvolver algoritmos de correção cromática.

Edwin H. Land, fundador da Polaroid®, é um dos pesquisadores mais proeminentes na área de constância cromática (Ebner, 2007). É atribuída a Land a criação do termo *retinex*, formado pela mistura dos termos **retina** e **córtex**, por não saber se os descritores responsáveis pela computação da

constância cromática estavam localizados na retina do olho ou no córtex visual cerebral.

Com base na teoria de Land, vários algoritmos de constância cromática têm sido propostos (Ebner, 2007; Finlayson et al., 1998; Rahman et al., 1999). Nesta pesquisa, dois algoritmos de constância cromática foram pesquisados, além de uma operação de expansão linear de histograma. O objetivo da operação de expansão linear de histograma é possibilitar o aprimoramento do contraste e da visibilidade de detalhes em uma imagem. Cada operação é detalhada a seguir.

### **A.1.1.1. *GrayWorld***

A hipótese do mundo cinza (*GrayWorld assumption*) foi proposta por Buschsbaum (1980). A hipótese de que o iluminante poderia ser estimado pela computação de uma média da luz recebida como estímulo tem longa história e também foi proposta por Land (Judd, 1960). Buschsbaum (1980), sendo o primeiro a formalizar tal método, também assumiu que, na média, o mundo é cinza.

Supondo uma imagem com variação suficiente de cores, a média dos valores das componentes R, G e B deve tender a um nível de cinza comum. Algoritmos de balanceamento de cores aplicam a hipótese do mundo cinza, forçando um valor médio comum para cada canal R, G e B, que atenua efeitos de uma condição de iluminação não favorável. A aproximação do valor dos pixels pelo valor médio faz com que a cor da imagem fique próxima à cor da cena original. Tal conceito é mais bem entendido pela análise do pseudo-código apresentado a seguir.

```
GrayWorld(imagem)
mediaR = mediaCanal(imagem, R);
mediaG = mediaCanal(imagem, G);
mediaB = mediaCanal(imagem, B);
mediaGeral = media(mediaR + mediaG + mediaB);

maxR = maximo(imagem, R);
maxG = maximo(imagem, G);
maxB = maximo(imagem, B);
fatorEscala = maximo(maxR, maxB, maxG);

FOR i = 0 TO imagem.altura DO
FOR j = 0 TO imagem.largura DO
imagem[i,j,R] = ((imagem[i,j,R]*(mediaGeral/mediaR))/fatorEscala)*255;
imagem[i,j,G] = ((imagem[i,j,G]*(mediaGeral/mediaG))/fatorEscala)*255;
imagem[i,j,B] = ((imagem[i,j,B]*(mediaGeral/mediaB))/fatorEscala)*255;
```

Outros algoritmos de constância cromática foram propostos ao longo dos anos (Ebner, 2007) usando a hipótese do mundo cinza. No entanto, o método se mostra limitado quando a cena possui mais de um iluminante, sendo mais adequado quando a cena possui grande variação de cores. Tal limitação pode ser visualizada na Figura A.1.



Figura A.1 – Exemplo de falha do algoritmo *GrayWorld*.

### A.1.1.2. Retinex

O método *GrayWorld* é mais adequado a cenas nas quais a iluminação é uniforme ou há apenas uma fonte de luz. No entanto, há outros algoritmos que não assumem uma iluminação da cena, como o *Retinex*.

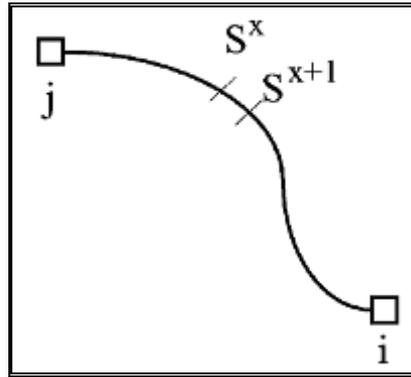
Rizzi, Gatta, e Marini (2004) fazem uma breve avaliação e revisão do modelo *Retinex* proposto por Land e McCann (Ebner, 2007; Finlayson et al., 1998; Rahman et al., 1999). A teoria *Retinex* assume que o sistema visual humano é composto por três sistemas retino-córtex (Rizzi, Gatta, e Marini, 2004), responsáveis pela absorção das baixas, médias e altas frequências do espectro visível. Cada sistema forma uma imagem diferente, determinando a luminosidade relativa das várias regiões da cena.

De acordo com Land e McCann (1971), as bordas presentes entre áreas adjacentes participam ativamente na percepção cromática. A razão (divisão) da luminosidade entre duas regiões foi escolhida como um fator que descreve o relacionamento entre elas. Portanto, se duas áreas possuem luminosidade muito diferentes, a razão (divisão) afasta-se de 1 (um), e se possuem luminosidades próximas, o valor se aproxima de 1. Se a razão da luminosidade for mensurada em várias regiões da imagem, a dominância cromática pode ser compensada, equalizando assim a luminosidade geral da cena.

A resposta visual  $S^x$  na coordenada  $x$  da imagem é dada por:

$$S_{R,G,B}^x = \int_{\lambda \in (400nm, 700nm)} E(\lambda) R^x(\lambda) \rho_{R,G,B}(\lambda) d\lambda \quad (\text{A.1})$$

em que  $x$  é a coordenada espacial,  $E(\lambda)$  é a distribuição espectral da potência da luz do iluminante,  $R^x(\lambda)$  é a reflectância no ponto  $x$ ,  $\rho_{R,G,B}(\lambda)$  é a sensibilidade espectral do pigmento do fotorreceptor. Na Figura A.2, é mostrada a idéia do caminho randômico e da computação da resposta visual  $S^x$  para a equalização da luminosidade da cena.



**Figura A.2 – Caminho entre j e i. A luminosidade relativa é computada como a razão entre as resposta visuais dos pontos  $x$  e  $x+1$ .**

A luminosidade relativa  $L$ , no ponto  $x$ , ao longo de um caminho randômico na imagem (mostrado na Figura A.2), para cada canal RGB, é dada pela seguinte equação:

$$L_{R,G,B}^{i,j} = \sum_{x \in \text{caminho}} \delta \log \frac{S^{x+1}}{S^x} \quad (\text{A.2})$$

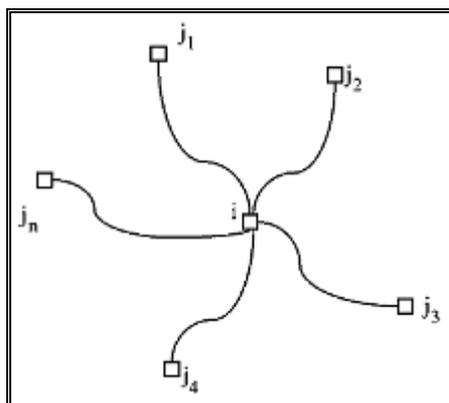
em que:

$$\delta = \begin{cases} 1 \Rightarrow \left| \log \frac{S^{x+1}}{S^x} \right| > \text{limiar} \\ 0 \Rightarrow \left| \log \frac{S^{x+1}}{S^x} \right| < \text{limiar} \end{cases} \quad (\text{A.3})$$

Ao final do algoritmo Retinex, o valor da intensidade  $S_{RGB}$  recalculada para cada ponto  $i$  é a média da luminosidade relativa  $L_{RGB}$  computada em um número  $N$  de caminhos randômicos, terminando no ponto  $i$ , especificada por:

$$S_{R,G,B}^i = \frac{\sum_{k=1}^N L_{R,G,B}^{i,j_k}}{N} \quad (\text{A.4})$$

Na Figura A.3 pode ser observada a idéia de tais caminhos randômicos:



**Figura A.3 – Exemplos de caminhos randômicos ( $j_1 \rightarrow j_n$ ) para a computação da luminosidade em um ponto  $i$ .**

Os cálculos podem ser computados de forma independente para cada canal do espaço RGB. Este algoritmo depende de vários fatores, tais como: (i) a quantidade e aleatoriedade dos caminhos escolhidos para a medição da resposta visual  $S_x$ ; (ii) valores-limite de  $\delta$  que tornam possíveis a computação de mudanças graduais na iluminação; e (iii) um mecanismo de recomposição para que, se uma área mais clara for identificada, durante a computação de um caminho, a luminosidade relativa seja corretamente computada, fazendo com que a área mais clara da imagem seja tomada como a referência de branco da imagem, a fim de que a média dos valores seja corretamente computada.

Dependendo dos parâmetros e da imagem de entrada, o algoritmo *Retinex* pode prover aguçamento, compensação de iluminação, atenuação de sombras, melhoria na constância cromática da imagem ou compressão da faixa dinâmica (Sharma, 2002).

Várias versões do algoritmo *Retinex* foram propostas ao longo do tempo: Single Scale *Retinex* (SSR), Multi-Scale *Retinex* (MSR), Multi-Scale *Retinex* with Color Restoration (MSRCR) (Barnard e Funt, 1999). O Multi Scale *Retinex* é o método que tenta transpor o hiato presente entre a cena capturada por uma câmera e a cena captada por um olho humano (Barnard e Funt, 1999). Este foi o método escolhido para os testes realizados no âmbito desta pesquisa.

No contexto do *Multi Scale Retinex* usado nesta dissertação, os parâmetros foram definidos experimentalmente, sendo: número de escalas = 3; pesos das escalas = {0.3, 0.4, 0.3}; valores de referência de desvio padrão para as escalas = {50, 100, 200}; Ganho = 180 e Offset = 140.

### A.1.1.3. Expansão de Contraste

A expansão de contraste, ou expansão linear de histograma, é uma operação pontual linear que expande a faixa de valores do histograma de uma imagem, utilizando toda a faixa de valores disponíveis (Bovik, 2005).

A expansão de contraste divide-se em duas partes: (i) a determinação dos extremos do histograma; e (ii) o mapeamento dos níveis de cinza presentes. Na primeira etapa, o valor máximo e mínimo de cada canal RGB é obtido:

$$\begin{aligned} a &= \text{mínimo}(\gamma) \\ b &= \text{máximo}(\gamma) \end{aligned} \tag{A.5}$$

Os valores entre a e b são então mapeados para valores entre 0 e  $k-1$ , em que  $k-1$  é o valor máximo de intensidade (adotado 255). Esta operação é realizada a partir das equações:

$$\begin{aligned} pa + l &= 0 \\ pb + l &= k - 1 \end{aligned} \tag{A.6}$$

em que:

$$p = \left( \frac{k-1}{b-a} \right) \tag{A.7}$$

e

$$l = -a \left( \frac{k-1}{b-a} \right) \tag{A.8}$$

Portanto, a expansão linear de contraste é dada por:

$$\delta(i, j) = \text{expansão}(\gamma(i, j)) = \left( \frac{k-1}{b-a} \right) (\gamma(i, j) - a) \tag{A.9}$$

A partir da expansão de contraste, há um aprimoramento da visibilidade de detalhes da imagem, melhorando a faixa dinâmica de imagens superexpostas (muito claras) ou subexpostas (muito escuras).

Na Figura A.4, pode ser visualizada a aplicação de cada um dos

algoritmos relatados nesta seção.



Figura A.4 – Correção cromática: (A) Original; (B) Expansão; (C) *GrayWorld*; e (D) *Retinex*.

### A.1.2. Conversão de Espaço Cromático

O propósito de um espaço cromático é facilitar a especificação de cores, gerando um padrão de representação das cores (Gonzalez e Woods, 2008). Vários espaços cromáticos foram criados nos últimos séculos (Poyton, 2003), cada um servindo a um propósito bem definido. O modelo cromático RGB é um modelo aditivo de cores, formado pelas componentes Vermelho (**R**ed), Verde (**G**reen) e Azul (**B**lue), combinadas de forma a reproduzir outras cores do espectro visível (Gonzalez e Woods, 2008). Tal modelo tem como propósito a reprodução de cores em dispositivos tais como monitores de computador, televisões e scanners.

O modelo RGB não define, espectroscopicamente, o significado exato de vermelho, verde ou azul. Assim, os resultados da combinação das principais componentes são relativos a média da percepção do olho humano. No entanto, a padronização por meio do modelo sRGB (*standard* RGB) fez com que as cores fossem reproduzidas fielmente em sistemas diferentes.

Com o intuito de melhorar as representações não foto-realísticas geradas, dois espaços cromáticos foram pesquisados: (i) *HSV*; e (ii) *YCbCr*. Para tal objetivo, apenas um canal é usado: canal *V*, no modelo *HSV*; e canal *Y*, no modelo *YCbCr*. A seguir, os espaços cromáticos pesquisados são brevemente detalhados.

### A.1.2.1. Espaço Cromático YCbCr

Um espaço cromático robusto a nuances de sombras (*shading*) e luzes (*highlights*) é desejável para a geração de representações não foto-realísticas. Nos modelos cromáticos do tipo luminância-crominância (como YIQ, YUV, YCrCb), divide-se a informação da imagem em uma componente de luminância (brilho, presente na componente Y) e duas componentes de crominância (cor ou tonalidade). O modelo YUV foi o modelo cromático originalmente adotado pelo sistema PAL (***Phase Alternating Line***) de televisão analógica, enquanto o YIQ é o modelo adotado pelo padrão NTSC (***National Television System Committee***). Por sua vez, o espaço cromático YCbCr, intimamente relacionado ao YUV, é utilizado no padrão de compressão MPEG e na codificação digital de imagens do padrão JPEG.

As vantagens da conversão do espaço RGB para o YCbCr é que tal conversão possibilita aos dois sinais de cor serem sub-amostrados sem perda perceptível de detalhes. Além disto, a luminância não precisa ser codificada usando uma grande quantidade de bits, pois vários valores em uma escala de 255 tons de cinza tornam-se redundantes, já que o sistema visual humano não consegue distinguir níveis de cinza muito próximos (Ebner, 2007). Ou seja, a conversão para o espaço YCbCr reduz a redundância na representação dos dados presente na codificação RGB. A equação de conversão de RGB para YCbCr é:

$$\begin{cases} Y = 0,2989 \times R + 0,5866 \times G + 0,1145 \times B \\ Cb = -0,1688 \times R - 0,3312 \times G + 0,5000 \times B \\ Cr = 0,5000 \times R - 0,4184 \times G - 0,0816 \times B \end{cases} \quad (\text{A.10})$$

Na Figura A.5, são apresentados os três canais resultantes da conversão de uma imagem RGB para o espaço cromático YCbCr.



Figura A.5 – Conversão para o espaço cromático YCbCr.

### A.1.2.2. Espaço Cromático HSV

O espaço cromático HSV é um modelo que descreve as cores como percebidas por seres humanos. HSV é acrônimo para Matiz (**H**ue), Saturação (**S**aturation) e Valor (**V**alue). O matiz é essencialmente a componente cromática percebida pelos humanos. O valor provê a intensidade do brilho das cores em termos de quanta luz é refletida ou emitida por um objeto (Ebner, 2007).

O modelo cromático HSV é comumente usado em aplicativos de manipulação de imagens e aplicativos para desenho, sendo usadas comumente nesse tipo de aplicação algumas metáforas comuns à pintura realizada com tinta, pincel e tela.

O espaço cromático HSL (ou HLS ou HSI) tem as componentes Matiz (**H**ue), Saturação (**S**aturation) e **L**uminosidade (*Lightness*, *Luminance* ou *Luminosity*) ou Intensidade (I). Enquanto o espaço HSV é visto como um cone de cor (Figura A.6), o espaço HSL pode ser visto como uma pirâmide exagonal. Os dois espaços são transformações não lineares do cubo RGB (Ebner, 2007).

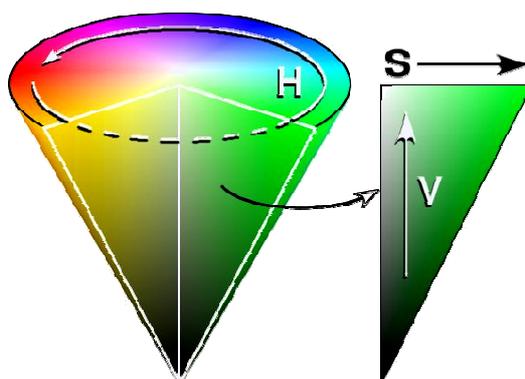


Figura A.6 – Cone HSV

A equação de conversão do espaço RGB para o espaço HSV é:

$$\begin{aligned}
 H = \begin{cases} \left[ 0 + \frac{(G-B)}{(\max-\min)} \right] \times 60 \Rightarrow R = \max \\ \left[ 2 + \frac{(B-R)}{(\max-\min)} \right] \times 60 \Rightarrow G = \max \\ \left[ 4 + \frac{(R-G)}{(\max-\min)} \right] \times 60 \Rightarrow B = \max \end{cases} & \quad (\text{A.11}) \\
 S = (\max - \min) / \max \\
 V = \max
 \end{aligned}$$

sendo *max* e *min* respectivamente iguais ao máximo e ao mínimo dos valores RGB no intervalo [0, 1]. Na Figura A.7, são apresentados os três canais resultantes da conversão de uma imagem RGB para o espaço cromático YCbCr.

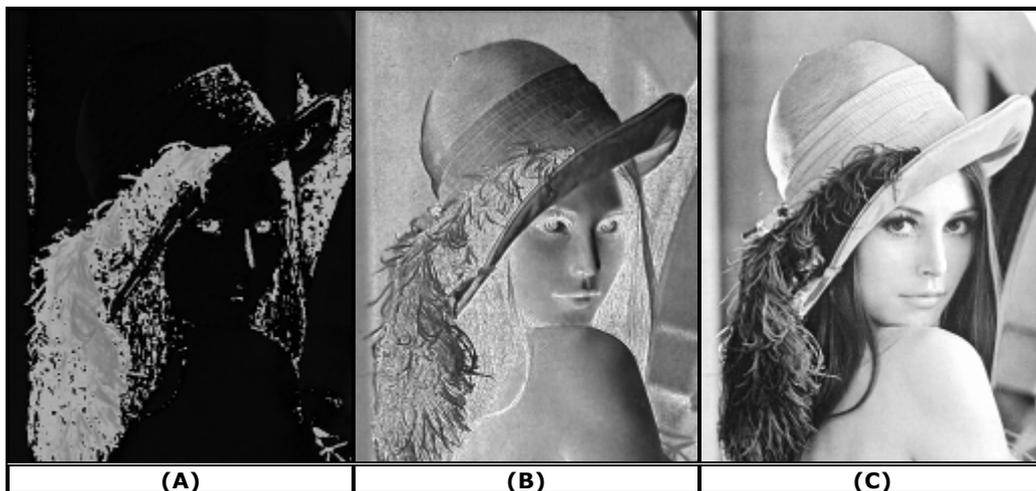


Figura A.7 – Conversão para o espaço cromático HSV.

### A.1.3. Suavização

Os filtros de suavização são usados para homogeneizar a imagem, atenuando detalhes finos, tais como bordas, razão pela qual são utilizados para reduzir o ruído presente na imagem (Gonzalez e Woods, 2008). O efeito da suavização é usado, por vezes, em etapas de pré-processamento, nas quais é necessária a remoção de detalhes da imagem, antes de uma extração de objetos, bem como a conexão de pequenas fissuras (*gaps*) em linhas ou curvas. Por conseguinte, os filtros de suavização atuam atenuando as altas frequências, produzindo uma homogeneização geral da cena, podendo ter graus diferentes de efeito, dependendo dos pesos usados na

janela de convolução e, principalmente, das dimensões desta janela.

No domínio espacial, o uso de filtros tem como consequência a variação no valor digital de um pixel da cena original, segundo a influência de seus vizinhos, a qual é função dos valores positivos, negativos ou nulos fornecidos pelo usuário e atribuídos ao conjunto de pesos que caracterizam a "máscara" ou "janela" de filtragem utilizado. A janela é deslocada sobre a imagem, linha a linha, coluna a coluna. Através da combinação dos valores de entrada ou pesos, se promoverá um maior ou menor realce da cena, segundo as direções preferenciais de interesse (Paradella, 1990).

Podem ser geradas máscaras de quaisquer dimensões. Para uma máscara de dimensões  $M \times N$  pixels, o resultado para o pixel  $(i, j)$  da imagem será:

$$r(i, j) = \sum_{m=1}^M \sum_{n=1}^N \varphi(m, n) \cdot t(m, n) \quad (\text{A.12})$$

sendo  $(m, n)$  o valor de brilho do pixel, endereçado de acordo com a posição da máscara, e  $t(m, n)$  a máscara de entrada para esta localização.

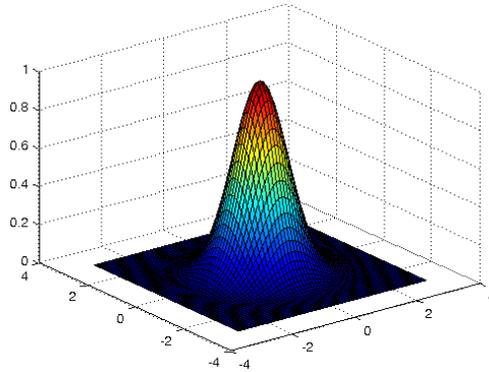
No entanto, como bordas são caracterizadas por transições abruptas nos valores de intensidade dos pixels, um processo de suavização tem o efeito de desfocar as bordas da imagem. A utilização desses filtros nesta pesquisa teve como propósito a redução de detalhes na imagem de entrada, tendo sido pesquisados: (i) o filtro da mediana; e (ii) o filtro de suavização gaussiana. A redução de detalhes e remoção de alguns artefatos menos importantes faz parte da abordagem de abstração da imagem para a geração de representações não foto-realísticas.

### **A.1.3.1. Suavização Gaussiana**

O operador gaussiano de suavização é um operador de convolução bi-dimensional usado para remoção de ruído e para suavização de imagens. A suavização gaussiana é semelhante à filtragem da média. No entanto, a máscara usada na suavização obedece a uma função gaussiana. Para uma superfície 2D, uma função gaussiana isotrópica (circularmente simétrica) se apresenta na forma:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (\text{A.13})$$

A distribuição da função com média (0,0) e  $\sigma=1$  é ilustrada graficamente na Figura A.8:



**Figura A.8 – Visualização da distribuição gaussiana em 3D.**

A suavização gaussiana corresponde à convolução de uma imagem com uma função gaussiana, na qual os pesos da máscara usada no processamento devem ser discretizados e aproximados antes da realização da convolução. Após o cálculo de tal máscara, métodos-padrão de convolução podem ser usados. Dada a natureza isotrópica da função gaussiana, a convolução em duas dimensões pode ser decomposta em duas convoluções em uma dimensão.

Quando aplicada a duas dimensões, a função produz uma superfície cujos contornos são círculos concêntricos que formam uma distribuição gaussiana a partir do ponto central. Os valores de tal distribuição são usados na construção de uma matriz de convolução que é aplicada à imagem original. Cada valor de um novo pixel recebe a média ponderada de seus vizinhos, em contraste com a média uniforme dada pelo filtro da média.

O grau de suavização é determinado pelo desvio padrão da função, ou seja, quanto maior desvio padrão, mais desfocado será o resultado. No entanto, quanto maior o desvio, maior será a máscara de convolução necessária para que os valores sejam representados com boa precisão.

Escolhendo apropriadamente o filtro gaussiano, é possível confiar e estabelecer limites para qual intervalo de frequências espaciais estarão presentes na imagem após a suavização, o que não é possível quando se usa o filtro da média. Este comportamento tem implicações em alguns detectores de borda, tais como o Canny, que utiliza um filtro similar em um dos passos para a detecção de bordas (Paradella, 1990).

### **A.1.3.2. Suavização Usando Filtro da Mediana**

O filtro da mediana é um filtro de ordem estatística, ou seja, é um filtro não linear cuja resposta é baseada na ordenação de pixels em uma região considerada da imagem. O filtro da mediana é o mais conhecido filtro estatístico da ordem (Gonzalez e Woods, 2008). Assim, como o próprio nome deixa entrever, o valor de um pixel é substituído pela mediana dos níveis de intensidade na vizinhança daquele pixel:

$$\hat{f}(x, y) = \underset{(s,t) \in S_{xy}}{\text{mediana}} \{g(s, t)\} \quad (\text{A.14})$$

O valor do pixel na posição  $(x, y)$  é incluído na computação da mediana. O filtro da mediana é popular pela excelente capacidade de redução de ruídos em imagens com ruído randômico, bem como pelo menor grau de atenuação da imagem quando comparado a filtros lineares. A principal função do filtro da mediana é forçar com que pixels com intensidades distintas sejam mais parecidos com seus vizinhos. Devido à natureza do filtro, a quantidade de detalhes da imagem é reduzida. No entanto, as bordas (transições abruptas de valores adjacentes) presentes na imagem são preservadas, dependendo do tamanho da máscara usada.

Na Figura A.9, são apresentados exemplos de aplicações das suavizações apresentadas, variando o tamanho da janela, no caso da suavização gaussiana, e variando o tamanho da janela e o número de iterações, no caso da suavização a partir do filtro da mediana.



Figura A.9 – Suavização: (A) Original; (B) Gaussiana 3x3; (C) Gaussiana 5x5; (D) Duas iterações do filtro da mediana 5x5; e (E) Duas iterações do filtro da mediana 3x3.

## A.2. Técnicas para Pós-Processamento

Foram estudadas duas técnicas a serem usadas na etapa de pós-processamento: (i) *correção de brilho e contraste*; e (ii) *transformação histogrâmica* por operador exponencial. A modificação de brilho e contraste na etapa de pós-processamento é feita de forma estática com valores fixos, obtidos experimentalmente. Após a modificação do brilho e contraste, procede-se a aplicação de uma transformação histogrâmica para realce do contraste da imagem, a partir de um operador exponencial.

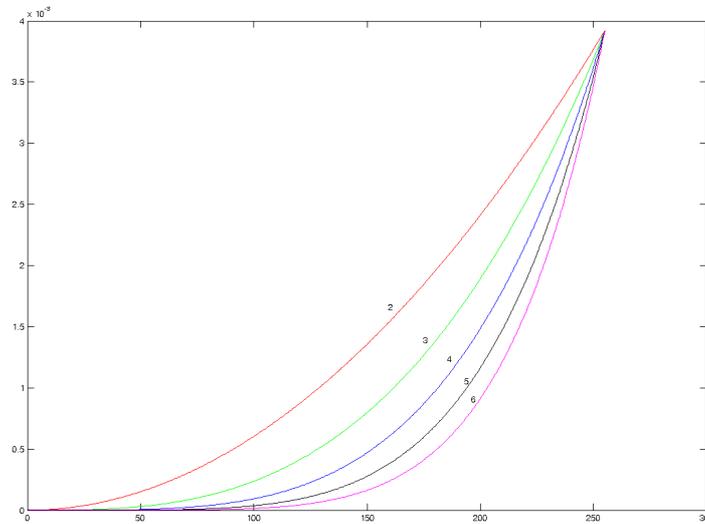
### A.2.1. Transformação Histogrâmica

O operador exponencial pode ser aplicado a imagens não-coloridas com o propósito de alterar a distribuição dos pixels na faixa dinâmica de uma imagem, assim como ocorre ao se usar um operador logarítmico. O operador exponencial, ao contrário do operador logarítmico, enfatiza os valores de pixels de alta intensidade (Gonzalez e Woods, 2008).

O operador exponencial corresponde a um processamento pontual da imagem, no qual a função de mapeamento é uma curva exponencial. No caso do operador "elevado a potência de", os valores dos pixels da imagem de saída é igual ao valor da imagem de entrada (base) elevado a uma potência fixa:

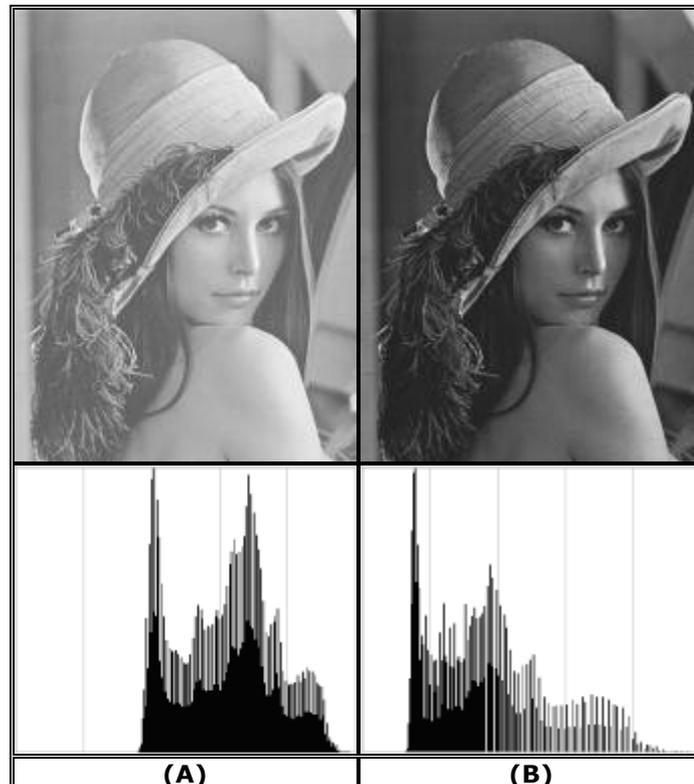
$$Q(i, j) = cP(i, j)^r \quad (\text{A.15})$$

Para valores de  $r$  iguais a 2, 3, 4, 5 e 6, são apresentadas na Figura A.10 as curvas das funções exponenciais correspondentes.



**Figura A.10 – Curvas de potência para operador exponencial.**

Tal operação, no contexto desta pesquisa, teve o propósito de realçar as linhas de bordas mais importantes, fazendo com que mais detalhes sejam percebidos visualmente no esboço gerado. Tal processamento pode ser visualizado na Figura A.11.



**Figura A.11 – Operador exponencial: (A) Imagem cinza; e (B) Após transformação**

Algumas técnicas de processamento digital de imagens úteis à geração de representações não foto-realísticas foram apresentadas neste Apêndice. As técnicas foram divididas em: (i) *correção cromática* (GrayWorld, *Retinex* e expansão linear do contraste); (ii) *conversão de espaço cromático* (YCbCr e HSV); (iii) *suavização* (filtro gaussiano e filtro da mediana); (iv) *ajuste de brilho e contraste*; e (v) *transformação histogrâmica* por operador exponencial.

Na etapa de pré-processamento, deve haver uma melhoria da iluminação da cena em questão, a partir de algoritmos de constância cromática ou de uma expansão de contraste. Após a melhoria das cores da cena, deve haver uma conversão do espaço cromático RGB para um espaço cromático mais adequado à geração de representações não foto-realísticas. Em seguida, deve ocorrer uma suavização da imagem para a supressão de detalhes menos importantes para a geração de representações não foto-realísticas, seguida da melhoria do aspecto visual do esboço e da remoção de ruído.

Na etapa de pós-processamento, o brilho e contraste da imagem de bordas devem ser modificados e deve haver um ajuste, a partir de um operador exponencial, com fins à melhoria do resultado visual.

# Apêndice B

## Resultados Adicionais Obtidos a partir do Uso da Abordagem Proposta

Este apêndice contém alguns resultados obtidos a partir do uso da abordagem proposta para geração de representações não foto-realísticas robustas que se assemelham a esboços semi-detalhados, com avaliação centrada em faces humanas.



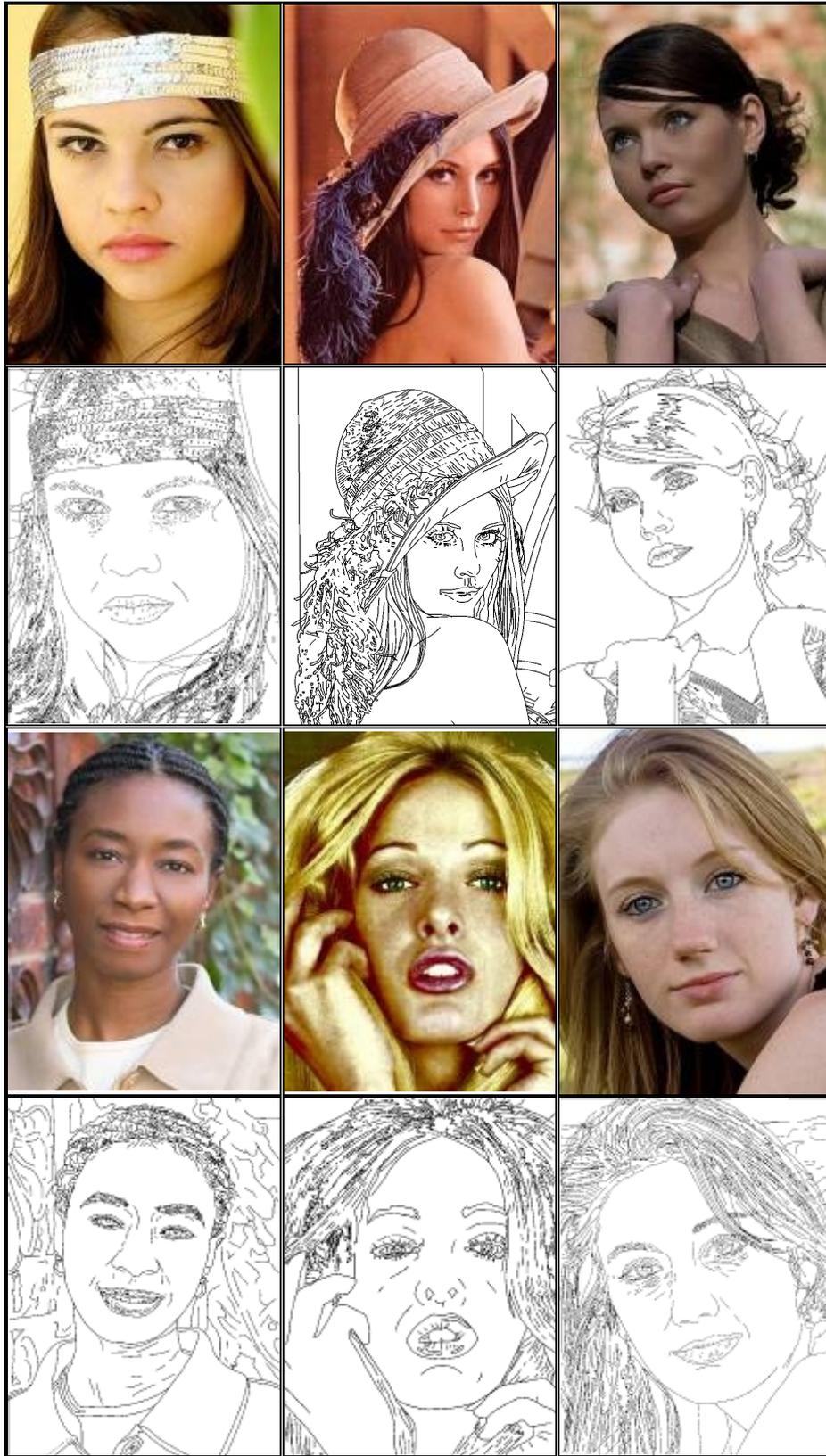


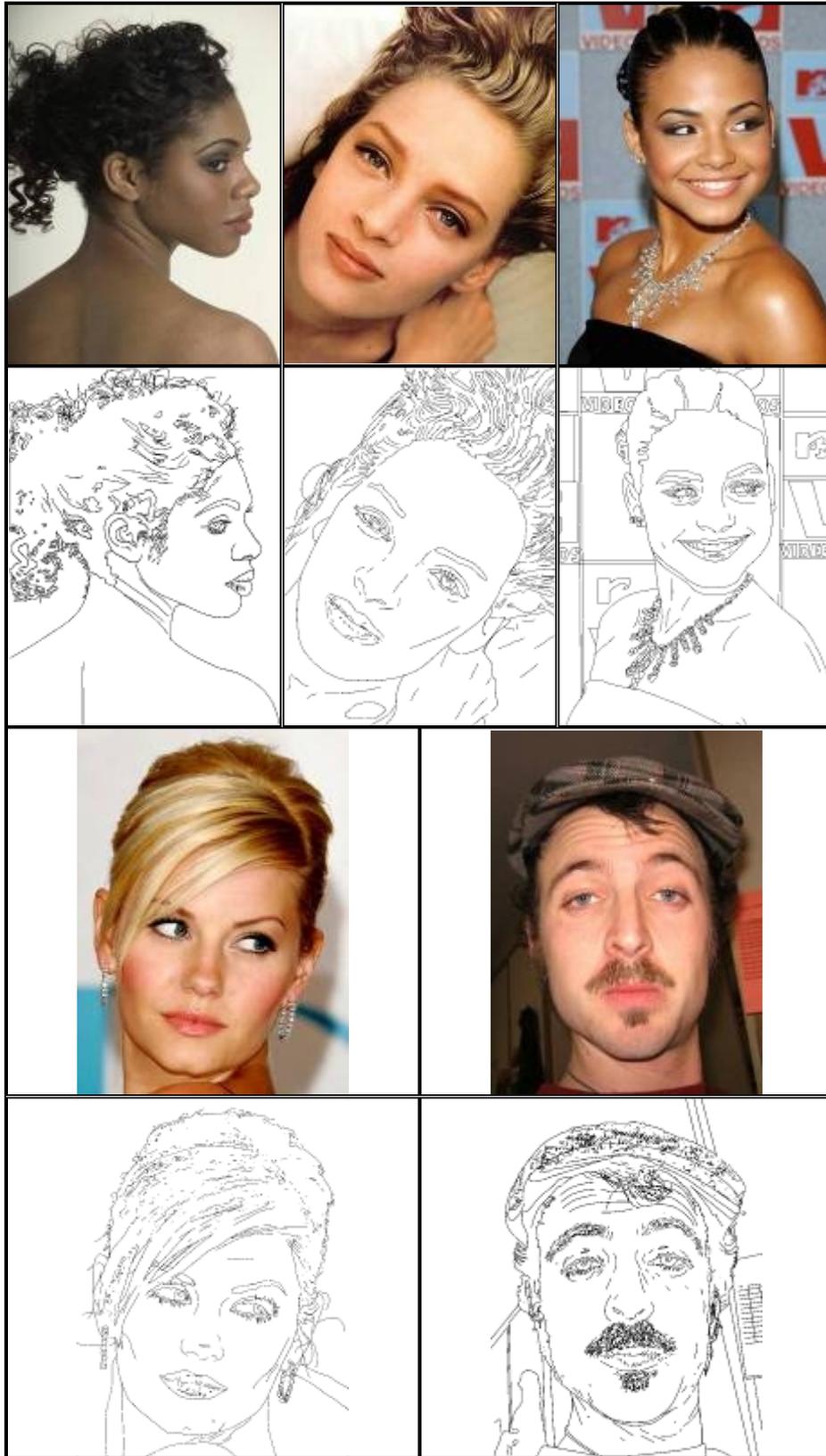
# Apêndice C

## Imagens Utilizadas no Experimento Objetivo

Este apêndice contém as imagens e os respectivos *ground-truths*, anotados manualmente, que compuseram as etapas de treinamento da rede neural MLP baseada em *ground-truth*, avaliação de classificadores usando a curva ROC e avaliação objetiva de detectores de borda. As imagens são apresentadas em sua forma original, seguida da anotação manual em forma de imagem binária.







# Apêndice D

## Teste de Tukey para Valores de FoM

Este apêndice contém o resultado do teste de Tukey para valores de FoM. Na Tabela D.2, é fundamentada a escolha de um grupo de algoritmos a serem usados na abordagem proposta, por meio do teste de Tukey. Os grupos comparados estão descritos na Tabela D.1.

**Tabela D.1 – Grupos comparados no teste de Tukey.**

GrayWorld				Retinex				Expansão de histograma			
Canal Y		Canal V		Canal Y		Canal V		Canal Y		Canal V	
Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.	Gauss.	Medi.
<b>G1</b>	<b>G2</b>	<b>G3</b>	<b>G4</b>	<b>G5</b>	<b>G6</b>	<b>G7</b>	<b>G8</b>	<b>G9</b>	<b>G10</b>	<b>G11</b>	<b>G12</b>

**Tabela D.2 – Teste de Tukey (*Honestly Significant Differences*) para valores de FoM.**

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,1$
G1 vs G7	0,106	3,384	2,960	0,031	<b>Sim</b>
G1 vs G8	0,104	3,304	2,960	0,039	<b>Sim</b>
G1 vs G5	0,098	3,113	2,960	0,067	<b>Sim</b>
G1 vs G6	0,097	3,090	2,960	0,071	<b>Sim</b>
G1 vs G3	0,081	2,587	2,960	0,235	Não
G1 vs G4	0,080	2,561	2,960	0,247	Não
G1 vs G9	0,048	1,519	2,960	0,882	Não
G1 vs G10	0,045	1,420	2,960	0,919	Não
G1 vs G2	0,040	1,267	2,960	0,959	Não
G2 vs G7	0,067	2,117	2,960	0,519	Não
G2 vs G8	0,064	2,037	2,960	0,575	Não
G2 vs G5	0,058	1,846	2,960	0,705	Não
G2 vs G6	0,057	1,823	2,960	0,720	Não
G2 vs G3	0,041	1,320	2,960	0,947	Não
G2 vs G4	0,041	1,293	2,960	0,954	Não
G2 vs G9	0,008	0,252	2,960	1,000	Não
G2 vs G10	0,005	0,153	2,960	1,000	Não
G10 vs G7	0,062	1,964	2,960	0,626	Não
G10 vs G8	0,059	1,884	2,960	0,680	Não
G10 vs G5	0,053	1,693	2,960	0,797	Não
G10 vs G6	0,052	1,670	2,960	0,810	Não
G10 vs G3	0,037	1,167	2,960	0,976	Não
G10 vs G4	0,036	1,140	2,960	0,979	Não
G10 vs G9	0,003	0,099	2,960	1,000	Não
G9 vs G7	0,059	1,865	2,960	0,692	Não
G9 vs G8	0,056	1,785	2,960	0,743	Não

G9 vs G5	0,050	1,594	2,960	0,849	Não
G9 vs G6	0,049	1,571	2,960	0,860	Não
G9 vs G3	0,034	1,068	2,960	0,987	Não
G9 vs G4	0,033	1,042	2,960	0,989	Não
G4 vs G7	0,026	0,824	2,960	0,998	Não
G4 vs G8	0,023	0,744	2,960	0,999	Não
G4 vs G5	0,017	0,553	2,960	1,000	Não
G4 vs G6	0,017	0,529	2,960	1,000	Não
G4 vs G3	0,001	0,026	2,960	1,000	Não
G3 vs G7	0,025	0,797	2,960	0,999	Não
G3 vs G8	0,023	0,717	2,960	0,999	Não
G3 vs G5	0,017	0,526	2,960	1,000	Não
G3 vs G6	0,016	0,503	2,960	1,000	Não
G6 vs G7	0,009	0,294	2,960	1,000	Não
G6 vs G8	0,007	0,214	2,960	1,000	Não
G6 vs G5	0,001	0,023	2,960	1,000	Não
G5 vs G7	0,009	0,271	2,960	1,000	Não
G5 vs G8	0,006	0,191	2,960	1,000	Não
G8 vs G7	0,003	0,080	2,960	1,000	Não
<b>Valor crítico d de Tukey:</b>			<b>4,187</b>		

# Apêndice E

## Métricas para Avaliação de Detectores de Bordas

Neste apêndice, há uma breve fundamentação teórica das métricas destinadas à avaliação de detectores de borda, bem como um breve delineamento das métricas objetivas usadas no Capítulo 4.

O grande número de detectores de bordas existentes serve de motivação para a concepção de uma forma sistemática de avaliação dos resultados da detecção de bordas, com vistas a atestar ou classificar qual detector (ou combinação de detectores) produz melhor resultado (Wang et al., 2006). No contexto de tal classificação, há abordagens subjetivas e objetivas. A avaliação subjetiva, inerentemente ambígua, usa a percepção humana para avaliar a qualidade das bordas. Em métodos objetivos, resultados numéricos são comparados (Wang et al., 2006).

Avaliações objetivas de detectores de bordas podem ser categorizadas em função da necessidade ou não de *ground-truth* (Prieto e Allen, 2003). Sem o uso de *ground-truth*, há métodos que consideram a coerência das bordas detectadas, ou seja, há a medição da continuidade e espessura das bordas detectadas. Adicionalmente, outros métodos de avaliação de bordas usam um *ground-truth* sintético, obtido após a junção de várias bordas obtidas por múltiplos detectores. Quanto às imagens de teste, estas podem ser sintéticas ou reais (Heath et al., 1997).

Há uma dificuldade inerente ao processo de avaliação de bordas, pois, conforme revisada por vários autores (Hoover et al. 1995; Palmer, Dabis e Kittler, 1996; Wang et al., 2006; Jiang et al., 2006), a medição objetiva do desempenho de um detector de bordas está intimamente ligada ao escopo da aplicação do detector. Portanto, métricas de propósito geral podem levar a resultados não confiáveis, mas avaliações que dependem do domínio de

aplicação do detector reduzem a generalidade do método de avaliação.

Prieto e Allen (2003) estudaram métricas clássicas para a avaliação de bordas: *Closest Distance Metric (CDM)*, *Figure of Merit (FoM)* e *Peak Signal-to-Noise Ratio (PSNR)*. Os autores ressaltaram problemas associados às métricas clássicas, tais como o fato de a métrica FoM permitir que haja uma correspondência múltipla entre pixels detectados como borda com o equivalente no *ground-truth*. Tal fato serviu de motivação para que Prieto e Allen (2003) propusessem uma nova métrica, denominada *Pixel Correspondence Metric (PCM)*.

PCM contempla, além da distância da borda detectada à borda correspondente no *ground-truth*, a força da borda, tida como informação relevante para melhor estimativa da similaridade entre imagens. É importante lembrar que tal informação é perdida no processo de binarização e é por vezes ignorada na maioria das métricas.

O algoritmo proposto por Prieto e Allen (2003) adota a medida de melhor compatibilização (*matching*) de grafos bipartidos ponderados, constituindo, como relatado pelos próprios autores, um desafio de implementação computacional não trivial.

Adicionalmente, os resultados obtidos por Prieto e Allen (2003) revelaram inconsistências na métrica FoM, quando avaliada usando imagens ruidosas, já que esta produziu valores iguais para imagens com grau de ruído diferentes. Há ainda outra inconsistência com valores da métrica PSNR, a qual produz valores intimamente relacionados às dimensões da imagem. A seguir, são apresentadas três métricas de similaridade usados no presente trabalho: (i) Relação Sinal/Ruído Máxima (**Peak Signal to Noise Ratio** – PSNR); (ii) Figura de Mérito de Pratt (**Pratt's Figure of Merit** - FoM) (Pratt, 1978); e (iii) Similaridade Estrutural (**Structural SIMilarity** - SSIM) (Wang, Lu e Bovik, 2002; Wang e Simoncelli, 2005).

### **E.1. Relação Sinal/Ruído Máxima (PSNR)**

A Relação Sinal/Ruído Máxima (**Peak Signal to Noise Ratio** – PSNR) é uma métrica de distorção que pode ser usada para avaliar a similaridade entre

duas imagens, usando uma comparação baseada em pixel, ou seja, não leva em consideração a vizinhança (Prieto e Allen, 2003). A PSNR é uma métrica simples e popular. No entanto, a simplicidade dos cálculos pode levar a avaliações equivocadas. Esta métrica é geralmente usada na avaliação da reconstrução de imagens após compressão.

O erro médio quadrático (**Mean Square Error - MSE**) representa o erro quadrático médio entre a imagem comprimida e a original. Para computar o valor da PSNR, calcula-se o MSE a partir da expressão:

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad (4.2)$$

na qual  $I_1$  e  $I_2$  são imagens com  $M$  linhas e  $N$  colunas. Após o cálculo do MSE, a PSNR pode ser obtida a partir da expressão:

$$PSNR = 10 \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (4.3)$$

na qual  $MAX_I$  é o valor máximo que pode ser atribuído a um pixel, e.g., quando a representação da imagem usa 8 bits, o valor máximo será 255.

Um valor de PSNR maior significa maior similaridade entre imagens. As métricas para comparação de bordas seguem a idéia de estimar a similaridade entre as imagens pela adição de erros causados pelos *pixels* posicionalmente compatíveis e *pixels* posicionalmente incompatíveis dentro de uma determinada vizinhança. A métrica PSNR equivale à medição de similaridade quando a vizinhança é igual a zero, ou seja, a métrica não contempla deslocamento de bordas. Valores mais elevados de PSNR implicam menor presença de ruído sendo, portanto, um aspecto desejável na medição da qualidade de imagens após compressão. Mesmo sendo uma métrica popular, os critérios usados ao avaliar uma imagem usando PSNR nem sempre proveêm uma medida precisa da qualidade visual, especialmente quando são comparadas imagens de borda.

## **E.2. Similaridade Estrutural (SSIM)**

A Similaridade Estrutural (**Structural SIMilarity - SSIM**) é um métrica de

similaridade que usa padrões locais de intensidades de pixels, previamente normalizadas em relação a luminância e contraste (Wang, Lu e Bovik, 2002). SSIM é uma métrica de referência completa, ou seja, mede a qualidade da imagem baseada na comparação com uma imagem de referência de alta qualidade e sem distorções (Wang, Lu e Bovik, 2002). A métrica é calculada com base em máscaras ou padrões de tamanho fixo, de dimensões  $N \times N$ . A medida SSIM de duas máscaras  $x$  e  $y$  é dada por:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2cov_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4.4)$$

sendo:  $\mu_x$  a média de  $x$ ;  $\mu_y$  a média de  $y$ ;  $\sigma_x^2$  a variância de  $x$ ;  $\sigma_y^2$  a variância de  $y$ ;  $cov_{xy}$  a covariância de  $y$ ;  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$  duas variáveis para estabilizar a divisão com denominador comum, sendo  $L$  a faixa dinâmica dos valores dos pixels, tipicamente  $2^{\#bitsporpixel} - 1$ ;  $k_1 = 0,01$  e  $k_2 = 0,03$  por padrão.

A avaliação da qualidade das imagens é feita apenas no canal de luminância. O resultado é um valor decimal no intervalo  $[-1; 1]$ . O valor 1 só é atingido no caso de duas imagens idênticas serem avaliadas. Tipicamente, é usada uma máscara de tamanho  $8 \times 8$ , podendo o cálculo ser realizado *pixel-a-pixel* ou apenas em um sub-grupo de todas as possíveis máscaras, a fim de reduzir o tempo necessário para o cálculo.

Mesmo o PSNR sendo uma métrica popular, um pequeno movimento nas bordas da imagem gerada pelo detector e aquela obtida pela anotação do *ground-truth* levam a um valor de similaridade muito baixo (Prieto e Allen, 2003). A métrica SSIM propõe, então, melhorias ao PSNR, buscando aprimorar a resposta da métrica pela melhor simulação da resposta do sistema visual humano a comparações de imagens de borda. No entanto, o foco da métrica é a avaliação da qualidade da imagem, sendo esta largamente utilizada na avaliação de algoritmos de compressão de vídeo (Wang, Lu e Bovik, 2002).

### **E.3. Figura de Mérito (FoM)**

A Figura de Mérito (**Figure of Merit – FoM**) de Pratt (Pratt, 1978) é uma medida relativa que usa as distâncias entre as bordas detectadas e o

*ground-truth*, permitindo, portanto, uma medição do erro de localização entre a borda ideal e a borda detectada (Prieto e Allen, 2003). Os valores do FoM variam no intervalo [0; 1], sendo o valor 1 obtido quando há uma compatibilidade perfeita entre a imagem e o *ground-truth*.

A medida relativa (ou comparação baseada em vizinhança) usa a localização/tamanho/forma dos objetos da imagem, comparando as bordas obtidas por um detector com uma imagem que contém as bordas ideais. Conforme anteriormente ressaltado, as bordas ideais foram obtidas pela marcação manual, tendo sido utilizada a imagem original como fundo. A métrica FoM é expressa como:

$$FoM = \frac{1}{I_N} \sum_{i=1}^{I_A} \left( \frac{1}{(1 + a \times d^2(i))} \right) \quad (4.5)$$

na qual  $I_N = \max(I_I, I_A)$ ,  $I_I$  representa o número de pixels de borda ideal,  $I_A$  corresponde ao número de pixels da imagem avaliada,  $a$  é um fator de escala constante (usualmente 1/9),  $d(i)$  é o valor máximo da distância entre a posição do pixel da imagem real e da posição do pixel no *ground-truth*.

# Apêndice F

## Imagens Utilizadas no Experimento Subjetivo

Neste apêndice, são apresentadas as imagens utilizadas no experimento subjetivo, bem como uma breve descrição da licença usada pelos proprietários das imagens, permitindo o uso previamente autorizado.

A descrição da licença *creative commons* presente no site Flickr ([www.flickr.com/creativecommons](http://www.flickr.com/creativecommons)) atesta: *Creative Commons* é uma empresa sem fins lucrativos que oferece uma alternativa aos direitos autorais integrais. Resumidamente, em tal licença:

- **Atribuição** significa: Você permite que outros copiem, distribuam, exibam e executem seu trabalho protegido por direitos autorais, e os trabalhos derivados feitos com base nele, mas somente se estes lhe derem crédito;
- **Não comercial** significa: Você permite que outros copiem, distribuam, exibam e executem seu trabalho protegido por direitos autorais, e os trabalhos derivados feitos com base nele, mas somente para propósitos não comerciais;
- **Não a obras derivadas** significa: Você permite que outros copiem, distribuam, exibam e executem somente cópias literais do seu trabalho, e não trabalhos dele derivados; e
- **Compartilhamento pela mesma licença** significa: Você permite que outros distribuam trabalhos derivados somente sob uma licença idêntica àquela que rege o seu trabalho.

As *tags* (etiquetas) pesquisadas com base na nuvem de *tags* para a composição dessa pequena base de imagens foram: *baby; children; family; model; people face; children face; portrait.*

As URLs das imagens são:

- [http://www.flickr.com/photos/anakin\\_skywalker/3556288500/](http://www.flickr.com/photos/anakin_skywalker/3556288500/)
- <http://www.flickr.com/photos/piulet/3553976741/>
- <http://www.flickr.com/photos/craigpitchers/3551665791/>
- <http://www.flickr.com/photos/spicyrogue/3547358314/>
- <http://www.flickr.com/photos/tvanardenne/3546258529/>
- <http://www.flickr.com/photos/conorkeller/3543754999/>
- <http://www.flickr.com/photos/arturnuta/3540665479/>
- <http://www.flickr.com/photos/waechor/485065438/>
- <http://www.flickr.com/photos/geomangio/712941469/>
- <http://www.flickr.com/photos/beija-flor/15709083/>
- <http://www.flickr.com/photos/beija-flor/425218010/>
- <http://www.flickr.com/photos/kristin-and-adam/281936249/>
- <http://www.flickr.com/photos/shutterdreams/2329940116/>
- <http://www.flickr.com/photos/charl22/82263592/>
- <http://www.flickr.com/photos/phildowsing/313982722/>
- <http://www.flickr.com/photos/phildowsing/313982717/>
- <http://www.flickr.com/photos/didier-lg/919170669/>
- <http://www.flickr.com/photos/jervissalvador/2541007474/>
- <http://www.flickr.com/photos/12392252@N03/1498466356/>

As imagens são apresentadas na Tabela F.1

**Tabela F.1 – Imagens utilizadas no experimento subjetivo.**



