

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

TESE

UMA ABORDAGEM HÍBRIDA PARA A RENDERIZAÇÃO

NÃO FOTORREALÍSTICA DE VÍDEOS DIGITAIS

FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA

ORIENTADORES:

JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ

HERMAN MARTINS GOMES

CAMPINA GRANDE, PARAÍBA

MARÇO – 2014

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

**UMA ABORDAGEM HÍBRIDA PARA A RENDERIZAÇÃO
NÃO FOTORREALÍSTICA DE VÍDEOS DIGITAIS**

FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA

Tese submetida à Coordenação do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande, Campus I, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

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 – Brasil

Março de 2014

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A778a	<p>Arruda, Francisco de Assis Pereira Vasconcelos de.</p> <p>Uma abordagem híbrida para a renderização não fotorrealística de vídeos digitais. -- Campina Grande, 2014.</p> <p>245 f. : il. color.</p> <p>Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2014.</p> <p>“Orientação: Prof. Dr. José Eustáquio Rangel de Queiroz, Prof. Ph.D. Herman Martins Gomes”.</p> <p>Referências.</p> <p>1. Processamento de Imagens. 2. Representações não Fotorrealísticas. 3. Inteligência Artificial – Aprendizagem de Máquina. I. Título.</p> <p>CDU 004.932(043)</p>
-------	--

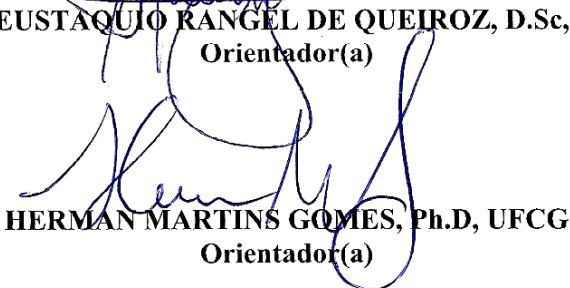
**"UMA ABORDAGEM HÍBRIDA PARA A RENDERIZAÇÃO NÃO FOTORREALÍSTICA
DE VÍDEOS DIGITAIS"**

FRANCISCO DE ASSIS PEREIRA VASCONCELOS DE ARRUDA

TESE APROVADA EM 28/03/2014



JOSÉ EUSTAQUIO RANGEL DE QUEIROZ, D.Sc, UFCG
Orientador(a)



HERMAN MARTINS GOMES, Ph.D, UFCG
Orientador(a)



JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG
Examinador(a)



BRUNO MOTTA DE CARVALHO, Ph.D, UFRN
Examinador(a)



ELMAR UWE KURT MELCHER, Dr., UFCG
Examinador(a)



LEONARDO VIDAL BATISTA, D.Sc, UFPB
Examinador(a)

CAMPINA GRANDE - PB

RESUMO

A renderização *não fotorrealística* (NPR - *Non-Photorealistic Rendering*) ou *estilizada* (*Stylized Rendering*), constitui uma subárea importante e em expansão, comum às áreas da Computação Gráfica (CG) e do Processamento Digital de Imagens (PDI). A NPR visa à geração de imagens ou vídeos que se destacam pela expressividade do resultado final e pela exibição de características que traduzem emoções. Destacam-se como importantes desafios nesta área: (i) variações nas condições de iluminação do vídeo de entrada; (ii) requisitos computacionais elevados para a renderização de vídeos; (iii) dificuldades na configuração dos parâmetros característicos das técnicas propostas, a fim de se produzir resultados equivalentes sobre uma gama mais ampla de gêneros de vídeos de entrada; (iv) dificuldades na renderização de vídeos de forma temporalmente coerente; e (v) dificuldades no processo de avaliação dos resultados produzidos. Esta tese insere-se neste contexto, tendo como foco a definição de uma abordagem original para a renderização não fotorrealística de vídeos ao estilo de cartuns. A abordagem proposta, validada a partir de uma aplicação de *software*, consiste na agregação de seis módulos que apresentam alterações quanto à abordagem fundamentada na suavização por difusão anisotrópica, especificamente no que diz respeito: (i) à modificação da equação de parada-na-aresta da técnica original; (ii) ao uso de bordas destinadas à estilização e orientação do processo de suavização; (iii) à combinação das abordagens de detecção de bordas e suavização por difusão anisotrópica; e (iv) ao aumento da homogeneidade cromática intrarregional e da saturação do resultado obtido. Foi conduzido um processo de avaliação experimental, a partir do qual a abordagem proposta foi comparada a quatro abordagens concorrentes. Após um processo de votação por usuários, a análise dos dados coletados foi conduzida a partir da administração de testes F ANOVA fator único, seguido de testes de comparação de médias, visando a evidenciar as melhorias associadas à estilização produzida pela abordagem proposta. Tendo sido mais bem avaliada, em relação às abordagens comparadas, no tocante à preferência visual direta, à melhor definição de bordas e ao efeito de estilização resultante e à produção de cintilação e artefatos espúrios equivalentes àqueles produzidos pelas abordagens concorrentes comparadas.

Palavras-chave: Processamento de imagens, representações não fotorrealísticas, inteligência artificial – aprendizagem de máquina.

ABSTRACT

Non-photorealistic rendering (NPR) or stylized rendering is an important growing area of Computer Graphics and Image Processing. NPR is mainly concerned with the generation of images or videos that can have an expressive appeal with regards to the display of features that convey emotions. Some research challenges include: (i) coping with variations in the lighting conditions of the input video; (ii) demanding computational requirements for video rendering; (iii) difficulties in parameter adjustment for each compared approach for a wide range of input videos; (iv) processing videos in a temporally coherent manner; (v) difficulties in the evaluation process. In such a context, this thesis aims at the conception of an original approach to generate stylized videos. The proposed approach was validated by means of a software application, consisting of a six-step pipeline, based on the rendering of each video frame by a modified anisotropic diffusion technique. More specifically, the modifications include: (i) the replacement of the original edge stopping function; (ii) the usage of edges to stylize and guide the diffusion process; (iii) the combination of the detected edges and smoothing by anisotropic diffusion; and (iv) the increase of the saturation and intraregional chromatic homogeneity of the obtained result. A comparative, subjective experimental evaluation of the proposed approach and four concurrent ones was performed, which was based on a user voting scheme. The goal of the evaluation was to highlight improvements in the results of the proposed approach. One-Way ANOVA F-tests followed with multiple means comparison tests, were performed on the results. The proposed approach has obtained better results compared to the four other approaches, regarding the direct visual preference, improved edge definition and stylization. In addition, the results of the proposed approach contain spurious artifacts and flickering which are equivalent with those produced by the compared approaches.

Keywords: image processing, non-photorealistic rendering, artificial intelligence – machine learning.

LISTA DE QUADROS E TABELAS

Quadro 2.1	-	Resumo das técnicas destinadas à renderização não fotorrealística de vídeos digitais.....	59
Quadro 2.2	-	Sumário dos aspectos relativos à validação e avaliação de técnicas para renderização de imagens e vídeos.	62
Tabela 4.1	-	Sumário das características dos vídeos selecionados.	126
Tabela 5.1	-	Perfil dos participantes do experimento realizado nesta pesquisa.....	155
Tabela 5.2	-	Grupos de parâmetros para cada abordagem comparada.....	156
Tabela 5.3	-	Sumário dos testes F ANOVA, considerando avaliações intra-abordagens.....	158
Tabela 5.4	-	Sumário dos testes de Tukey-Kramer, considerando avaliações intra-abordagens.	158
Tabela 5.5	-	Sumário dos testes F ANOVA, considerando avaliações interabordagens.	160
Tabela 5.6	-	Grupos de similaridade do teste de Tukey-Kramer, considerando avaliações interabordagem.	160
Tabela 5.7	-	Sumário dos testes F ANOVA, considerando avaliações interabordagens com maiores valores intra-abordagens.	161
Tabela 5.8	-	Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.....	162
Tabela D.1	-	Parâmetros da abordagem proposta por Winnemöller, Olsen e Gooch (2006).....	210
Tabela D.2	-	Parâmetros da abordagem proposta por Kyprianidis, Kang e Döllner (2009).	210
Tabela D.3	-	Parâmetros da abordagem proposta por Gastal e Oliveira (2011).	211
Tabela D.4	-	Parâmetros da abordagem proposta por Kyprianidis e Döllner (2008).	212

Tabela G.1	-	Teste ANOVA para valores da abordagem Transformada de Domínio.	219
Tabela G.2	-	Teste ANOVA para valores da abordagem Kuwahara Anisotrópico.	219
Tabela G.3	-	Teste ANOVA para valores da abordagem Bilateral+DoG.	219
Tabela G.4	-	Teste ANOVA para valores da Flow Bilateral+FDoG.	220
Tabela G.5	-	Teste ANOVA para valores da Abordagem Proposta.	220
Tabela H.1	-	Grupos comparados nos testes ANOVA e Tukey-Kramer.	221
Tabela H.2	-	Grupos de similaridade obtidos no teste de Tukey-Kramer, para aspectos intra-abordagem.	222
Tabela I.1	-	Teste F ANOVA interabordagem para o Aspecto 1.	223
Tabela I.2	-	Teste F ANOVA interabordagem para o Aspecto 2.	223
Tabela I.3	-	Teste F ANOVA interabordagem para o Aspecto 3.	223
Tabela J.1	-	Teste de Tukey-Kramer para o Aspecto 1.	224
Tabela J.2	-	Teste de Tukey-Kramer para o Aspecto 2.	224
Tabela J.3	-	Teste de Tukey-Kramer para o Aspecto 3.	225
Tabela J.4	-	Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.	225
Tabela K.1	-	Teste F ANOVA interabordagem, Aspecto 1, considerando maiores valores intra-abordagem.	226
Tabela K.2	-	Teste F ANOVA interabordagem, Aspecto 2, considerando maiores valores intra-abordagem.	226
Tabela K.3	-	Teste F ANOVA interabordagem, Aspecto 3, considerando maiores valores intra-abordagem.	226
Tabela L.1	-	Teste de Tukey-Kramer para o Aspecto 1.	227
Tabela L.2	-	Teste de Tukey-Kramer para o Aspecto 3.	227
Tabela L.3	-	Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.	228
Quadro M.1	-	Tempo de execução médio para cada módulo, considerando a renderização de 20 imagens.	229

LISTA DE FIGURAS

Figura 1.1	- Exemplos de pontilhismo e entalhe: (A) Don Draper, ao estilo hedcut; (B) Resultado obtido por Kim et al. (2008); e (C) Resultado obtido por Melo et al. (2007).	20
Figura 1.2	- Exemplos de renderizações estilísticas em jogos eletrônicos: (A) Team Fortress 2 (VALVE, 2007); e (B) Street Fighter IV (CAPCOM, 2008).	21
Figura 1.3	- Excerto de cena do filme Amor além da vida: (A) Imagem do cenário antes do processo de estilização; (B) Cena estilizada ao estilo expressionista.	22
Figura 1.4	- Exemplos de renderizações que modificam projetos CAD: (A) Piranesi; (B) PalleteCAD; e (C) Google SketchUP.	24
Figura 2.1	- Taxonomia adotada nesta pesquisa para a área de NPR.	32
Figura 2.2	- Progressão temporal da área de NPR, na qual se destaca a semiautomação das abordagens propostas no início dos anos 1990, seguida da automaização a partir do emprego de técnicas de alto nível (e.g., visão computacional) e da filtração com preservação de bordas e passando à ênfase na interação com o usuário e à avaliação da estética dos resultados de renderização produzidos.	36
Figura 2.3	- Ênfase no movimento.	39
Figura 2.4	- Arquitetura da abordagem proposta por Wang et al. (2004).	41
Figura 2.5	- Arquitetura da abordagem proposta por Collomosse, Rowntree e Hall (2005).	43
Figura 2.6	- Arquitetura a abordagem proposta por Huang (2006).	44
Figura 2.7	- Arquitetura da abordagem proposta por Winnemöller, Olsen e Gooch (2006).	45
Figura 2.8	- Arquitetura da abordagem proposta por Hong et al. (2007).	47
Figura 2.9	- Resultados obtidos por Bousseau et al. (2007).	50
Figura 2.10	- Arquitetura da abordagem proposta por Lee, Lee e Yoon (2009).	51

Figura 2.11	- Máscaras utilizadas no filtro Kuwahara: (A) Original, proposta por Kuwahara et al. (1976); e (B) Máscaras propostas por Kyprianidis, Kang e Döllner (2009).	52
Figura 2.12	- Arquitetura do sistema proposto por Kagaya et al. (2010).	53
Figura 2.13	- Resultado da renderização obtida com a aplicação da abordagem proposta por Bhat et al. (2010): (A) Original; e (B) Renderizada.....	54
Figura 2.14	- Resultado da renderização obtida com a aplicação da abordagem proposta por Gastal e Oliveira (2011): (A) Original; e (B) Renderizada.....	55
Figura 2.15	- Abordagem generalizada proposta por Kyprianidis (2013) para renderização de imagens não fotorrealísticas ao estilo de cartum.	57
Figura 2.16	- Escala pictórica de classificação emocional: (A) Excitação; (B) Valência; e (C) Dominância.....	65
Figura 2.17	- Excerto da interface computacional utilizada por Li, Mould e Davies (2013) durante a coleta de dados de tempo de resposta à renderizações NPR.	67
Figura 2.18	- Excerto da interface computacional utilizada por Li, Mould e Davies (2013) durante a coleta de dados de qualidade de imagens NPR: (A) Estrutural; e (B) Estética.....	67
Figura 2.19	- Modelo arqueológico utilizado na pesquisa realizada por Isenberg et al. (2006).	73
Figura 2.20	- Excerto da interface computacional usada por Seshadrinathan et al. (2010) durante a coleta de dados a respeito da qualidade de vídeos.....	76
Figura 2.21	- Excerto da interface utilizada pelo grupo MSU para avaliação da percepção visual de vídeos.....	78
Figura 2.22	- Visão geral do fluxo de execução de uma sessão experimental. ..	79
Figura 2.23	- Visão geral de dois métodos para avaliação subjetiva da qualidade de imagens: (A) Estímulo simples; e (B) Estímulo duplo.	80
Figura 2.24	- Visão geral de dois métodos para avaliação subjetiva da qualidade de imagens: (A) Escolha forçada; e (B) Julgamento pareado.....	81

Figura 2.25	-	Difusão anisotrópica com diferentes funções de parada-na-aresta.	87
Figura 2.26	-	Difusão anisotrópica com diferentes funções de parada-na-aresta.	88
Figura 2.27	-	Difusão anisotrópica após grande número de iterações.....	88
Figura 2.28	-	Comparação visual de técnicas de suavização: (A) Original; (B) Suavização Gaussiana; (C) filtro da mediana 5x5; (D) Perona-Malik; e (E) RAD.....	89
Figura 3.1	-	Abordagem proposta, fundamentada na suavização por difusão anisotrópica.	92
Figura 3.2	-	Imagens utilizadas para exemplificar o uso da abordagem proposta: (A) Henrik Andersson; (B) Macaco-prego; e (C) Ilha de Capri.	95
Figura 3.3	-	Resultado da aplicação da CEF para realce de características locais.	97
Figura 3.4	-	Aplicação do CEF com valores de parâmetros diferenciados.....	98
Figura 3.5	-	Abordagem para detecção de bordas utilizando redes neurais. ..	99
Figura 3.6	-	Alterações realizadas no detector de bordas neural: (A) Resultado original obtido por Arruda; Queiroz; Gomes (2012); (B) Resultado após modificação de valores de parâmetros.	101
Figura 3.7	-	Resultado da aplicação da detecção de bordas que utiliza redes neurais.....	101
Figura 3.8	-	Resultado da aplicação do processo de difusão anisotrópica modificado.....	103
Figura 3.9	-	Módulo de realce de regiões de sombra.	104
Figura 3.10	-	Resultado da aplicação do processo de realce de regiões de sombra.....	105
Figura 3.11	-	Resultado da aplicação da filtragem NLM a três diferentes imagens de entrada.	106
Figura 3.12	-	Resultado do ajuste de saturação aplicado as três diferentes imagens de entrada.	107
Figura 3.13	-	Coleção de imagens renderizadas com o uso da abordagem proposta: (A) Imagem original; e (B) Imagem renderizada.....	108
Figura 3.14	-	Síntese gráfica de todas as etapas da abordagem proposta.....	109

Figura 3.15	-	Abordagem proposta aplicada a imagens corrompidas com ruído Gaussiano.	111
Figura 3.16	-	Abordagem proposta aplicada à imagem corrompida com ruído sal e pimenta.	111
Figura 4.1	-	Primeiro quadro de cada vídeo utilizado no processo de avaliação.	123
Figura 4.2	-	Ambiente de testes.	130
Figura 4.3	-	Interface gráfica utilizada no experimento.	131
Figura 5.1	-	Difusão anisotrópica robusta com funções de parada na aresta distintas: (A) função de Tukey original; (B) função de Tukey modificada; e (C) Função de Dennis e Welsch (1978).	138
Figura 5.2	-	Comparação de resultados: (A) Original; (B) Suavização Gaussiana; (C) Suavização por filtro Bilateral; (D) Difusão anisotrópica robusta; (E) Resultado da difusão anisotrópica utilizada na abordagem proposta; (F) Resultado da utilização dos seis módulos da abordagem proposta.	139
Figura 5.3	-	Exemplo de filtragem utilizando a CEF: (A) $\tau = 2$ e $\rho = 5$; e (B) $\tau = 6$ e $\rho = 7$	142
Figura 5.4	-	Adoção de valores inadequados para os parâmetros τ e ρ da CEF.	142
Figura 5.5	-	Comparação de ponderação de bordas com diferentes valores de brilho: (A) $B=10$; e (B) $B=50$	144
Figura 5.6	-	Resultado da filtragem NLM para conjuntos de valores diferenciados: (A) Imagem de entrada; (B) $R_b=3$, $R_v=3$; (C) $R_b=3$, $R_v=5$; e (D) $R_b=7$, $R_v=9$	145
Figura 5.7	-	Aplicação da abordagem proposta por Winnemöller, Olsen e Gooch (2006) com valores de parâmetros diferenciados: (A) $\omega=2$ e $\sigma_{rad}=5$; e (B) $\omega=10$ e $\sigma_{rad}=10$	149
Figura 5.8	-	Aplicação da abordagem proposta por Kyprianidis, Kang e Döllner (2009) com valores de parâmetros diferenciados: (A) $\alpha=2$; $N=8$ e $\sigma_r=8$; e (B) $\alpha=10$; $N=8$ e $\sigma_r=18$	150
Figura 5.9	-	Aplicação da abordagem proposta por Gastal e Oliveira (2011) com valores de parâmetros diferenciados: (A) $\sigma_{radial}=25$; $\sigma_{espacial}=0,4$ e $N=3$; e (B) $\sigma_{radial}=60$; $\sigma_{espacial}=1,0$ e $N=21$	152

Figura 5.10	-	Aplicação da abordagem proposta por Kyprianidis e Döllner (2008) com valores de parâmetros diferenciados: (A) $n_a=4$; $n_e=1$ e $n_{bins}=20$; e (B) $n_a=10$; $n_e=10$ e $n_{bins}=7$	153
Figura 5.11	-	Renderizações dos grupos de valores de parâmetros mais bem avaliados: (A) Original; (B) Kuwahara Anisotrópico (G1: $\alpha=1$, $N=3$, $\sigma_r=6$); (C) Bilateral+DOG (G2: $\omega=3\times 3$, $\sigma_{rad}=0,3$); (D) Flow Bilateral+FDoG (G1: $n_a=4$, $N_{bins}=8$); (E) Transformada de Domínio (G1: $\sigma_{rad}=25$, $\sigma_{esp}=0,3$, $N=3$); e (F) Abordagem proposta (G2: $\rho_{cef}=5$, $B=50$, ψ_{bordas} =Welsch).	159
Figura 5.12	-	Gráfico de preferência visual direta.	163
Figura A.1	-	Abordagem proposta, fundamentada na segmentação espaçotemporal de quadros de vídeo.	188
Figura A.2	-	Quadros de entrada.	192
Figura A.3	-	Mean-shift aplicado a quadros consecutivos de um vídeo.	193
Figura A.4	-	Mean-shift topológico aplicado a quadros consecutivos de um vídeo.	194
Figura A.5	-	K-médias aplicado a quadros consecutivos de um vídeo.	194
Figura A.6	-	Cortes normalizados de grafos aplicado a quadros consecutivos de um vídeo.	195
Figura A.7	-	GlobalPb+owt aplicado a quadros consecutivos de um vídeo. ...	195
Figura A.8	-	Mean-shift topológico após a renderização utilizando a abordagem proposta.	196
Figura A.9	-	Representação de vetores de deslocamento em algoritmos de fluxo óptico: (A) Esparso; (B) Legenda para denso; e (C) Denso.	199
Figura A.10	-	Exemplo de imagem usada na avaliação de fluxo óptico (BAKER et al., 2007): (A) Imagem 1; (B) Imagem 2; e (C) Fluxo detectado.	200
Figura A.11	-	Lucas-Kanade/Horn-Schunck, por Bruhn, Weickert e Schnörr (2005): (A) Imagem 1; (B) Imagem 2; e (C) Fluxo detectado..	201
Figura C.1	-	Resultados da aplicação de quantização cromática.	208
Figura C.2	-	Resultados da aplicação de quantização cromática e dithering..	208

LISTA DE SIGLAS E ABREVIATURAS

ANOVA	AN alysis Of VA riance
API	A pplication P rogram I nterface
AVC	A dvanced V ideo C oding
CAD	C omputer- A ided D esign
CAMShift	C ontinuously A daptive M ean- S hift
CEF	C oherence- E nhancing F iltering
CG	C omputação G ráfica
CIE	C ommission I nternationale de L' é clairage
<i>codec</i>	c oder- d ecoder
CPU	C entral P rocessing U nit
DoG	D ifference of G aussians
DSCQS	D ouble S timulus C ontinuous Q uality S cale
DSIS	D ouble S timulus I mpairment S cale
EDISON	E dge D etection and I mage Segmentati ON
ETF	E dge T angent F low
F	F ixation indices (Estatística F)
FDoG	F low-Based D ifference of G aussians
FPGA	F ield- P rogrammable G ate A rray
FPS	F rames P er S econd
GDDR	G raphics D ouble D ata R ate
gl	g raus de l iberdade
GML	G raphics and M edia L ab
GMM	G aussian M ixture M odels
GPU	G raphics P rocessing U nit
HCI	H uman- C omputer I nteraction
HRC	H ypothetical R eference C ircuit
HSV	H ue, S aturation, V alue
IAPS	I nternational A ffective P icture S ystem
IB-AR	I mage- B ased A rtistic- R endering
IC	I nterpolated C onvolution
ITU-R	I nternational T elecommunication U nion R adiocommunication

JPEG	J oint P hotographic E xperts G roup
L.I.V.E.	Laboratory for I mage and V ideo E ngineering
LCD	Liquid- C rystal D isplay
MATLAB	MAT rix LAB oratory
MHI	M otion H istory I mage
MLP	M ulti- L ayer P erceptron
MPEG	M oving P icture E xperts G roup
MQ	M édia dos Q uadrados
MSE	M ean S quare E rror
MSU	M oscow S tate U niversity
NC	N ormalized C onvolution
NEC	N ippon E lectric C ompany
NLM	N on- L ocal M eans
NPR	N on- P hotorealistic R endering
OpenCV	O pen Source C omputer V ision (Intel Library)
PDI	P rocessamento D igital de I magens
RAD	R obust A nisotropic D iffusion
RAM	R andom- A ccess M emory
RC	R ecursive F iltering
RGB	R ed, G reen, B lue
SCACJ	S timulus C omparison A djectival C ategorical J udgement
SQ	S oma dos Q uadrados
TCLE	T ermo de C onsentimento L ivre e E sclarecido
WSJ	W all S treet J ournal
XDoG	e X tended D ifference-of- G aussians

SUMÁRIO

Capítulo 1	Considerações Iniciais.....	14
1.1	Motivação e Caracterização do Problema	16
1.2	Objetivos da Pesquisa	27
1.2.1	Geral	27
1.2.2	Específicos	28
1.3	Metodologia de Pesquisa	28
Capítulo 2	Levantamento Bibliográfico.....	30
2.1	Aspectos Gerais da Área de Renderização Não Fotorrealística....	31
2.2	Breve Histórico da Renderização Não Fotorrealística	35
2.3	Avaliação Subjetiva na área de Renderização Não Fotorrealística	60
2.4	Suavização para a Renderização Não Fotorrealística	83
2.5	Considerações Finais do Capítulo.....	89
Capítulo 3	Abordagem Proposta.....	91
3.1	Abordagem Fundamentada na Suavização por Difusão Anisotrópica	92
3.2	Limitações da Abordagem Proposta	110
3.3	Considerações Finais do Capítulo.....	112
Capítulo 4	Metodologia de Avaliação de Renderizações Não Fotorrealísticas de Vídeos	114
4.1	Planejamento do Experimento.....	117
4.1.1	Seleção dos Critérios a serem Adotados no Experimento.....	117
4.1.2	Seleção dos Vídeos a serem Utilizados no Experimento.....	120
4.1.3	Seleção das Abordagens a serem Comparadas.....	126
4.1.4	Caracterização dos Participantes do Experimento	127
4.2	Elaboração do Material de Teste e Condução da Sessão Experimental	129
4.3	Aspectos relativos à Análise dos Resultados Experimentais.....	132
4.4	Considerações Finais do Capítulo.....	133
Capítulo 5	Apresentação e Discussão dos Experimentos e Resultados ...	135
5.1	Experimentos envolvendo a Abordagem Proposta.....	136
5.1.1	Experimentos relativos à Suavização por Difusão Anisotrópica	137
5.1.2	Experimentos relativos à CEF	141
5.1.3	Experimentos relativos à Sobreposição de Bordas	143
5.1.4	Experimentos relativos à Filtragem NLM.....	144

5.2	Resultados da Avaliação de Renderizações de Vídeos	146
5.2.1	Configuração dos Parâmetros das Abordagens Comparadas.....	146
5.2.2	Delineamento do Perfil dos Participantes do Experimento.....	154
5.2.3	Resultados da Avaliação Comparativa entre Abordagens	155
5.3	Considerações Finais do Capítulo.....	163
Capítulo 5	Considerações Finais	165
6.1	Visão Contextual da Pesquisa Realizada	165
6.2	Contribuições da Pesquisa	168
6.3	Investigações Futuras.....	169
Referências Bibliográficas.....		172
Apêndice A	Experimentos envolvendo Segmentação, Rastreamento de Objetos e Efeitos de Movimento para a Renderização Não Fotorrealística de Vídeos	187
A.1	Abordagem Fundamentada na Segmentação Espaço-temporal de Vídeos.....	187
A.2	Experimentos relativos às Técnicas de Segmentação	192
A.3	Experimentos envolvendo a Computação do Fluxo Óptico.....	196
A.4	Experimentos de Efeitos de Movimento em Vídeo	202
Apêndice B	Resultados Obtidos a Partir da Utilização da Abordagem Proposta.....	204
Apêndice C	Resultados Obtidos a Partir de Técnicas Destinadas à Quantização Cromática.....	207
Apêndice D	Caracterização das Abordagens Comparadas	209
D.1	Abordagem proposta por Winnemöller, Olsen e Gooch (2006).....	209
D.2	Abordagem proposta por Kyprianidis (2009).....	210
D.3	Abordagem proposta por Gastal e Oliveira (2011)	211
D.4	Abordagem proposta por Kyprianidis e Döllner (2008)	211
Apêndice E	Documentação Pré-teste	213
Apêndice F	Questionário Pós-teste	217
Apêndice G	Testes F ANOVA Fator Único Intra-abordagens	219
Apêndice H	Testes de Tukey-Kramer Intra-abordagens	221
Apêndice I	Testes F ANOVA Fator Único Interabordagens.....	223
Apêndice J	Testes de Tukey-Kramer Interabordagens	224
Apêndice K	Testes F ANOVA Fator Único Interabordagens com Maiores Valores Intra-Abordagens	226
Apêndice L	Testes de Tukey-Kramer Interabordagens com Maiores Valores Intra-abordagens	227
Apêndice M	Tempo de Execução da Abordagem Proposta	229
Apêndice N	Detalhamento do Módulo de Detecção de Bordas	231

Capítulo 1

Considerações Iniciais

A renderização *não fotorrealística*¹, usualmente conhecida como NPR (*Non-Photorealistic Rendering*) ou renderização *estilizada*, constitui uma área comum à Computação Gráfica (CG) e ao Processamento Digital de Imagens (PDI) que visa à geração de imagens ou vídeos, os quais se destacam pela expressividade do resultado final. Tais imagens ou vídeos podem exibir características que expressam emoções, apresentando representações visuais flexíveis, obtidas a partir da interpretação arbitrária de características presentes na imagem ou vídeo de entrada, ao invés da adoção total das propriedades realísticas dos objetos da cena (BARILE, CIESIELSKI e TRIST, 2008).

Em contraste à busca por síntese de imagens realísticas, pesquisas na área da renderização não fotorrealística buscam elucidar os princípios e conceitos empregados na concepção de peças de arte tradicionais, tais quais a pintura, ilustração, desenho gráfico, tendo-se como alvo renderizações expressivas e estilizadas (KYPRIANIDIS, 2013).

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; CHUI, 2007; KYPRIANIDIS; DÖLLNER, 2008); (ii) a assistência a desenhistas de histórias em quadrinhos (HWANG et al., 2006); (iii) a animação de estilos clássicos não fotorrealísticos (GOOCH, 2001); (iv) a geração de caricaturas com fins de entretenimento e apelo publicitário (GOOCH; REINHARD; GOOCH, 2004; SHET et al., 2005); (v) a transferência e automatização de estilos de desenho (BARLA et al., 2006); (vi) visualização ilustrativa de dados (YU, 2013); (vii) auxílio na produção de diagramas explodidos (LI et al., 2008); e (viii) renderização estilizada de jogos digitais (VALVE, 2007; GEARBOX, 2009).

¹ Conforme o Dicionário Houaiss da Língua Portuguesa (2011), o termo *realística* refere-se ao que é relativo a ou próprio do realismo, enquanto *realista* refere-se tanto ao que é relativo a ou próprio do realismo quanto ao que é relativo à monarquia. Embora sejam sinônimas, no tocante ao realismo de uma cena, o termo *realística* afigurou-se mais adequado ao propósito da área de pesquisa, por sua conotação única.

Enquanto abstrações da realidade, os cartuns² e as histórias em quadrinhos constituem uma fonte de humor, além de uma forma de representação da realidade, sendo também uma forma versátil de comunicação (BUSSERT, 2005). As histórias em quadrinhos envolvem diferentes elementos, tais como diálogos, onomatopéias, narrativas de momentos épicos ou histórias fantásticas e, até mesmo, o relato de elementos do conhecimento humano, e.g., *Biologia Evolutiva* (EAKIN, 2002), *História* (KENDRICKS, 2000) e *Física* (KAKALIOS, 2002). Chun et al. (2006) ressaltaram que a arte sequencial é estudada e praticada há muito tempo e constitui o gênero mais antigo de renderização não fotorrealística.

Os desenhos animados trazem à vida os estáticos cartuns e histórias em quadrinhos. Wang et al. (2004) propuseram um método para a transformação de um vídeo real em um vídeo ao estilo de cartuns, ressaltando, todavia, o aparecimento de artefatos indesejáveis quando técnicas de NPR adequadas a imagens estáticas são aplicadas a quadros consecutivos de vídeo.

De posse de um vídeo real, sua conversão em um produto não fotorrealístico ao estilo dos cartuns e desenhos animados, duas alternativas possíveis se destacaram, conforme documentadas na literatura especializada, a saber: (i) a conversão do vídeo em uma história em quadrinhos (HWANG et al., 2006; RYU et al., 2008; HONG et al., 2010; SAWADA; TOYOURA; MAO, 2013; XU et al., 2013); e (ii) a conversão do vídeo real em outro vídeo renderizado ao estilo de um cartum (SAUVAGET; BOYER, 2009; KYPRIANIDIS; KANG; DÖLLNER, 2009; BHAT et al. 2010; CRIMINISI et al., 2010; KAGAYA et al., 2010; KYPRIANIDIS, 2013).

Diante do exposto, objetivou-se nesta pesquisa a definição de uma abordagem original para a conversão de vídeos reais em vídeos renderizados de forma não fotorrealística, ao estilo de cartuns. A abordagem foi validada a partir de uma aplicação de *software* capaz de gerar representações não fotorrealísticas de vídeos digitais. Os resultados produzidos pela abordagem proposta foram validados por meio de um processo de votação, inspirado nos estudos realizados por Isenberg et al. (2006), Mandryk, Mould e Li (2011), Winnemöller, Olsen e Gooch (2006), Santella e DeCarlo (2004).

Adicionalmente, também está no âmbito desta pesquisa uma publicação sobre a renderização de esboços de forma não fotorrealística (ARRUDA, QUEIROZ e GOMES, 2012). Tal abordagem para a geração de esboços foi utilizada na presente pesquisa

² Neste documento, preferiu-se adotar a grafia aportuguesada *cartum*, ao invés da grafia *cartoon*, de origem inglesa. Conforme o Dicionário Houaiss da Língua Portuguesa (2001), o termo *cartum* refere-se tanto ao desenho humorístico/caricatural e às histórias em quadrinhos, quanto aos desenhos animados do cinema e às vinhetas cômicas da televisão.

enquanto módulo para a geração do mapa de bordas, a qual é utilizada para guiar um processo de suavização por difusão anisotrópica.

O presente documento está estruturado em seis capítulos. O restante do presente capítulo contempla, na Seção 1.1, as motivações para a pesquisa ora documentada, englobando o desenvolvimento de técnicas para a renderização não fotorrealística que resultaram em produtos empregados fora dos domínios puramente acadêmicos. Na Seção 1.2, são descritos os objetivos da pesquisa, incluindo as hipóteses formuladas. Na Seção 1.3, descreve-se a metodologia de pesquisa utilizada para alcançar o objetivo geral delineado.

O restante deste documento foi estruturado como segue. No Capítulo 2, é apresentada a fundamentação bibliográfica de diferentes técnicas e abordagens destinadas à renderização não fotorrealística de imagens e vídeos, incluindo uma breve perspectiva histórica da área. Ainda no Capítulo 2, apresentam-se conceitos relativos à suavização por difusão anisotrópica, assim como à avaliação subjetiva de imagens e vídeos resultantes de processos de NPR.

A abordagem proposta para a NPR é descrita no Capítulo 3. No Capítulo 4 é apresentada uma metodologia destinada à avaliação de renderização não fotorrealística de vídeos digitais. Os resultados experimentais obtidos a partir da avaliação da abordagem proposta, fundamentada em um processo de votação com usuários humanos, são apresentados no Capítulo 5. Finalmente, no Capítulo 6, são apresentadas as considerações finais, as contribuições advindas da pesquisa realizada e atividades sugeridas para investigações futuras.

1.1 Motivação e Caracterização do Problema

Como vertente ativa na área do Processamento Digital de Imagens, a renderização não fotorrealística (NPR) visa à geração de imagens simples e compreensíveis a partir de cenas complexas (MARKOSIAN, 2000). Mais precisamente, NPR reúne uma série de técnicas que buscam a comunicação visual e a ilustração conceitual, ao invés da fidelidade à realidade (SAYEED; HOWARD, 2006), envolvendo conceitos artístico-científicos e sendo guiada mais pela percepção humana do que pela precisão física (GOOCH, 2001). Adicionalmente, Santella (2005) afirmou que NPR é uma área excessivamente ampla, posto que pode englobar qualquer técnica direcionada à criação de imagens que não sejam imitações fidedignas da realidade.

Gooch et al. (2010) afirmaram que na área de CG e na área de visualização estilizada, renderização refere-se ao processo de conversão de dados em imagens. A renderização fotorrealística denota a geração de imagens fundamentadas em

simulação de aspectos da realidade, do mundo físico, tendo com objetivo a criação de imagens indistinguíveis de fotografias de cenas do mundo real. Em contrapartida, a renderização não fotorrealística denota a geração de imagens por meio de processos artísticos, limitadas por restrições estéticas, ao invés de simulações de aspectos físicos do mundo real.

Kyprianidis (2013) afirmou que, nos dias atuais, a área de NPR expandiu-se enquanto área ativa de pesquisa, englobando a pesquisa relativa à técnicas que resultam em renderizações de estilos variados, tais quais a renderização de objetos tridimensionais (LEE; KWON; LEE, 2006; DECARLO; RUSINKIEWICZ, 2007), diagramas explodidos (LI et al., 2008), falsa cor (POULI; REINHARD, 2011), pintura (ZENG et al., 2009) e renderização com redução da paleta de cores (XU; KAPLAN, 2008).

Borgo et al. (2011) apresentaram uma revisão abrangente de técnicas destinadas à transformação de vídeos em representações artísticas ou não fotorrealísticas, contemplando também outros métodos que visam à criação de sumários e representações resumidas de vídeos. Os autores propuseram uma taxonomia, fundamentada no objetivo ou tipo de vídeo de saída gerado, com fins à classificação dos conceitos e técnicas de abstração ou representações de vídeos. Adicionalmente, os autores afirmaram que, até recentemente, vídeos têm sido utilizados enquanto meio para a apresentação de resultados. Contudo, a produção de uma quantidade crescente de vídeos digitais suscitou a criação de algoritmos de Computação Gráfica e Visualização Computacional que recebem vídeos como entrada, os manipula e converte em diferentes formas de saída, de acordo com o propósito a que se destinem (e.g., apresentação artística, composição, seleção de quadros-chave).

Em geral, a manipulação e a conversão de vídeos digitais de entrada visam à geração de resultados (imagens ou vídeos) criativos, destinados à apreciação e ao entretenimento. Destacam-se duas vertentes: (i) a *reapresentação de vídeos*, que se fundamenta na manipulação de primitivas geométricas (e.g., forma e distribuição de objetos) ou de atributos ópticos (e.g., cor e iluminação) para a produção de um vídeo de saída que captura a essência do vídeo de entrada, de uma forma artística ou expressiva, e.g., simulando um desenho animado; e (ii) a *modelagem de cenas*, que envolve tipicamente a reconstrução tridimensional da cena capturada, na qual o modelo de interesse é manipulado ou combinado com outros modelos, a fim de ser processado de forma similar aos modelos 3D sintéticos. O presente trabalho enquadra-se na **primeira categoria**, i.e., ***reapresentação de vídeos***.

Borgo et al. (2011) destacaram ainda que o impulso para a transformação de vídeos deveu-se ao aumento do emprego de aplicações multimídia, em especial as aplicações destinadas ao consumidor final, incluindo filmes e jogos digitais. Os autores ressaltaram que a visualização de um vídeo tem como propósito a criação de novas representações visuais de elementos-chave apresentados no vídeo de origem, de modo a capturar, extrair e sumarizar, no vídeo de saída, eventos e características importantes do vídeo de entrada. Tal visualização não tem como propósito principal a automatização de decisões fundamentadas no conteúdo do vídeo de origem, mas assistir ao usuário na difícil tarefa de se ater à informação relevante que os vídeos, em geral, contêm.

Borgo et al. (2011) afirmaram que o grande desafio em se aplicar diretamente técnicas de abstração de imagens (NPR), quadro-a-quadro, em vídeos digitais, é a manutenção da coerência temporal. Tal visão também é compartilhada por Wang et al. (2004), Collomosse, Rowntree e Hall (2005), Gomes et al. (2009) e Bénard, Bousseau e Thollot (2011).

A coerência temporal refere-se à exibição correta de elementos presentes em vídeos estilizados, tais como pinceladas artísticas ou contornos bem delineados dos objetos da cena (GOMES; SANTOS; CARVALHO, 2007). De acordo com Gomes et al. (2009), a falta de coerência temporal resulta em artefatos indesejáveis no resultado da renderização, e.g., o efeito *shower-door* e o efeito de cintilação (*flickering*). O efeito *shower-door* (GOMES; SANTOS; CARVALHO, 2007 *apud* MEIER, 1996) diz respeito à visualização da animação como se esta ocorresse através de um vidro texturizado e os objetos retratados afigurarem-se dissociados da textura aplicada ao tema renderizado. O efeito de cintilação (*flickering*) ocorre quando o mesmo objeto é retratado de forma diferente, mesmo em quadros adjacentes, e.g., as pinceladas que retratam uma forma geométrica mudam bruscamente entre quadros adjacentes. O efeito de cintilação pode surgir devido a mudanças na iluminação da cena ou decorrente de falhas no processo de segmentação utilizado.

É importante ressaltar que uma taxonomia diferente daquela apresentada por Borgo et al. (2011) encontra-se na revisão abrangente da área de NPR feita por Kyprianidis et al. (2013). Tal taxonomia agrupa técnicas de NPR de duas formas, sejam estas agrupadas por tipos de algoritmos empregados (e.g., filtragem de imagem com preservação de bordas, segmentação, filtragem morfológica) ou agrupadas por tipo de mecanismo (e.g., visão computacional, técnicas que empregam algoritmos gulosos, minimização de uma função objetivo) utilizado. Tal taxonomia afigurou-se mais adequada aos propósitos da pesquisa descrita neste documento.

Adicionalmente, a referida taxonomia é mais recente (2013) do que a classificação proposta por Borgo et al. (2011). A pesquisa ora documentada insere-se no contexto da taxonomia proposta por Kyprianidis et al. (2013) na categoria: *Renderização fundamentada no processamento e filtragem de imagens* → *Pertencente ao domínio espacial* → *Fundamentada no conteúdo das regiões* → *Usando difusão anisotrópica*. Detalhes da classificação proposta por Kyprianidis et al. (2013) são apresentados no Capítulo 2.

Ressalta-se, como motivação para esta pesquisa, 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* (SOHN, 1999), *Waking Life* (FOX, 2001) e *A Scanner Darkly* (WARNER, 2006). Além das indústrias cinematográficas e de jogos eletrônicos, há igualmente potencial de aplicação da renderização não fotorrealística de vídeos às áreas de propaganda e *marketing*, conforme retratado nos seguintes exemplos: anúncio de vinho³, vídeo promocional⁴, videoclipe⁵ e propaganda de banco⁶, dentre outros.

Cabe ainda ressaltar que na maioria dos exemplos supracitados foi requerido algum nível de intervenção humana nos processos de renderização não fotorrealística, o que reafirma a importância do desenvolvimento de métodos automáticos que produzam vídeos com maior apelo estético. A seguir, são destacados alguns produtos nos quais foram aplicadas técnicas de NPR, os quais resultaram de iniciativas puramente acadêmicas.

Ressalva-se que os exemplos apresentados a seguir são tão somente um excerto das várias aplicações que utilizam técnicas de NPR. Com o intuito de ampliar a definição de NPR, a fim de englobar tais produtos, Winnemöller (2013) afirma que: a renderização não foto realística pode ser definida enquanto área que assiste, com o uso de ferramentas computacionais, a síntese, criação e reprodução de manifestações artísticas. Ainda que muito ampla, esta definição engloba as áreas nas quais são ressaltados os usos de NPR: (i) entretenimento, por meio de filmes e jogos eletrônicos; (ii) técnicas de visualização e re-apresentação da informação; e (iii) ferramentas de NPR para dispositivos móveis. Nesse contexto, a definição de produto, contida nesta seção, engloba programas de computador comercializados para profissionais ou não-profissionais, jogos eletrônicos digitais e filmes produzidos para o cinema.

³ <http://www.youtube.com/watch?v=JxJmRDXjPTo>

⁴ <http://www.youtube.com/watch?v=QUiM1YQKdMm>

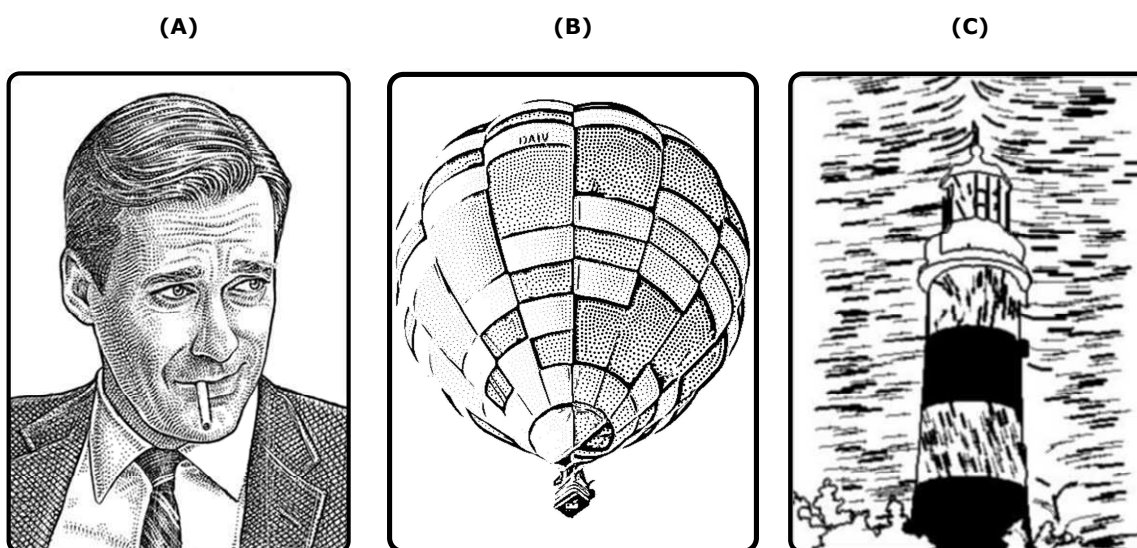
⁵ http://www.youtube.com/watch?v=_QBISF26Kd4

⁶ <http://www.youtube.com/watch?v=K7IFSiqxsVI>

Winnemöller (2013) afirmou que a reprodução automatizada ou assistência ao usuário na produção de estilos artísticos configura um dos possíveis usos da NPR. Um artista pode usar uma ferramenta computacional como auxílio à criação manual de uma peça artística, a um custo menor e em menor espaço de tempo. Uma animação ao estilo de cartuns pode demandar do artista a produção de 12 a 24 quadros manualmente pintados, resultando em cerca de 1 segundo de vídeo. A conversão automatizada de um filme com atores humanos ou a conversão de uma animação 3D para o estilo cartum pode se beneficiar de técnicas de NPR, auxiliando o artista a reduzir o tempo de produção, conforme proposto por Lake et al. (2000), Nienhaus e Döllner (2004), Winnemöller, Olsen e Gooch (2006), Bhowmick e Klette (2013).

Winnemöller (2013) destacou a pesquisa feita no âmbito da NPR para a criação de imagens aos estilos pontilhista e de entalhe, nos quais se destacam Melo et al. (2007). No contexto do estilo pontilhista, há o estilo *hedcut*, nome dado a um estilo tradicional que é marca distinta e registrada do *Wall Street Journal* (WSJ), documentada em publicações relativamente recentes (e.g., KIM et al., 2008; SON et al., 2011). Três exemplos, um de figura produzida por um artista ao estilo pontilhista *hedcut*, outro de imagem renderizada por uma técnica NPR e um terceiro de imagem renderizada a partir da técnica de entalhe podem ser visualizados na Figura 1.1(A), (B) e (C), respectivamente. Destaca-se a presença, na figura produzida ao estilo *hedcut*, de irregularidades nos pontos postos no desenho, os quais conferem aspecto diferenciado do resultado que apresenta textura com pontos regulares quando a abordagem proposta por Kim et al. (2008) foi utilizada.

Figura 1.1 – Exemplos de pontilhismo e entalhe: (A) Don Draper, ao estilo *hedcut*; (B) Resultado obtido por Kim et al. (2008); e (C) Resultado obtido por Melo et al. (2007).

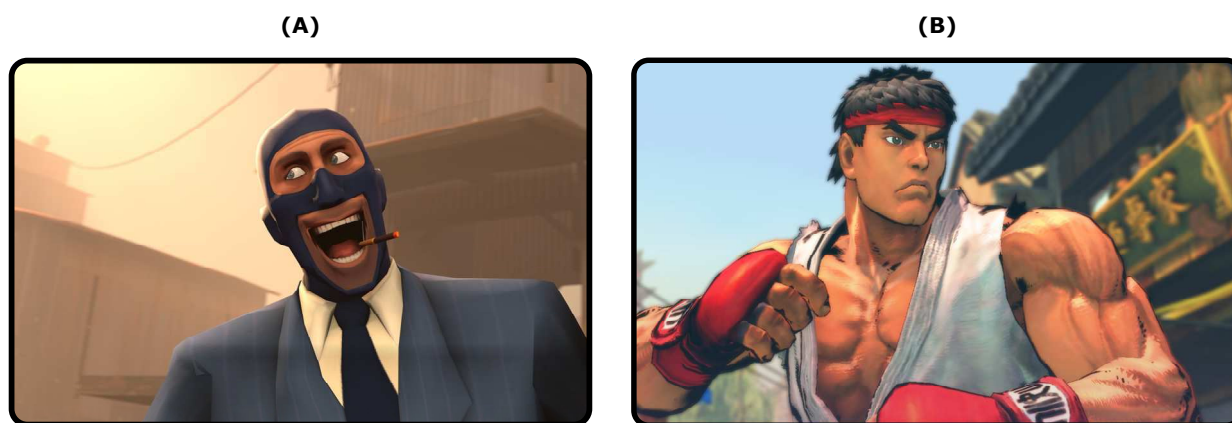


Winnemöller (2013) ressaltou que a indústria do cinema e a indústria de jogos

eletrônicos digitais têm constatado que as técnicas empregadas na área de NPR constituem um fator de diferenciação na produção do conteúdo. O autor ainda afirmou que houve um amadurecimento na área da Computação Gráfica, em conjunção com o barateamento dos custos com *hardware*. Tais fatores levaram as empresas produtoras de jogos eletrônicos digitais a competir para oferecer os melhores efeitos realísticos, tais como o desfoque resultante do movimento (*motion-blur*), as sombras e penumbras (*soft-shadows*), os efeitos de oclusão no ambiente (*ambient occlusion*) e a translucência subsuperficial (*sub-surface scattering*).

Entretanto, Winnemöller (2013) afirmou que há, atualmente, saturação na quantidade de avanços em técnicas que visam ao fotorrealismo nos motores de renderização de jogos eletrônicos (*game engines*), fazendo com que tais avanços sejam cada vez mais sutis. Devido a isto, alguns produtores de jogos eletrônicos têm diferenciado suas criações a partir do uso da renderização estilística, em especial efeitos de *cel shading*⁷: *Team Fortress 2* (VALVE, 2007), *Grand Theft Auto V* (ROCKSTAR, 2013), *Street Fighter 4* (CAPCOM, 2008), *XIII* (UBISOFT, 2003), *Prince of Persia* (UBISOFT, 2008), *Borderlands* (GEARBOX, 2009). Dois exemplos de jogos digitais que utilizam renderização estilística podem ser vistos na Figura 1.2.

Figura 1.2 – Exemplos de renderizações estilísticas em jogos eletrônicos: (A) Team Fortress 2 (VALVE, 2007); e (B) Street Fighter IV (CAPCOM, 2008).



O estilo cartum, no contexto considerado anteriormente, é prevalente nas técnicas usadas pela indústria de jogos eletrônicos e de filmes (WINNEMÖLLER, 2013), pois permite que a coerência temporal seja atingida com a aplicação da mesma técnica em quadros consecutivos, um fator importante em jogos digitais, nos quais a pré-renderização não é possível, já que as ações dos jogadores não podem ser previstas e renderizadas de antemão.

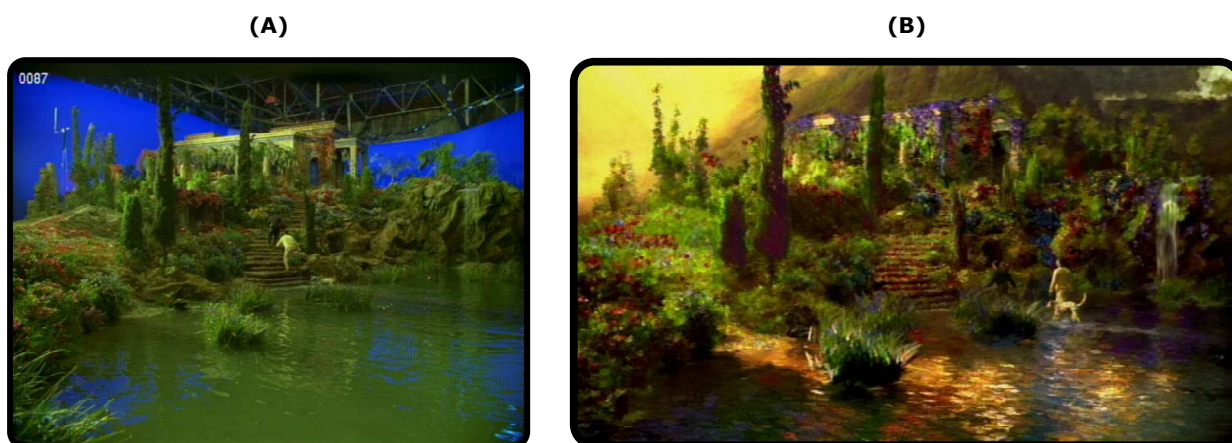
⁷ Conjunto de técnicas empregadas na renderização de imagens tridimensionais, onde são empregados contornos escuros ao redor dos objetos da cena, texturas simplificadas, iluminação com menos cores, de modo que a cena retratada se assemelhe ao de desenhos animados.

Apesar de tal prevalência, Winnemöller (2013) ressaltou que o primeiro filme longa metragem a usar ferramentas de NPR foi o filme *Amor além da Vida* (POLYGRAM, 1998), no qual o personagem interpretado pelo ator Robin Williams cria um mundo de sonhos na pós-vida, cujas cenas lembram uma pintura expressionista, tendo sido renderizadas com um auxílio de um sistema (*MotionPaint*) fundamentado nas pesquisas realizadas por Litwinowicz (1997). A cena filmada com atores reais formava a etapa inicial de processamento. A cena inicial recebia as pinceladas artísticas que formavam a cena renderizada, as quais eram movidas de acordo com o fluxo óptico advindo da cena original.

O sistema fundamentado na pesquisa de Litwinowicz (1997) facilitou o trabalho da equipe artística atuante no filme, cujos artistas beneficiaram-se do auxílio do mecanismo de fluxo óptico para que as pinceladas fossem propagadas para os quadros seguintes, ao invés de terem que pintar, meticulosa e manualmente, cada quadro da cena retratada.

Entretanto, o sistema não era completamente automatizado e o método para a obtenção de fluxo óptico era imperfeito (métodos de fluxo óptico atuais ainda possuem imperfeições, de acordo com Baker et al., 2007). Ressalta-se ainda que foram empregados *lasers* para traçar e mapear o ambiente no qual a ação com atores reais foi filmada, o que gerou um modelo tridimensional da cena. A junção da informação obtida a partir da utilização dos *lasers*, o fluxo óptico ao longo dos quadros da cena e os marcadores posicionados no cenário ajudaram no alinhamento da cena filmada com o modelo tridimensional, então empregado para auxiliar a renderização estilística da cena. Um excerto da cena original, assim como um excerto da cena estilizada, pode ser visto na Figura 1.3.

Figura 1.3 – Excerto de cena do filme *Amor além da vida*: (A) Imagem do cenário antes do processo de estilização; (B) Cena estilizada ao estilo expressionista.



A técnica de NPR proposta por Litwinowicz (1997) fornecia aos diretores de arte

um método para a especificação de parâmetros locais de cada pincelada e cada pincelada era propagada nas cenas subsequentes, mediante o emprego do fluxo óptico. A especificação de parâmetros locais para cada pincelada reduzia o trabalho necessário para renderizá-las no quadro seguinte. A propagação das pinceladas para quadros subsequentes reduziu a intervenção manual necessária em cada quadro do filme. O filme recebeu um prêmio oscar pelos efeitos visuais empregados, em 1998. Winnemöller (2013) argumentou que tal feito deveu-se, em grande parte, às cenas renderizadas ao estilo expressionista.

Rosin e Collomosse (2013) destacaram outra utilização de NPR, para além da produção e de entretenimento, que é a visualização de dados (YU, 2013) ou discussão de idéias em estágios iniciais de um projeto arquitetônico. Técnicas modernas de NPR permitem que projetistas inovem ao usar renderizações CAD, as quais podem beneficiar a comunicação e exploração de idéias nos estágios iniciais do desenvolvimento do projeto. A aplicação de técnicas de NPR pode levar a produção de vários estilos de renderizações, e tal produção seria onerosa se fosse realizada de forma manual. Neste sentido, Winnemöller (2013) destacou três produtos que realizam tais renderizações, descritos brevemente a seguir.

O *Piranesi*⁸ é um produto que tem como fundamento a pesquisa realizada por Richens (1999), tendo sido um dos primeiros sistemas comercialmente disponíveis que disponibilizou técnicas de NPR para arquitetos, aplicando renderizações estilizadas a projetos CAD.

O *PaletteCAD*⁹ (PALETTECAD, 2013) é um sistema focado no planejamento de *design* de interiores que integra o desenvolvimento de projetos CAD com módulos de renderização artística. O sistema oferece modos pré-definidos de renderização NPR, incluindo esboços compostos apenas por bordas, efeito de esboços feitos com caneta de feltro, efeito de aquarela (fundamentado na pesquisa de Luft et al., 2008).

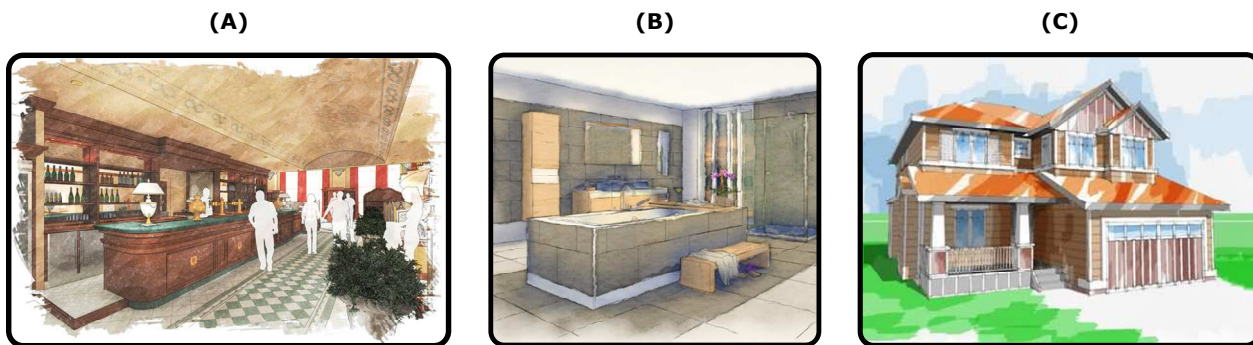
Por fim, o *SketchUP* é um sistema desenvolvido pela Google (SKETCHUP, 2000), com foco na modelagem simples e intuitiva de espaços tridimensionais. Em contraponto a renderizações sofisticadas de CAD, o sistema produz, nativamente, renderizações estilizadas.

Exemplos das renderizações produzidas pelos três sistemas supracitados são ilustrados na Figura 1.4.

⁸ www.piranesi.co.uk

⁹ www.palettacad.com

Figura 1.4 – Exemplos de renderizações que modificam projetos CAD: (A) Piranesi¹⁰; (B) PalletteCAD; e (C) Google SketchUP.



Winnemöller (2013) ressaltou um exemplo de aplicação de técnicas de NPR em um produto de *software* para dispositivos móveis: o ToonPAINT (2011). O aplicativo usa técnicas de NPR para renderizar estilisticamente imagens digitais, transformando-as em renderizações ao estilo de cartuns. A tecnologia subjacente ao aplicativo teve como fundamento as pesquisas realizadas por Winnemöller, Olsen e Gooch (2006) e por Kyprianidis e Döllner (2008). A transposição dos fundamentos da componente de pesquisa a um aplicativo comercial, aliada à retroalimentação de usuários do aplicativo aos pesquisadores, revela o interesse comercial no tocante à pesquisa na área de NPR.

Na pesquisa realizada por Winnemöller, Olsen e Gooch (2006) para a abstração de imagens e vídeos digitais, as imagens renderizadas são produzidas a partir de filtragem com o filtro bilateral (TOMASI; MANDUCHI, 1998), utilizado em conjunção com pseudoquantização do canal de luminância, sobrepostos pelas bordas detectadas pelo filtro DoG (*Difference of Gaussians*). O sistema utiliza recursos disponíveis em placas de processamento gráficas dedicadas (GPU – *Graphics Processing Unit*), que permitiram a renderização de vídeo em tempo real. Uma extensão da pesquisa destes autores foi proposta por Kyprianidis e Döllner (2008), na qual o detector de bordas originalmente sugerido foi substituído por um modelo de Diferença de Gaussianas Baseada no Fluxo (FDoG - *Flow-Based Difference of Gaussians*). O detector fundamenta-se na computação do fluxo das bordas, empregando e suavizando um campo de vetores (ETF - *Edge Tangent Flow*).

Com efeito, a pesquisa de Winnemöller, Olsen e Gooch (2006) e as extensões posteriores propostas por Kyprianidis e Döllner (2008) influenciaram a pesquisa relatada nesta tese, de modo que seu detalhamento é apresentado no Capítulo 2.

Winnemöller (2013) ressaltou que o aplicativo *ToonPAINT* foi proposto para usuários com habilidades artísticas limitadas, fator que ampliou a quantidade de

¹⁰ picasaweb.google.com/PiranesiGallery/PiranesiGallery - vista bar by emanuele

possíveis usuários. O autor ainda ressaltou que o sistema foi concebido para que o usuário pudesse manter certo grau de controle sobre o resultado apresentado, i.e., não foi desenvolvida uma solução completamente automática. Empregando técnicas de NPR, o aplicativo em questão apresenta as formas, silhuetas e sombreamento do objeto presentes na imagem de entrada. O sombreamento da imagem pode ser ajustado de acordo com parâmetros da técnica XDoG (WINNEMÖLLER; OLSEN; GOOCH, 2006; WINNEMÖLLER, 2011). As cores das renderizações produzidas pelo *ToonPAINT* não são atribuídas automaticamente, cabendo ao usuário atribuir as cores desejadas às regiões da imagem renderizada mediante a tela sensível ao toque do dispositivo móvel, comportamento semelhante àquele adotado em livros infantis para colorir.

Adicionalmente, outro produto computacional que também usa técnicas de renderização estilística é o *plug-in PixelBender Toolkit*¹¹, especificamente desenvolvido para o programa Adobe Photoshop, o qual também incorpora filtros destinados à renderização estilística. Por meio do filtro *Oilpaint*, tal *plug-in* destina-se a estilizar imagens ao estilo de pintura a óleo. Há um grupo do Yahoo Flickr especialmente dedicado à exposição dos resultados obtidos por tal *plug-in*¹².

Diante do exposto, ressaltaram-se ao longo desta seção os impactos causados por técnicas de NPR na produção de peças de entretenimento e em outros cenários específicos. Adicionalmente, os esforços na área de NPR concentraram-se, ao longo do tempo, na replicação de estilos artísticos que demandariam muita habilidade do usuário não profissional, tais como pintura a óleo ou aquarela.

Assim sendo, a pesquisa ora documentada focaliza a rerepresentação do conteúdo original do vídeo de forma não fotorrealística. Na revisão bibliográfica realizada, contida no Capítulo 2, duas abordagens destacam-se na pesquisa de NPR, as quais se fundamentam no uso de: (i) algoritmos de segmentação espaçotemporal do vídeo de entrada; e (ii) algoritmos de suavização dos quadros do vídeo de entrada. Inicialmente, foram pesquisados algoritmos de segmentação espaçotemporal de vídeos, os quais não produziram resultados satisfatórios, nem do ponto de vista subjetivo, nem do ponto de vista da coerência temporal. Os experimentos fundamentados no emprego da segmentação espaçotemporal estão relatados no Apêndice A.

A pesquisa que se seguiu focalizou-se na abstração de vídeos, fundamentada na

¹¹ <http://www.adobe.com/devnet/pixelbender.html>

¹² <http://www.flickr.com/groups/1558914@N23/>

suavização de seus quadros, com a preservação das bordas das regiões que os compõem. As técnicas de suavização de imagens e vídeos com preservação de bordas possibilitam a remoção de detalhes em regiões de baixo contraste da imagem/quadro, sem afetar as descontinuidades, de modo a preservar a estrutura geral da imagem ou dos quadros de vídeo. Tais técnicas permitem a suavização das cores presentes no interior das regiões que compõem a imagem ou os quadros de vídeo, enquanto preservam as cores e descontinuidades existentes entre tais regiões.

Dentre as abordagens para a abstração de imagens ou quadros de vídeo, destacaram-se, durante a pesquisa realizada, a suavização por difusão anisotrópica (PERONA; MALIK, 1990) e a suavização a partir do uso do filtro bilateral (TOMASI; MANDUCHI, 1998). Tais filtros de suavização foram propostos com o intuito de remover ruído em imagens fotorrealísticas, embora haja aplicabilidade de tais filtros na obtenção de imagens abstraídas. Fundamentado nestas observações, assim como em resultados documentados na literatura (WINNEMÖLLER; OLSEN; GOOCH, 2006; KYPRIANIDIS; DÖLLNER, 2008; KANG; LEE; CHUI, 2009), destaca-se a importância da suavização de imagens ou vídeos, com preservação de bordas, em especial de melhorias que podem ser incorporadas às técnicas de suavização por difusão anisotrópica, que constitui um tópico de pesquisa importante na área de abstração de imagens.

Apesar de toda a pesquisa já desenvolvida até então, a geração de representações não fotorrealísticas, na forma de vídeos digitais, de forma automática, temporalmente coerente, ainda constitui um problema cuja solução não foi esgotada. Assim sendo, Kyprianidis, Kang e Döllner (2009) relataram dificuldades na abstração de regiões de imagens que se apresentam muito texturizadas, enquanto Kyprianidis et al. (2013) afirmaram que ainda não há um método geral que apresente boa coerência temporal para uma gama variada de vídeos de entrada.

Assim sendo, a solução proposta neste documento fundamenta-se, primordialmente, na combinação de técnicas destinadas à abstração de vídeos, assim como na validação dos resultados a partir da realização de investigações experimentais envolvendo usuários, dada a escassez de experimentos dessa natureza na literatura pesquisada, em especial com enfoque à avaliação de vídeos renderizados.

Dificuldades e desafios associados à pesquisa devem-se aos seguintes fatores: (i) variações nas condições de iluminação do vídeo de entrada; (ii) requisitos computacionais elevados para a renderização de vídeos; (iii) dificuldades no ajuste de

parâmetros das técnicas propostas que atendam a uma ampla gama de vídeos de entrada; e (iv) dificuldades no processo de avaliação subjetiva dos resultados produzidos.

1.2 Objetivos da Pesquisa

A partir da motivação apresentada, da ênfase dada a produtos que utilizam técnicas de renderização não fotorrealística e da pesquisa da renderização estilística documentada por outros autores, em especial aquela que delinea desafios de pesquisa a serem elucidados, pôde-se formular a seguinte questão de pesquisa: *é possível gerar, de forma automática, renderizações estilizadas que têm como ponto de partida imagens ou vídeos digitais, de modo que os resultados sejam mais bem avaliados do que resultados produzidos por métodos atualmente existentes?*

O enfoque da pesquisa ora documentada deu-se na renderização de vídeos digitais, na qual a aplicação da técnica proposta foi realizada quadro-a-quadro. É importante ressaltar que a renderização estilística de apenas uma imagem digital é factível na pesquisa proposta e também é englobada pela questão de pesquisa apresentada.

A seguir, apresentam-se as hipóteses que se pretendeu confirmar ou refutar, à luz da questão de pesquisa apresentada:

1. A combinação da filtragem anisotrópica com a ponderação por um detector de bordas neural, em uma abordagem destinada à renderização não fotorrealística de vídeos gera resultados que são mais bem avaliados por um conjunto de usuários, quando comparados a resultados de técnicas concorrentes;

2. Os resultados gerados pela abordagem proposta para a renderização não fotorrealística contêm artefatos espúrios ou cintilação no vídeo resultante que são equivalentes àqueles produzidos por abordagens encontradas na literatura da área;

3. A abordagem proposta produz resultados que possuem melhor definição de bordas e de estilização, quando comparados a abordagens encontradas na literatura da área.

A partir destas hipóteses, foram definidos os objetivos geral e específico desta pesquisa, apresentados a seguir.

1.2.1 Geral

Propor e validar uma abordagem original destinada à abstração de vídeos digitais, que possibilite a geração de vídeos renderizados não fotorrealisticamente, ao estilo de cartuns.

1.2.2 Específicos

A abordagem proposta tem como diferencial a agregação de módulos que incorporam alterações na abordagem típica, fundamentada na suavização de quadros de vídeo por difusão anisotrópica (PERONA; MALIK, 1990; BLACK et al., 1998), mais especificamente no que diz respeito: (i) à alteração da equação de parada-na-aresta da técnica de suavização por difusão anisotrópica original (PERONA; MALIK, 1990); (ii) ao uso de um detector de bordas proposto por Arruda, Queiroz e Gomes (2012) para estilizar e orientar a suavização por difusão anisotrópica modificada; (iii) à combinação das abordagens de detecção de bordas e suavização, visando a reduzir o número de iterações necessárias para a obtenção de quadros do vídeo abstraído; e (iv) à combinação original e modular de um arcabouço destinado à produção de imagens e vídeos renderizados de forma não fotorrealística.

Fundamentados no objetivo geral, destacam-se os seguintes objetivos específicos:

- Investigar métodos e técnicas correntes destinadas à renderização estilística de vídeos reais em vídeos renderizados de forma não fotorrealística;
- Desenvolver melhorias nos algoritmos para a conversão de vídeos reais em vídeos ao estilo de cartuns;
- Propor uma abordagem destinada à conversão de vídeos reais em vídeos renderizados de forma não fotorrealística que incorpore os algoritmos investigados e/ou propostos;
- Desenvolver uma aplicação de software fundamentada na abordagem proposta, capaz de gerar vídeos renderizados não fotorrealisticamente e temporalmente coerentes, a partir de vídeos digitais; e
- Validar os resultados produzidos pela aplicação de software implementada, mediante um processo de avaliação subjetiva, fundamentada em experimentos envolvendo usuários.

1.3 Metodologia de Pesquisa

A metodologia adotada na presente pesquisa é descrita a seguir. Inicialmente, realizou-se uma revisão bibliográfica na área de renderização não fotorrealística, envolvendo tanto a renderização de imagens quanto a renderização de vídeos, dando enfoque à classificação e delimitação da pesquisa realizada, acrescido de uma breve perspectiva histórica da área de renderização não fotorrealística.

Após a revisão bibliográfica, foram definidas duas abordagens para o cumprimento dos objetivos propostos na Seção 1.2. A primeira abordagem fundamentou-se na segmentação espaçotemporal de vídeos e, após a realização de experimentos, a mesma foi descartada, em virtude dos resultados obtidos, os quais foram considerados insatisfatórios, conforme relatado no Apêndice A. A segunda abordagem, fundamentada na suavização de vídeos por difusão anisotrópica, levou em consideração inovações desejáveis às técnicas revisadas na literatura, tendo apresentado melhores resultados visuais em relação à proposta anterior. Portanto, tal abordagem foi alvo de melhorias/modificações, configurando assim a abordagem proposta para renderização de vídeos digitais, conforme relatado no Capítulo 3.

Em seguida, implementaram-se os módulos que compõem a aplicação de *software*, fundamentados na abordagem proposta. Tais módulos foram desenvolvidos separadamente e, posteriormente, integrados em um sistema. Após a integração dos módulos, conjuntos de valores de parâmetros foram testados, tendo sido o sistema calibrado. Os testes e parâmetros utilizados na etapa de calibração também foram alvos de avaliação.

A metodologia de avaliação proposta nesta pesquisa é detalhada no Capítulo 4. O detalhamento dos resultados experimentais obtidos, assim como sua comparação com resultados produzidos por outras técnicas de renderização não fotorrealística de imagens e vídeos, são apresentados no Capítulo 5. No Capítulo 6, são apresentadas as considerações finais e as proposições para a realização de pesquisas futuras.

Capítulo 2

Levantamento Bibliográfico

O presente capítulo tem como propósito-chave a exposição e discussão, no âmbito do Processamento Digital de Imagens, das abordagens mais recentes empregadas na renderização não fotorrealística. Em função da análise das abordagens estudadas, intentou-se propor uma abordagem original para a conversão de vídeos reais em vídeos renderizados não fotorrealisticamente ao estilo de cartuns. Sumaria-se neste capítulo a fundamentação para a abordagem ora proposta, a qual será descrita no Capítulo 3.

Visando a facilitar o entendimento da área na qual a pesquisa realizada se insere, este capítulo encontra-se dividido nas seguintes seções:

- Seção 2.1: são apresentados os aspectos gerais, o escopo da pesquisa realizada e uma classificação taxonômica da área de renderização não fotorrealística.

- Seção 2.2: é apresentada uma progressão temporal da área de renderização não fotorrealística, na qual são ressaltadas as pesquisas (artigos científicos e livros publicados) que influenciaram, de algum modo, a abordagem proposta. Tais pesquisas são apresentadas em ordem cronológica, visando a facilitar o acompanhamento dos avanços relativos à área de pesquisa. Nesta seção, também se detalham algumas abordagens que foram alvo de comparação experimental, inclusive sua arquitetura ou exemplos de resultados obtidos a partir de sua utilização.

- Seção 2.3: é apresentado um levantamento bibliográfico de métodos para a avaliação de resultados de abordagens destinadas à renderização não fotorrealística. A apresentação de tal levantamento foi feita separadamente da progressão temporal apresentada na seção 2.2, visando à focalização exclusiva em estratégias adotadas na avaliação de resultados de NPR, o que serviu de fundamento para a composição do processo de avaliação apresentado nesta tese.

- Seção 2.4: é apresentada uma fundamentação dos conceitos relativos à suavização por difusão anisotrópica, em especial a apresentação de funções de

parada-na-aresta e de conceitos relativos às modificações inseridas na abordagem proposta.

2.1 Aspectos Gerais da Área de Renderização Não Fotorrealística

Kyprianidis et al. (2013) forneceram uma classificação taxonômica que engloba técnicas de renderização não fotorrealísticas, cuja área de abrangência engloba a *renderização artística baseada em imagens (IB-AR - Image-Based Artistic-Rendering)*. Tal área incorpora técnicas para a renderização de imagens estáticas, por focar na estilização de conteúdo bidimensional, tais como quadros individuais de vídeo digital ou imagens digitais, a qual se contrapõe à estilização de conteúdo tridimensional, além de também incorporar a estilização artística, em uma subárea da renderização expressiva para a comunicação visual.

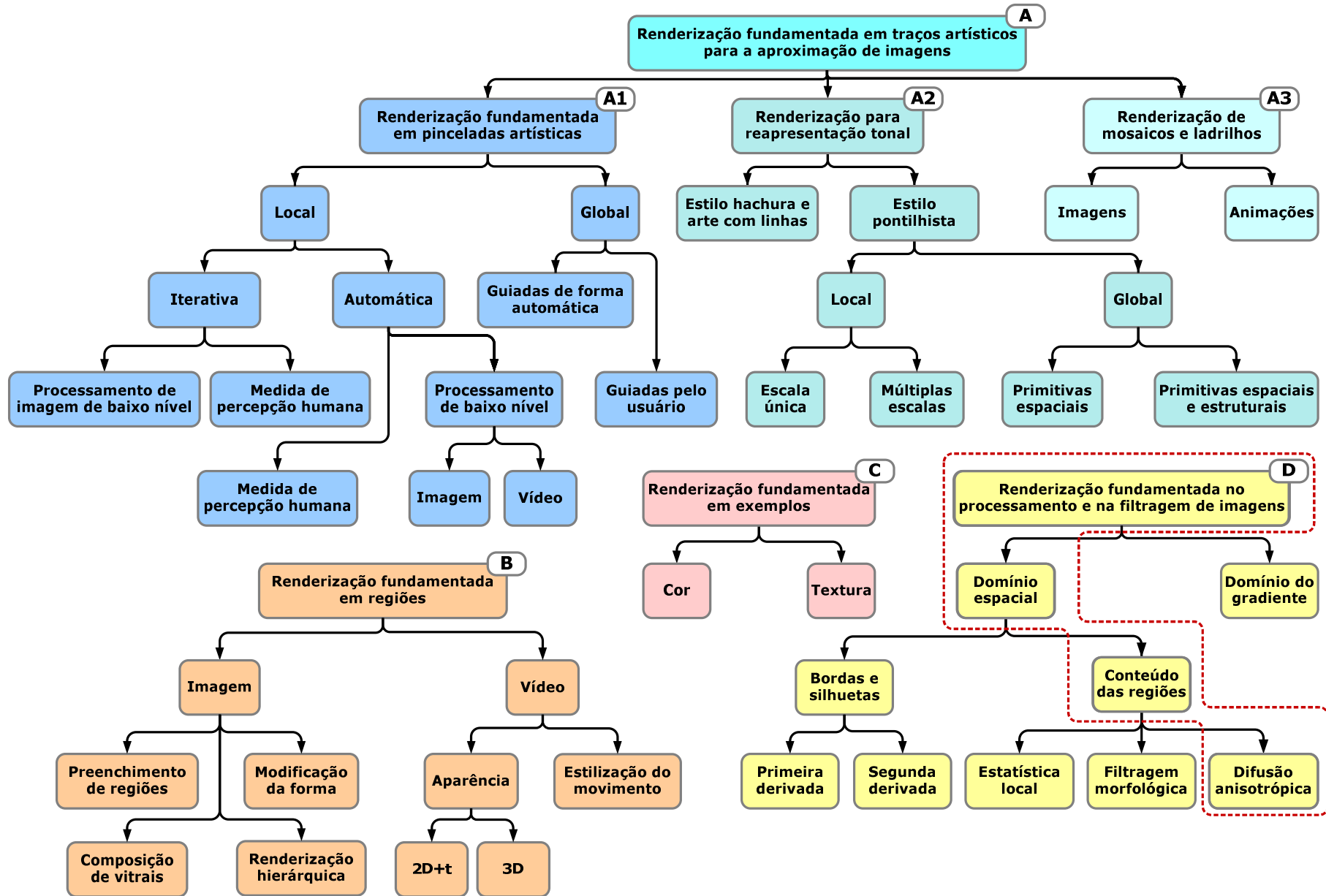
A classificação taxonômica ora apresentada enseja a delimitação do escopo da pesquisa ora realizada, assim como a demarcação de condições a serem observadas no tocante à avaliação da abordagem proposta para a renderização não fotorrealística de vídeos digitais. A classificação taxonômica que se afigurou mais apropriada foi aquela proposta por Kyprianidis et al. (2013), embora seja conveniente ressaltar que há outras classificações taxonômicas adotadas na área de NPR, e.g., Geng (2010), Borgo et al. (2011).

Na Figura 2.1 apresenta-se uma divisão taxonômica, fundamentada naquela proposta de Kyprianidis et al. (2013), na qual grupos de técnicas para renderização artística baseada em imagem são diferenciados por cores. Em tal divisão, a renderização pode fundamentar-se em: (a) traços artísticos; (b) regiões; (c) exemplos; e (d) processamento e filtragem de imagens.

Neste ponto, é pertinente ressaltar que a abordagem inicialmente proposta nesta pesquisa visava à segmentação espaçotemporal de vídeos, enquadrando-se, na classificação apresentada, no grupo das técnicas fundamentadas na abstração de regiões (utilizando a segmentação para a renderização de vídeos, apresentada na Figura 2.1B). O percurso da pesquisa desenvolvida com o uso daquela abordagem focada na segmentação está apresentado no Apêndice A.

A pesquisa retratada neste documento circunscreve-se no âmbito das abordagens que usam filtragem espacial da imagem de entrada, fundamentada no gradiente da imagem e utilizando difusão anisotrópica para abstração (destacada na linha vermelha pontilhada na Figura 2.1D). A seguir, é apresentada uma breve descrição de cada grupo da classificação apresentada.

Figura 2.1 – Taxonomia adotada nesta pesquisa para a área de NPR.



Grupo 2.1A - *Renderização fundamentada em traços artísticos*: primordialmente, os sistemas para renderização artística fundamentavam a renderização na composição sistemática e incremental de traços artísticos virtuais, na qual a cor, a orientação e o tamanho do traço podem ser definidos de forma automática ou semiautomática (HAEBERLI, 1990; LITWINOWICZ, 1997; TREAVENTT; CHEN, 1997). Tais sistemas revestem uma tela virtual de acordo com um estilo de pincelada de interesse do usuário, e.g., pincelada estilo crayon. Avanços nesta área de pesquisa levaram à simulação do meio no qual a pintura é realizada, i.e., à emulação da difusão do pigmento na tela (CURTIS et al., 1997).

Grupo 2.1A1 - *Renderização fundamentada em pinceladas artísticas*: formato prevalente na renderização fundamentada em traços artísticos, no qual as pinceladas artísticas são dispostas de forma local na tela virtual, i.e., ponderam-se os valores dos pixels adjacentes para a renderização. As técnicas que se enquadram neste grupo podem depender da interação do usuário (GOOCH; REINHARD; GOOCH, 2004; KAGAYA et al., 2010) ou do emprego de uma medida de percepção humana (SANTELLA; DECARLO, 2002). Além disto, as técnicas de renderização fundamentadas em pinceladas artísticas podem atuar de forma automática, a partir do emprego de técnicas de processamento de imagens de baixo nível destinado à renderização de imagens (HAGGERTY, 1991; HERTZMANN, 1998) ou de vídeos (LITWINOWICZ, 1997; HAYS; ESSA, 2004), assim como a partir do emprego de técnicas de visão computacional (COLLOMOSSE; HALL, 2005a). Na perspectiva global, na qual a renderização das pinceladas artísticas depende da minimização de uma função objetiva, as pinceladas podem ser guiadas semiautomaticamente pelo usuário (HERTZMANN, 2001; TRESSET; LEYMARIE, 2005) ou automaticamente (COLLOMOSSE; ROWNTREE; HALL, 2005).

Grupo 2.1A2 - *Renderização para a reapresentação tonal*: engloba técnicas de renderização que não atribuem cores aos pixels com base na variação tonal, mas decompõem as imagens em regiões (e.g., a partir do emprego de diagramas de Voronoi). Tais formas de renderização englobam o estilo pontilhista, com abordagem local e em escala única (OSTROMOUKHOV; HERSCH, 1999) ou multiescala (STREIT, 2009). O estilo pontilhista pode ser utilizado de forma global, fundamentando-se em primitivas espaciais (MOULD, 2007) ou primitivas espaciais e estruturais (MARTÍN et al., 2011; LI; MOULD, 2011) das imagens de referência. Há ainda técnicas que produzem renderizações ao estilo hachura e arte com linhas (MI; DECARLO; STONE, 2009; INGLIS; KAPLAN, 2011).

Grupo 2.1A3 - *Renderização de mosaicos e ladrilhos*: mosaicos, formados por ladrilhos ou lajotas de cores, materiais e formatos singulares são organizados visando à composição de padrões e formas artísticas. A aplicação de tais estilos específicos pode gerar diferentes formas de renderização de imagens (SCHLECHTWEG; GERMER; STROTHOTTE, 2005; XU; KAPLAN, 2008; HURTUT et al., 2009) ou de animações (KANG et al., 2011).

Grupo 2.1B - *Renderização fundamentada em regiões*: este grupo engloba técnicas que se fundamentam no uso de algoritmos de visão computacional ou de algoritmos de segmentação destinada à renderização de imagens. No tocante a renderização de vídeos, utilizam-se regiões bidimensionais em conjunção com aspectos temporais (2D+tempo) (COLLOMOSSE; ROWNTREE; HALL, 2005; BOUSSEAU et al., 2006) ou se utilizam representações tridimensionais (WANG, 2004). A estilização do movimento também está presente (LIU et al., 2005; WANG et al., 2006). Nesta categoria, há ainda a estilização de regiões de imagens, com o preenchimento de regiões (XU; KAPLAN, 2008), a renderização por modificação da forma (SONG et al., 2008), a renderização por composição (HALL et al., 2007), a renderização hierárquica (ZHAO; ZHU, 2010), a renderização ao estilo de cartuns (COLLOMOSSE; ROWNTREE; HALL, 2005; WANG et al., 2004), a composição de vitrais (SETLUR; WILKINSON, 2006) e a renderização com caneta de feltro (O'DONOVAN; MOULD, 2006).

Grupo 2.1C - *Renderização fundamentada em exemplos*: neste grupo estão presentes técnicas que usam um mapeamento entre uma imagem de referência e a representação desta imagem feita por um artista. Após o mapeamento, outras imagens são renderizadas de forma semelhante ao estilo do artista. Tais técnicas podem usar informações de cor (POULI; REINHARD, 2011) ou textura (ZHAO; ZHU, 2011) para compor a imagem a ser renderizada.

Grupo 2.1D - *Renderização fundamentada no processamento e filtragem de imagens*: os filtros empregados no processamento de imagens digitais também têm sido explorados no contexto da renderização estilizada. Entretanto, alguns deles têm sido propostos com ênfase na restauração e recuperação de imagens fotorrealísticas, contrapondo-se ao objetivo da renderização estilística, que visa à simplificação do conteúdo da imagem. As técnicas de filtragem podem ser divididas em duas categorias principais: (i) domínio espacial (prevalente na categoria de filtragem), no qual a alteração do valor do pixel ocorre em função dos valores de pixels adjacentes; e (ii) domínio do gradiente (BHAT et al., 2010), no qual a alteração do valor do pixel

ocorre em função de valores adjacentes em conjunção com a manipulação de diferenças entre pixels.

No domínio espacial, a maioria das técnicas pesquisadas atua de forma automática na produção da renderização, subdividindo-se usualmente em *estilização de linhas* e *estilização de regiões*. As técnicas para a estilização de linhas podem ser subdivididas de acordo com o tipo de detector de bordas utilizado, e.g., de *primeira derivada* (ORZAN et al., 2007) ou de *segunda derivada* (WINNEMÖLLER, 2011).

Por outro lado, as técnicas destinadas à estilização de regiões podem ser classificadas em função do emprego da difusão anisotrópica (WEICKERT, 1998; KYPRIANIDIS; DÖLLNER, 2008; KANG; LEE; CHUI, 2009; KYPRIANIDIS; KANG, 2011), da filtragem fundamentada em estatísticas locais da imagem (PAPARI; PETKOV; CAMPISI, 2007) ou da renderização de regiões a partir de filtragem morfológica (BOUSSEAU et al., 2007; CRIMINISI et al., 2010).

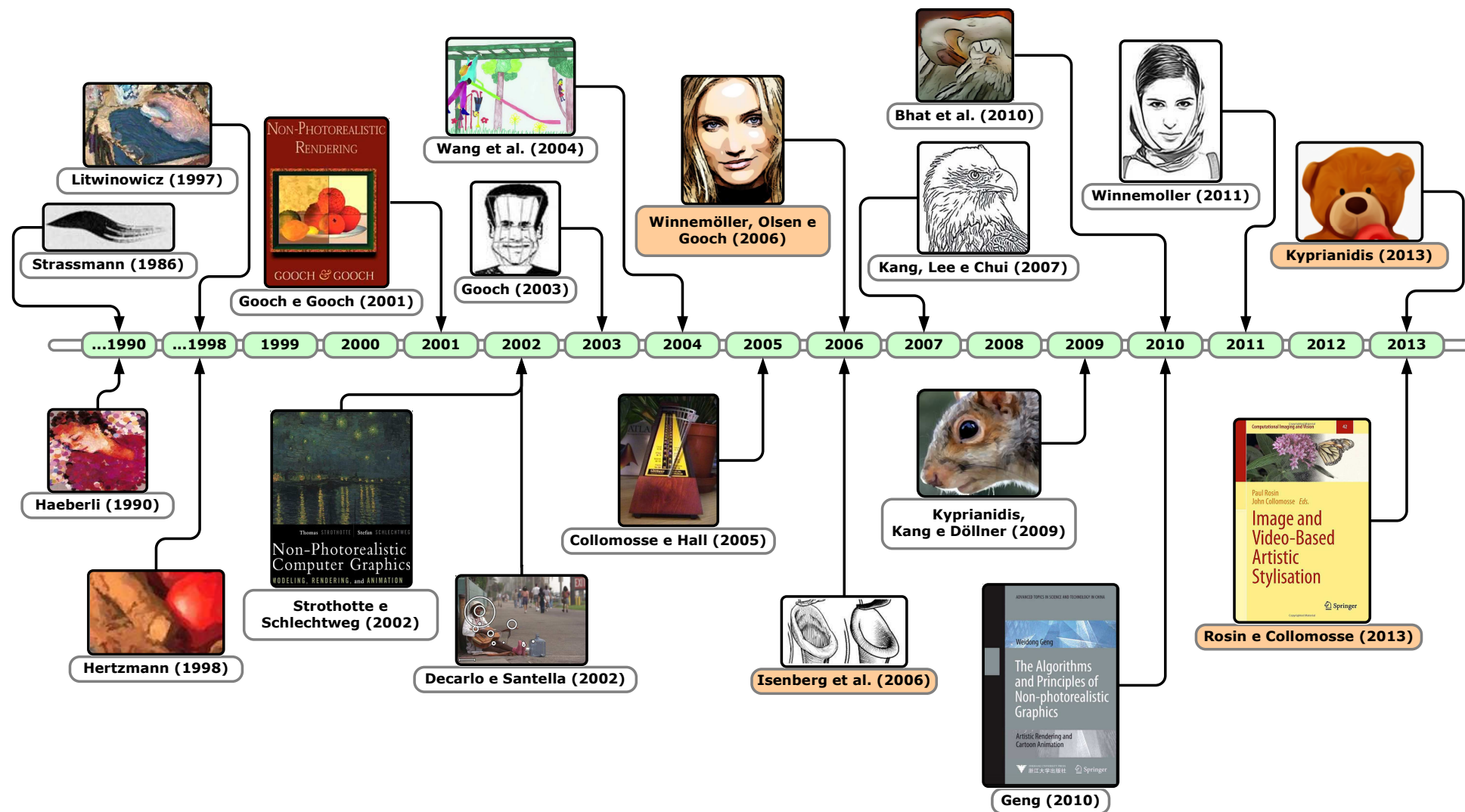
No contexto das técnicas de renderização não fotorrealística fundamentadas no processamento e filtragem de imagens, Kyprianidis et al. (2013) ressaltaram o uso de implementações do filtro bilateral (TOMASI; MANDUCHI, 1998), modificadas com fins à melhoria do desempenho computacional (e.g., uso de GPU). Adicionalmente, ressalta-se o emprego do filtro Kuwahara anisotrópico (KYPRIANIDIS; KANG, 2011) para a renderização não fotorrealística, o qual também se enquadra na categoria de difusão anisotrópica.

Conforme ressaltado anteriormente, a abordagem proposta na presente pesquisa utiliza uma variação da técnica da difusão anisotrópica, em conjunção com um detector de bordas fundamentado em redes neurais. Assim sendo, a presente pesquisa enquadra-se na seguinte categoria: *Renderização fundamentada na filtragem de imagens no domínio espacial, considerando o conteúdo das regiões e empregando a difusão anisotrópica*. A confrontação dos resultados obtidos a partir da abordagem proposta nesta pesquisa foi realizada em relação a outras abordagens pertencentes ao mesmo agrupamento taxonômico, a qual está relatada no Capítulo 5. Na próxima seção, apresenta-se uma progressão temporal da área de renderização não fotorrealística.

2.2 Breve Histórico da Renderização Não Fotorrealística

Nesta seção, é apresentado um breve histórico da área de renderização não fotorrealística. Na Figura 2.2 apresenta-se em uma visão geral da evolução da área de estilização artística, sob a forma de uma progressão temporal, adaptada da pesquisa original realizada por Kyprianidis et al. (2013).

Figura 2.2 – Progressão temporal da área de NPR, na qual se destaca a semiautomatização das abordagens propostas no início dos anos 1990, seguida da automatização a partir do emprego de técnicas de alto nível (e.g., visão computacional) e da filtragem com preservação de bordas e passando à ênfase na interação com o usuário e à avaliação da estética dos resultados de renderização produzidos.



As adaptações efetuadas em relação à pesquisa original (KYPRIANIDIS et al., 2013) foram realizadas com fins à inclusão de artigos científicos publicados entre 2010 e 2013, assim como de quatro livros que representam marcos bibliográficos para a área de NPR.

A pesquisa na área de NPR pode elucidar duas questões de cunho científico (HERTZMANN, 2011): (i) o estudo do processo de criação artística por computador; (ii) o estudo do modo como observadores reagem a resultados computacionais de NPR com carga artística elevada. As abstrações de imagens e vídeos produzidas a partir de técnicas de NPR visam à melhoria da percepção e da compreensão da cena retratada, permitindo a simplificação de objetos ou enfatizando características relevantes da cena considerada.

As primeiras iniciativas de NPR focalizaram a imitação e a transposição de estilos de pintura para imagens estáticas, concentrando-se exclusivamente na reprodução algorítmica do ato de desenhar e distribuir pinceladas artísticas na imagem resultante (HUANG, 2006). Uma das motivações para a automatização de técnicas de NPR é o provimento de meios para que usuários comuns produzam imagens estáticas ou vídeos que se assemelham a artefatos artísticos, uma vez que as ferramentas que provêem tais meios possibilitam a geração de imagens ou de vídeos abstraídos com pouca intervenção do usuário (GOMES et al., 2009).

Haggerty (1991) foi o precursor da reprodução de estilos artísticos de forma algorítmica, tendo proposto um procedimento para a renderização automática de imagens que distribuía pinceladas de forma pseudorrandômica, considerando as cores da imagem de entrada.

Em seguida, Litwinowicz (1997) propôs um sistema que distribuía pincelada de forma dependente do conteúdo da imagem, tendo sido o primeiro a propor uma técnica completamente automática de renderização NPR. O autor utilizou um algoritmo para a computação do fluxo óptico, visando a estimar o vetor de movimento e transladar as pinceladas pintadas no quadro inicial na mesma direção do movimento, nos quadros posteriores. Tal estratégia foi adotada, por exemplo, no filme *Waking Life*¹³.

Por sua vez, Hertzmann e Perlin (2000) desenvolveram uma abordagem baseada em exemplos, segundo a qual o estilo é aprendido a partir de um exemplo de

¹³ Disponível em www.imdb.com/title/tt0243017.

entrada, sendo reproduzido na nova imagem gerada.

O desenvolvimento acelerado das técnicas de NPR possibilitou a extensão do foco das pesquisas do domínio das imagens estáticas para o domínio das animações e vídeos. De uma forma geral, a renderização não fotorrealística em animações e vídeos pode ser realizada em: (i) animações sintéticas, produzidas a partir de primitivas 3D ou (ii) conjuntos de quadros de vídeo temporalmente adjacentes. De posse de uma animação sintética, a renderização não fotorrealística é facilitada, devido à presença das primitivas dos objetos da cena e das informações precisas dos movimentos nela presentes.

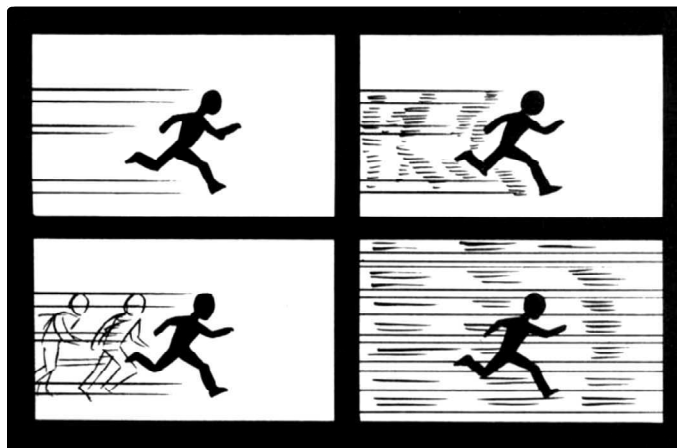
Quando o vídeo a ser renderizado não possui informações relativas às primitivas sintéticas presentes no espaço 3D, mas é constituído por uma sequência de imagens/quadros 2D, as abordagens para NPR baseiam-se em técnicas empregadas em imagens estáticas, nas quais as pinceladas são distribuídas quadro-a-quadro e alguma técnica para tentar evitar artefatos intraquadros é posteriormente utilizada.

No contexto da renderização não fotorrealística de vídeos, um método simplificado de translação de pinceladas artísticas foi proposto por Hertzmann e Perlin (2000), com o propósito de diferenciar quadros consecutivos de um vídeo, a partir do qual são repintadas apenas as áreas que sofreram alterações significativas, com relação a um limiar global definido pelo usuário.

Lake et al. (2000) apresentaram técnicas para enfatizar o movimento de objetos em vídeo, adicionando primitivas geométricas que são desenhadas, levando-se em consideração o movimento dos objetos em um vídeo sintético 3D. A adição de tais primitivas aos objetos do vídeo possui limitações, tais como a simplicidade do resultado e a não contextualização do movimento retratado na cena. Algumas técnicas destinadas à ênfase do movimento, enquanto efeito visual de complexidade mais elevada, necessitam de uma análise de alto nível do vídeo, o que não é feito, por exemplo, por ferramentas comerciais de edição de vídeo.

No contexto da ênfase no movimento, a pesquisa de Lake et al. (2000) encontra respaldo em McCloud (1993), o qual afirma que as linhas de movimento são utilizadas com maestria pela indústria de histórias em quadrinhos, em virtude de sua habilidade de retratar drama, a partir dos diferentes estilos em que aparecem. Efeitos para a ênfase do movimento considerados de baixa complexidade são o *slow motion* e o *motion blur*. Alguns exemplos de linhas utilizadas para dar ênfase ao movimento são ilustrados na Figura 2.3.

Figura 2.3 – Ênfase no movimento.



Fonte: McCloud (1993)

DeCarlo e Santella (2002) realizaram a estilização de fotografias a partir de um mecanismo de rastreamento do movimento dos olhos do observador. Deste modo, os autores obtiveram informações necessárias à geração de um mapa de atenção visual para a imagem considerada, a partir do qual foram inferidas informações acerca das regiões de grande impacto na imagem, posteriormente empregadas para criar a renderização estilizada.

As pesquisas realizadas por DeCarlo e Santella (2002), nas quais a atenção visual humana guiou a ênfase de pontos diferentes da imagem a ser renderizada, marcou uma divisão nas pesquisas na área de NPR (KYPRIANIDIS, 2013). Tal divisão ocorreu porque as técnicas anteriores àquela proposta por DeCarlo e Santella (2002) produziam pinceladas artísticas de acordo com o gradiente da imagem, conservando detalhes de alta frequência (bordas) na imagem renderizada. Técnicas posteriores consideraram o fluxo usado por artistas ao desenhar, no qual os detalhes mais salientes da imagem são mais bem retratados, enquanto detalhes alheios à cena são submetidos a um grau maior de abstração.

Agarwala (2002) propôs um sistema interativo que facilitava a criação de animações clássicas por parte de não profissionais, a partir de imagens e vídeos. O sistema funcionava a partir da segmentação manual da primeira imagem e, com base em contornos ativos (*snakes*), as bordas dos segmentos eram rastreadas ao longo do vídeo, quadro-a-quadro. Limitações desta abordagem incluem o grande esforço associado à segmentação e a limitação da natureza do vídeo de entrada, o qual devia incluir objetos bem definidos, de baixa complexidade e com fronteiras bem definidas. Adicionalmente, o autor propôs o emprego de outra abordagem, denominada *rotoscopia avançada*, segundo a qual, após o desenho de quadros-chave pelo artista, o sistema interpola os quadros desenhados, de acordo com um intervalo definido pelo

artista/usuário.

Hays e Essa (2004) propuseram uma modificação na abordagem descrita por Hertzmann e Perlin (2000), segundo a qual o uso das bordas guia o refinamento da pintura das pinceladas. Os autores objetivavam a concepção de uma estratégia que estendesse a renderização baseada em pixels para a renderização baseada em células da imagem, cuja determinação implicava uma análise de alto nível da cena, em contraste com as análises de baixo nível, realizadas nas técnicas precursoras. No tocante a análise de alto nível da cena, dois aspectos-chave devem ser considerados: (i) a segmentação; e (ii) o rastreamento (HUANG, 2006).

Conforme verificado na literatura da área, processos de segmentação de imagens e de vídeos foram incorporados ao arcabouço de técnicas destinadas à estilização de imagens e vídeos (WANG et al., 2004). Imagens estilizadas/abstraídas são, tipicamente, compostas por vastas regiões semanticamente significativas, delimitadas por bordas bem delineadas.

Visando a renderização de vídeos reais em vídeos renderizados não fotorrealisticamente, a identificação de tais regiões significativas torna-se um aspecto relevante, pois algoritmos de segmentação podem atuar de forma a dar suporte em tarefas realizadas pelo usuário ou, ainda, a realizar a tarefa de renderização de forma completamente autônoma. Por sua vez, o rastreamento pode ser feito, por exemplo, a partir de um algoritmo que se fundamente no filtro de Kalman (1960).

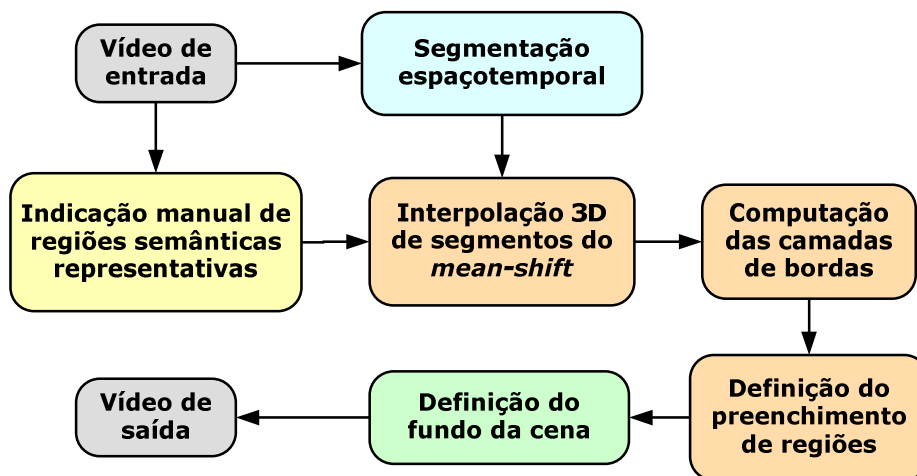
Inspirados nas pesquisas de DeCarlo e Santella (2002), Wang et al. (2004) propuseram uma abordagem destinada à transformação de um vídeo em cartum, ressaltando que o resultado abstraído deve possuir as seguintes características: (i) o resultado deve manter uma coerência espaço-temporal, evitando transições abruptas entre quadros; (ii) a abstração do conteúdo deve manter a representação semântica; e (iii) o artista deve ter controle sobre o estilo do resultado.

Wang et al. (2004) introduziram uma técnica semiautomática para a renderização de vídeos temporalmente coerentes ao estilo de *cartuns*, a qual é, sem dúvida, uma referência para a pesquisa da área. Os autores desenvolveram uma abordagem para a segmentação espaçotemporal baseada na técnica *mean-shift* (COMANICIU; MEER, 2002), tendo utilizado adicionalmente uma máscara (*kernel*) anisotrópica. Tal estratégia lhes possibilitou a obtenção de melhores resultados para segmentação de vídeo do que utilizando exclusivamente a técnica *mean-shift*. Os autores obtiveram subvolumes de atributos de baixo nível (*voxels* contíguos de cor similar), os quais foram utilizados como informação para criar o vídeo abstraído.

A abordagem proposta por Wang et al. (2004) estende a segmentação do domínio bidimensional para uma segmentação que provê, diretamente, primitivas tridimensionais para a manipulação, por parte do usuário, de regiões semânticas, a partir de uma interface gráfica. É disponibilizado um meio para o destaque das regiões a serem segmentadas, assim como para o ajuste da supersegmentação advinda do processo de uso do algoritmo *mean-shift*, minimizando a tarefa tediosa de segmentação manual quadro-a-quadro.

Adicionalmente, na técnica proposta por Wang et al. (2004), o usuário devia interagir com o sistema, a fim de indicar as regiões salientes quadro-a-quadro evitando, assim, a implementação de um algoritmo de rastreamento de regiões não rígidas (*non-rigid region tracking*). A interpolação entre quadros era baseada na interpolação guiada pela técnica *mean-shift* e os estilos de renderização do vídeo final resumiam-se à modificação da cor de um segmento e à mudança do fundo (*background*) da animação. A arquitetura do sistema proposto por Wang et al. (2004) pode ser visualizada na Figura 2.6

Figura 2.4 – Arquitetura da abordagem proposta por Wang et al. (2004).



No entanto, a utilização da abordagem proposta por Wang et al. (2004) demandava que o usuário dispusesse de várias horas para a conclusão da renderização, tal qual a renderização de um filme de 10 segundos (300 quadros), pois o *mean-shift* devia ser aplicado a todos os quadros do vídeo. Assim sendo, ressalta-se como limitação do sistema proposto por Wang et al. (2004) o tempo elevado para a produção do vídeo de saída, assim como o excesso de interatividade com o usuário, o qual necessitava, a cada 15 quadros do vídeo original, corrigir os segmentos detectados pela técnica dos autores.

Collomosse, Rowntree e Hall (2005) afirmaram que, no contexto de NPR, a transformação de um vídeo em *cartum* pelo uso de algoritmos de PDI é susceptível ao

fenômeno da incoerência temporal, se os quadros forem renderizados um a um. Os autores ressaltaram ainda que as abordagens adotadas para transformar vídeos em *cartuns* empregam usualmente algoritmos de visão computacional de propósito geral, tais como algoritmos para a detecção de bordas ou para a computação do fluxo óptico.

O sistema proposto por Collomosse, Rowntree e Hall (2005) teve como propósito a redução do esforço do usuário em processos de geração de vídeos ao estilo de *cartuns*, buscando minimizar diversos efeitos indesejáveis, tais como ruído, cintilação (*flickering*) e incoerência temporal entre quadros do vídeo. Todavia, o referido sistema ainda necessitava a seleção manual do objeto de interesse a ser segmentado e renderizado. Embora tenham adotado a segmentação semi-automática baseada na coerência cromática, adicionada ao processamento com o algoritmo *mean-shift*, os autores sugeriram que o processo de segmentação poderia ser melhorado a partir do emprego de informações de textura ou movimento.

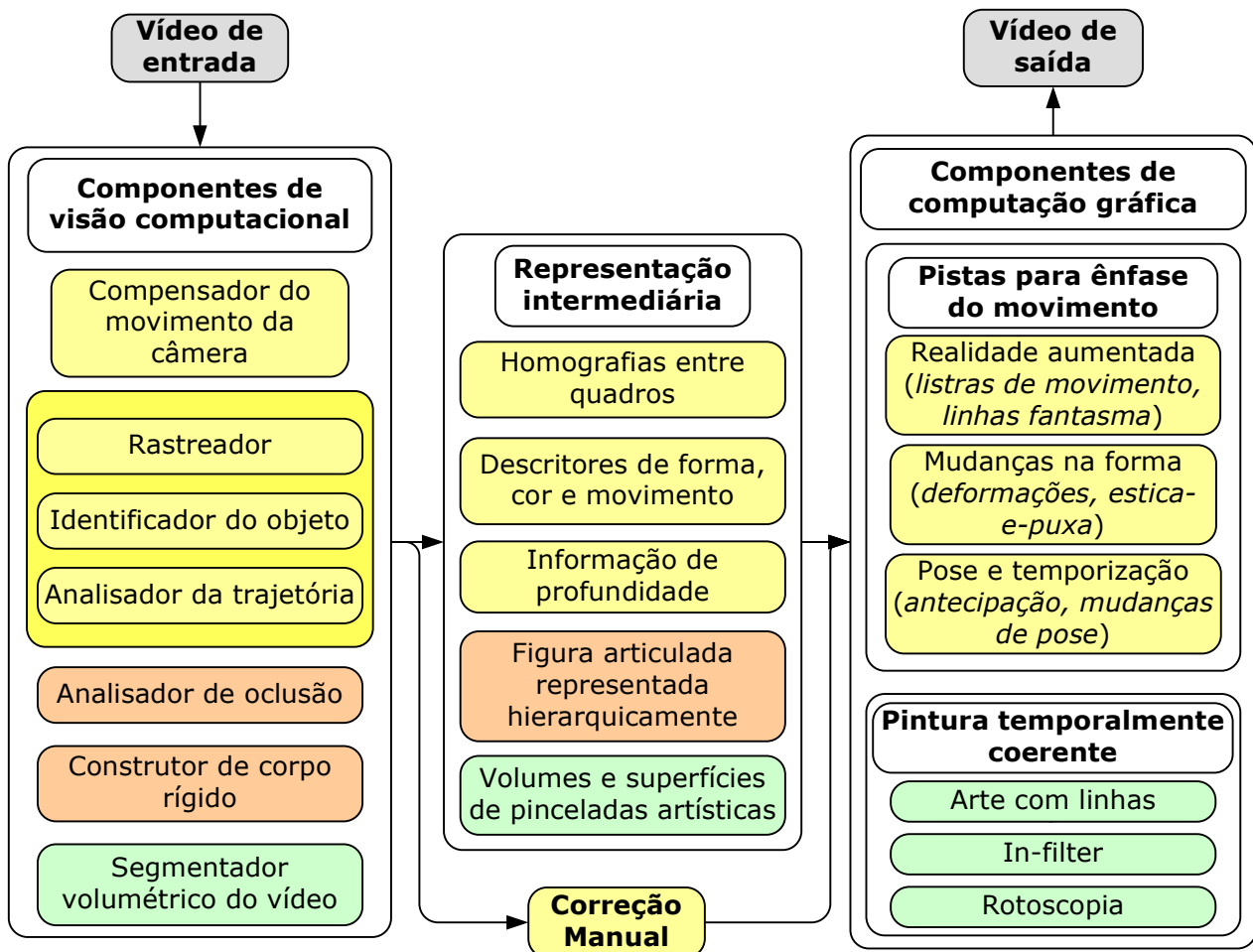
É conveniente observar que os autores incorporaram técnicas para a análise de vídeo, assim como técnicas de visão computacional, ao problema da renderização artística de vídeo, tendo argumentado que a análise do vídeo deveria ser o primeiro passo para a renderização. A informação saliente, assim como os limites dos objetos em cena e suas trajetórias, deveriam ser informações extraídas antes de qualquer tarefa de renderização. A partir de técnicas de visão computacional, Collomosse, Rowntree e Hall (2005) conseguiram enfatizar o movimento a partir de técnicas consagradas na produção de quadrinhos, tais como linhas de movimento (*streak-lines*), antecipação do movimento (*anticipation*) e deformação (*deformation*).

Tal contribuição possibilitou a análise das trajetórias dos objetos, assim como um tratamento mais refinado das oclusões e colisões entre objetos de interesse do vídeo de entrada. No entanto, o componente de visão computacional é penalizado por outra etapa no processamento: a intervenção manual do usuário para identificar e segmentar corretamente o objeto-alvo da renderização do vídeo. De acordo com os autores, a solução para tal problema residia no emprego de uma técnica mais automatizada de segmentação, em conjunção com um algoritmo de rastreamento de regiões não rígidas, a fim de melhorar a capacidade de análise do vídeo.

A arquitetura da abordagem proposta por Collomosse, Rowntree e Hall (2005) é ilustrada no diagrama em blocos da Figura 2.5, no qual a cor amarela identifica os componentes de visão computacional que são responsáveis pela construção dos elementos da representação intermediária, a cor laranja identifica os componentes de

computação gráfica que utilizam os elementos da representação intermediária, enquanto a cor verde representa os componentes de processamento de baixo nível empregados para dar suporte aos módulos de visão computacional e de computação gráfica.

Figura 2.5 – Arquitetura da abordagem proposta por Collomosse, Rowntree e Hall (2005).



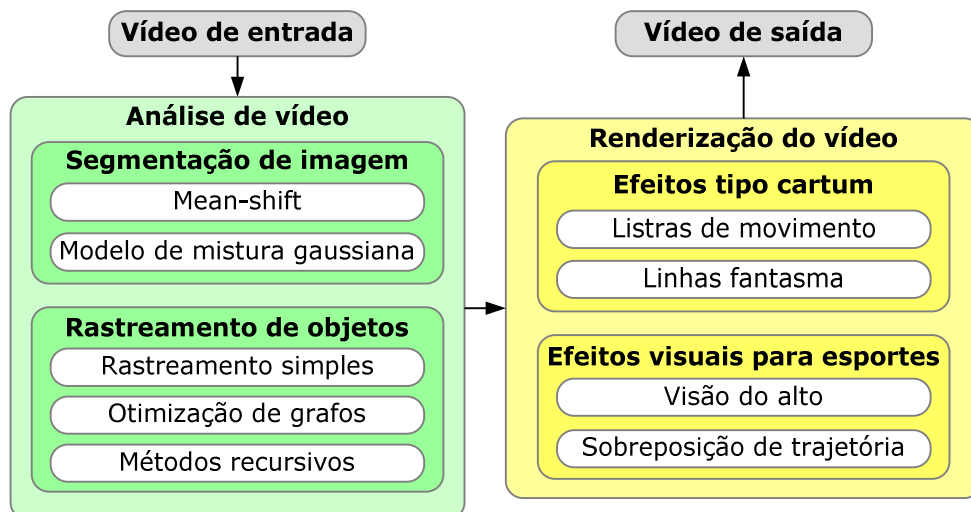
Na abordagem proposta por Collomosse, Rowntree e Hall (2005), as limitações residem nas restrições impostas ao tipo de movimento a ser tratado, assim como ao tipo de objeto a ser abstraído. O tipo de movimento deve ser planar, i.e., os objetos devem mover-se em apenas um plano e, adicionalmente, o objeto a ser renderizado não pode conter partes muito curvas ou cujo movimento, no vídeo original, curve o objeto original. A não observância de tais restrições pode implicar erros ou incoerências intraobjeto no vídeo resultante, conforme advertiram os autores.

Na abordagem proposta por Huang (2006), testes foram realizados em vídeos de interiores (*indoor videos*) e vídeos de exteriores (*outdoor videos*). Para os vídeos de interiores, duas subclasses foram consideradas: vídeos contendo objetos simples (e.g., uma pequena bola) e vídeos contendo objetos complexos (e.g., seres humanos). A técnica de limiarização por similaridade de cor foi empregada para a

extração de objetos simples, enquanto a técnica de segmentação *mean-shift* foi para a segmentação de objetos mais complexos, calibrada para os objetivos do autor. Outro aspecto importante do sistema proposto por Huang (2006) são as linhas de ênfase de movimento, as quais foram renderizadas a partir de *splines* cúbicas.

Para a categoria de vídeos de exteriores, a segmentação a partir de Modelos de Mistura Gaussiana (GMM – *Gaussian Mixture Models*) (REYNOLDS, 1995) foi empregada em vídeos de partidas de futebol, com o propósito de separar os jogadores e o campo. Após testes comparativos entre a segmentação dos vídeos de exteriores, a partir do *mean-shift* e GMM, Huang (2006) relatou que o GMM mostrou-se mais adequado aos objetivos do sistema, pois a técnica *mean-shift* é mais lenta do que a técnica GMM. O autor acrescentou que os parâmetros do *mean-shift* podiam diferir para cada conjunto de quadros, enquanto o modelo GMM podia ser calibrado apropriadamente para levar em consideração a categoria de vídeos a serem testados. Vale ressaltar que as câmeras utilizadas na captura dos vídeos usados nos testes foram dispostas estaticamente e o GMM foi adaptado para levar em consideração a cor da grama, comum em vídeos esportivos. A arquitetura proposta pelos autores é ilustrada na Figura 2.6.

Figura 2.6 – Arquitetura a abordagem proposta por Huang (2006).



Pesquisas futuras apontadas por Huang (2006) incluíam: (i) testes com outras categorias de vídeos externos; (ii) a adição de algoritmos para a compensação do movimento da câmera (*camera compensation*); (iii) a incorporação de técnicas destinadas à ênfase do movimento, tais como estica-e-puxa (*squash-and-stretch*), antecipação do movimento, *slow-in* e *slow-out*. Efeitos indesejáveis podem ser observados nos resultados apresentados por Huang (2006), a saber: (i) a necessidade da interação com o usuário na marcação dos pontos de interesse do objeto

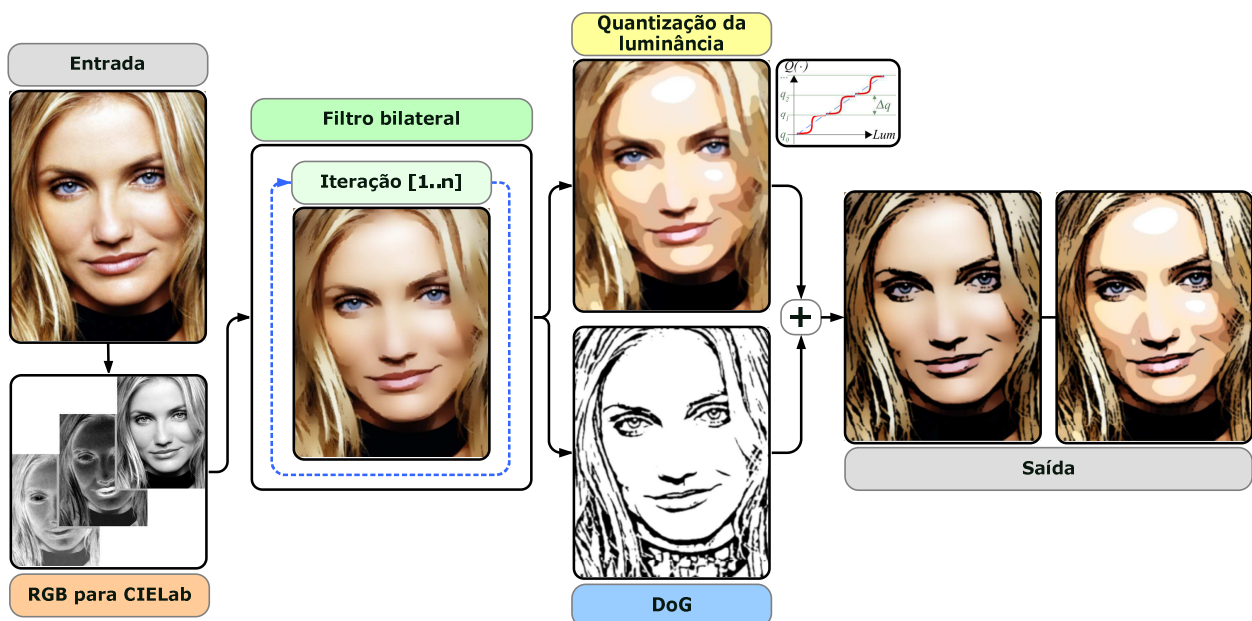
segmentado para a sobreposição das linhas de ênfase de movimento; (ii) a não naturalidade do efeito produzido pela sobreposição das linhas de ênfase de movimento; e (ii) a baixa qualidade da segmentação obtida com o uso do algoritmo *mean-shift* para a renderização do vídeo resultante.

Kim e Essa (2005) apresentaram um sistema semiautomático para adicionar efeitos que enfatizavam movimentos em vídeos, fazendo uso de técnicas de análise de movimento. Um passo importante da técnica proposta era a etapa de segmentação de uma região que fizesse parte do assunto ou tema da cena de vídeo. Após a segmentação, o usuário devia agrupar regiões semanticamente similares, a fim de que o efeito de ênfase do movimento fosse percebido.

O processo de segmentação utilizado pelos autores dividia-se em três etapas, a saber: (i) a filtragem espacial para a suavização com preservação das bordas (uso do filtro bilateral); (ii) o agrupamento de pixels com cor similar, resultando em supersegmentação (após a aplicação do *mean-shift*); e (iii) a união de segmentos adjacentes, de acordo com um limiar obtido experimentalmente. O espaço de cor adotado foi o *CIE Luv*, o qual, segundo os autores, afigurava-se perceptualmente mais apropriado para a tarefa.

Winnemöller, Olsen e Gooch (2006) propuseram uma abordagem destinada à abstração de vídeo em tempo real, objetivando a geração de um vídeo ao estilo de cartum, por meio de um processo de pseudoquantização de cores e superposição de bordas às cores abstraídas. Tal abordagem fundamentava-se na filtragem espacial dos quadros do vídeo, a partir do uso do filtro bilateral. A arquitetura da abordagem proposta por Winnemöller, Olsen e Gooch (2006) é ilustrada na Figura 2.7.

Figura 2.7 – Arquitetura da abordagem proposta por Winnemöller, Olsen e Gooch (2006).



Limitações da abordagem proposta por Winnemöller, Olsen e Gooch (2006) incluem dificuldades na abstração de imagens cujo contraste é baixo, assim como no controle do grau de abstração obtido, devido ao uso de uma função fixa para a realização da pseudoquantização das cores do quadro de entrada. A abordagem proposta por Winnemöller, Olsen e Gooch (2006) para a abstração de vídeos também foi adotada previamente por Hays e Essa (2004), os quais descreveram uma abordagem para a renderização de imagens e vídeos a partir da utilização do fluxo óptico para o desenho das pinceladas artísticas e o provimento de coerência temporal às pinceladas desenhadas. Os resultados obtidos apresentam alto grau de abstração, sendo o processo de refinamento das pinceladas desenhadas guiado por um detector de bordas.

Hwang et al. (2006) ressaltaram algumas das dificuldades encontradas na geração de histórias em quadrinhos: (i) a natureza essencialmente manual do trabalho de geração de imagens; (ii) a extensão do tempo necessário à finalização do produto; e (iii) a dificuldade de automatização do processo. Adicionalmente, os autores ressaltaram o fato de não existirem regras gerais a respeito da avaliação da qualidade ou os efeitos cognitivos dos resultados finais de tal processo. Assim sendo, os autores propuseram um sistema destinado a auxiliar o usuário a transformar vídeos em histórias em quadrinhos.

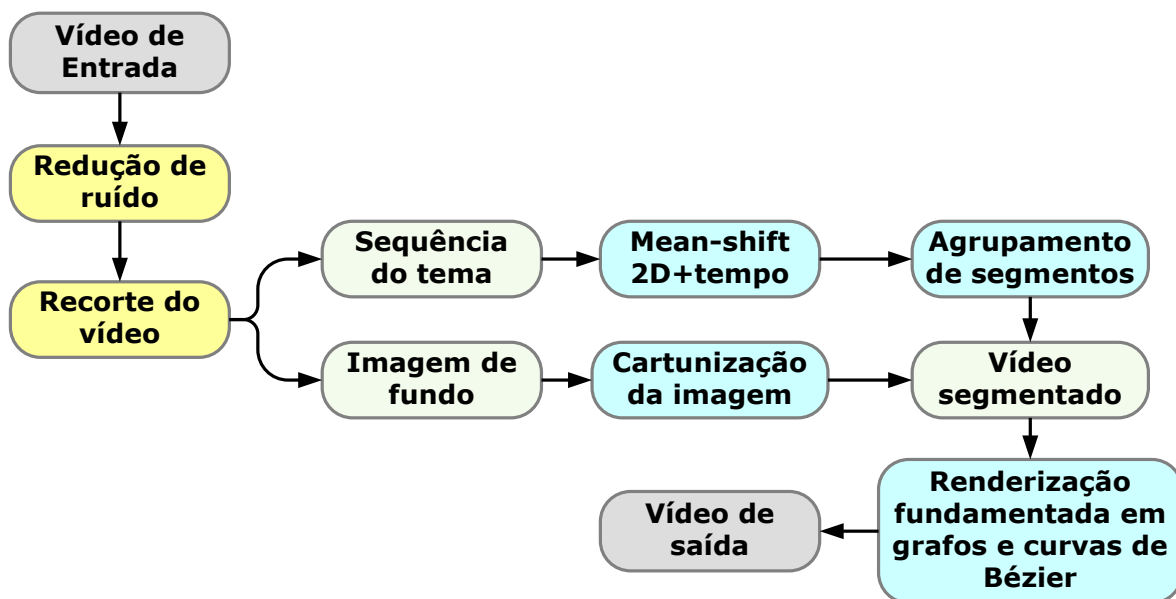
O sistema proposto por Hwang et al. (2006) incluiu o suporte ao posicionamento de balões de texto, efeitos estilizados que ressaltam o movimento linear ou rotacional de objetos, assim como texturas para o fundo do quadrinho. No entanto, o sistema proposto não realiza boa transformação dos quadros-chave em *cartuns*, além do que a segmentação dos quadros a serem transformados é feita de forma completamente manual. A técnica de segmentação usada pelos autores para realizar a abstração foi a técnica *mean-shift*, considerada pelos mesmos como sendo a técnica predominante para a segmentação de vídeos.

Hong et al. (2007) propuseram uma técnica fundamentada na redução de ruídos e também introduziram melhorias no algoritmo *mean-shift* para a segmentação de imagens, tendo apresentado um método para a conversão de uma sequência de vídeo em um vídeo ao estilo de *cartuns* e também se inspirado na pesquisa de DeCarlo e Santella (2002). Os passos do processo de conversão incluíam o pré-processamento e a divisão de cada quadro em fundo (*background*) e tema (*foreground*), a partir da aplicação do filtro *mean-shift* apenas no tema, sucedido da eliminação semiautomática ou automática de artefatos espúrios resultantes do

processo de filtragem. O vídeo final era, então, sintetizado a partir da fusão quadro-a-quadro do fundo com o tema previamente processado. Os autores aumentaram o desempenho computacional do algoritmo aplicando separadamente o algoritmo de segmentação no fundo e no tema do vídeo.

Mesmo com tais simplificações, os autores ressaltaram que o algoritmo *mean-shift* não se afigurava apropriado para a interação usuário-computador, pois o processamento não ocorria em tempo real. Em tal abordagem, todos os *pixels* de todos os quadros do vídeo deviam ser processados, uma tarefa que demandava tempo de processamento elevado, cerca de 4 minutos por quadro do vídeo. Outro fator limitante residia no fato do algoritmo proposto funcionar apenas para tomadas com ângulos fixos, o que podia ser melhorado com a utilização de um vetor de características do tema para o alinhamento temporal entre quadros. A arquitetura da abordagem proposta por Hong et al. (2007) é ilustrada na Figura 2.8.

Figura 2.8 – Arquitetura da abordagem proposta por Hong et al. (2007).



Wang (2007), após extensões da técnica proposta por Wang et al. (2004), corroborou com Hong et al. (2007), ao afirmar que a separação entre fundo e tema é uma etapa fundamental na edição de vídeos e que a segmentação espaçotemporal de uma sequência de quadros em um vídeo é uma tarefa difícil, pois o tema extraído em quadros consecutivos deve ser espacial e temporalmente coerente. As abordagens existentes recorrem ao usuário por serem penalizados pela falta de acurácia do resultado final.

De posse de uma ferramenta destinada à segmentação fundo/tema, de forma eficiente, Wang (2007) demonstrou uma variedade de aplicações para a técnica,

tendo proposto como aplicação para vídeos um processo de extração temporalmente coerente do tema, fazendo uso de uma nova abordagem fundamentada no algoritmo *mean-shift*, com o propósito de gerar supersegmentação, a qual era melhorada com o uso do algoritmo *graph-cut*, empregado com base em informações de cor.

Outra aplicação proposta pelo autor foi um sistema para a cartunização de vídeos, o qual utilizava os temas extraídos do vídeo como primitivas de processamento para gerar representações estilísticas com uma variedade de estilos, assim como exageros na forma e movimento dos objetos extraídos.

No entanto, Wang (2007) relatou a necessidade de intervenção do usuário em muitas das etapas associadas à obtenção do vídeo cartunizado, assim como ressaltou que a definição das formas e dos objetos a serem exagerados/deformados ainda se afigura um campo de pesquisa promissor, pois o reconhecimento de objetos baseado em aprendizagem é um campo de pesquisa ativo na área de visão computacional.

Wang (2007) acrescentou que o algoritmo poderia se beneficiar do conhecimento prévio de objetos para executar as deformações e que a segmentação por *mean-shift* e auxiliada por *graph-cut* poderia ser significativamente melhorada com a adição de informações de textura da imagem. Outra limitação da abordagem é o tempo elevado de processamento, relatado como 30 minutos de processamento computacional para o pré-processamento de 100 quadros.

Para tratar os problemas citados, Gomes et al. (2009) propuseram um método para melhorar a coerência temporal intraobjetos ao renderizar vídeos de forma não fotorrealística. A técnica proposta fundamenta-se na combinação de um método de segmentação e um método para a computação do fluxo óptico restrito aos segmentos detectados. O algoritmo de segmentação empregava o crescimento de regiões, permitindo que o usuário selecionasse objetos de interesse que formariam as sementes para o algoritmo de segmentação.

Tal interação facilitava as etapas subsequentes, pois o usuário podia selecionar as sementes em quadros diferentes, melhorando a separação dos segmentos e sua associação em objetos que possuíssem relação semântica. O algoritmo destinado à computação do fluxo óptico era, então, restringido às áreas dos objetos segmentados, melhorando a informação dos vetores de fluxo obtidos e melhorando a coerência temporal da estilização do objeto segmentado, ao longo dos quadros do vídeo de entrada. Os autores ressaltaram que a intervenção do usuário no processo de

segmentação atenuava problemas citados por Collomosse, Rowntree e Hall (2005), e.g., a segmentação de objetos pequenos ou que mudassem gradualmente de forma.

Bousseau et al. (2007) afirmaram que pinturas ao estilo de aquarela possuem características fundamentais que as tornam especialmente bonitas e evocativas, tais como as texturas diferenciadas advindas tanto da textura do papel quanto da granularidade dos pigmentos da pintura, além da insinuação de detalhes, mesmo com a abstração de cores e formas. Outros autores exploraram a renderização em aquarela (CURTIS et al., 1997; LUM; MA, 2001; VAN LAERHOVEN et al. 2004; BOUSSEAU et al. 2006).

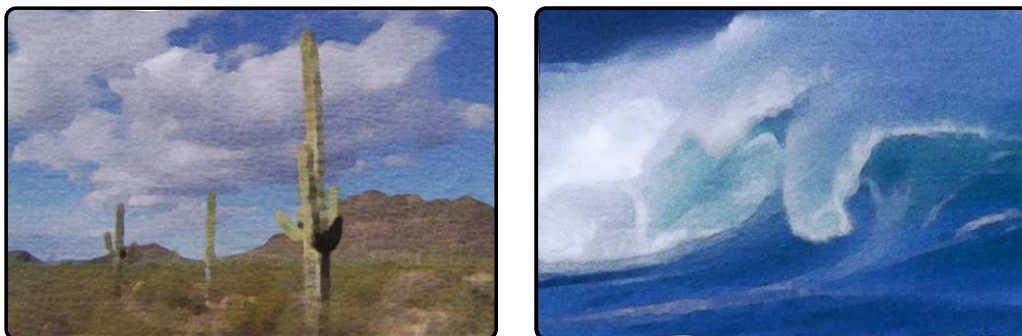
Bousseau et al. (2007) acrescentaram que as qualidades das pinturas em aquarela as tornam difíceis de reproduzir de uma forma temporalmente coerente, quando aplicada à renderização de um vídeo. Dois tipos de problemas quando da renderização de vídeos foram ressaltados pelos autores: (i) não há coerência temporal e as cores são modificadas randomicamente ao longo do tempo; ou (ii) o efeito de *shower-door* é exibido. Os autores ressaltaram ainda a dificuldade na adaptação de técnicas de NPR adotadas em imagens estáticas para vídeos, uma vez que não levam em consideração aspectos temporais para a renderização.

Bousseau et al. (2007) contribuíram para o melhoramento da técnica de renderização de aquarelas por meio da combinação de técnicas de medição do fluxo da animação combinada à computação do fluxo óptico, assim como a melhoria do cálculo do movimento das texturas retratadas no vídeo.

O cálculo do movimento das texturas foi realizado nas duas direções temporais do vídeo (avanço e retrocesso de cenas). Houve ainda a extensão de técnicas de morfologia matemática para filtragem temporal. Um exemplo do resultado obtido por Bousseau et al. (2007) pode ser visto na Figura 2.9, na qual se pode observar a ausência de distorção das cores, mesmo em imagens com movimentos complexos.

Apesar de satisfatórios, os resultados obtidos pelos autores apresentavam uma qualidade dependente da qualidade dos vetores obtidos após a computação do fluxo óptico. Perspectivas para a melhoria da referida abordagem incluíam: (i) a investigação de técnicas para a estimativa de movimento a partir da pós-otimização do fluxo óptico, e.g., conforme descrito por Sand e Teller (2006), na abordagem denominada *Particle Video*; e (ii) a tentativa de implementação da abordagem em GPU (*Graphics Processing Unit*).

Figura 2.9 – Resultados obtidos por Bousseau et al. (2007).



Posteriormente, Ryu et al. (2008) apresentaram um sistema destinado à conversão de sequências ou tomadas de filmes em tirinhas (*comic strips*) em preto-e-branco. Tal conversão utilizava a filtragem do vídeo a partir do filtro bilateral, responsável pela suavização da diferença das cores entre as áreas homogêneas, sendo preservadas as bordas entre as regiões. Após o processo de filtragem, os quadros eram segmentados via algoritmo *mean-shift*, dentro do espaço de cores *HSV*. Ressalta-se que o processo de extração de cenas, seguido da seleção dos quadros-chaves e da separação entre o fundo (*background*) e o tema (*foreground*), era um processo completamente *manual*. Mesmo assim, podiam ser empregadas técnicas para auxiliar o usuário na tarefa de segmentação, e.g., *Soft Scissors* (WANG; AGRAWALA; COHEN, 2007).

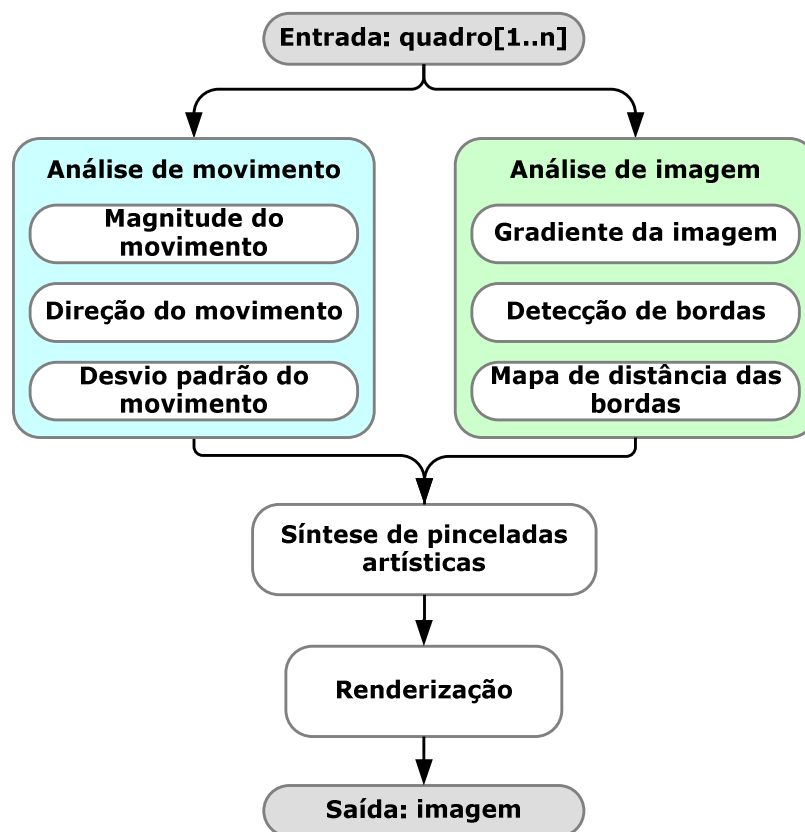
A maior limitação do sistema proposto por Ryu et al. (2008) reside na identificação de quais as cenas que deveriam compor os quadros-chaves que formavam uma única história em quadrinhos, processo que carecia de automatização no sistema proposto pelos autores. Após a segmentação, ocorria a renderização por estilização do fundo (*background*) e a inserção de balões de texto. O sistema proposto seria melhorado se extraísse quadros-chaves de forma automática, o que facilitaria a tarefa de renderização.

Sauvaget e Boyer (2009) afirmaram que algoritmos para a abstração de imagens obtêm como resultado imagens com pouca profundidade, dificultando a discriminação precisa do fundo e do tema, um problema recorrente em várias técnicas, tais como aquelas propostas por Wang et al. (2004) e por Winnemöller, Olsen e Gooch (2006). A fim de criar a ilusão de profundidade, comumente presente nas cenas de quadrinhos, os autores propuseram um arcabouço arquitetural que modificava cores da cena renderizada. Para tanto, várias etapas foram propostas neste arcabouço, dentre as quais se destaca o uso de uma técnica empregada na extração da estrutura da imagem e na análise de suas partes constituintes - a detecção de bordas, com posterior segmentação via algoritmo *mean-shift*. Outras

etapas envolviam a geração de um mapa de profundidade das regiões segmentadas, a estilização e a conversão de cores para enfatizar o tema da cena renderizada.

Lee, Lee e Yoon (2009) propuseram uma abordagem semiautomática para a renderização não fotorrealística de vídeos, cuja premissa era o emprego de informações de movimento do vídeo, visando à geração de uma única imagem abstraída. Mais precisamente, quadros de um vídeo eram empregados para gerar uma imagem abstraída, a partir de duas estratégias distintas, a saber: (i) análise de movimento; e (ii) análise de baixo nível. A análise de movimento levava em conta a magnitude, a direção e o desvio padrão do movimento. Por sua vez, na análise de imagens, considerava-se o gradiente da imagem, a detecção de bordas e mapa de distância das bordas. A contribuição principal dos autores é a combinação da análise de baixo nível com a análise do movimento em cena, visando à síntese das pinceladas artísticas na imagem resultante. Uma síntese gráfica da abordagem proposta por Lee, Lee e Yoon (2009) é ilustrada na Figura 2.10.

Figura 2.10 – Arquitetura da abordagem proposta por Lee, Lee e Yoon (2009).



Limitações da técnica proposta por Lee, Lee e Yoon (2009) incluíam: (i) a realização de testes da abordagem exclusivamente em vídeos de paisagens; (ii) a carência de expressividade das pinceladas artísticas utilizadas, no tocante à textura e formato utilizados; e a (iii) a necessidade de intervenção do usuário para fornecer um limiar de separação entre objetos dinâmicos e estáticos.

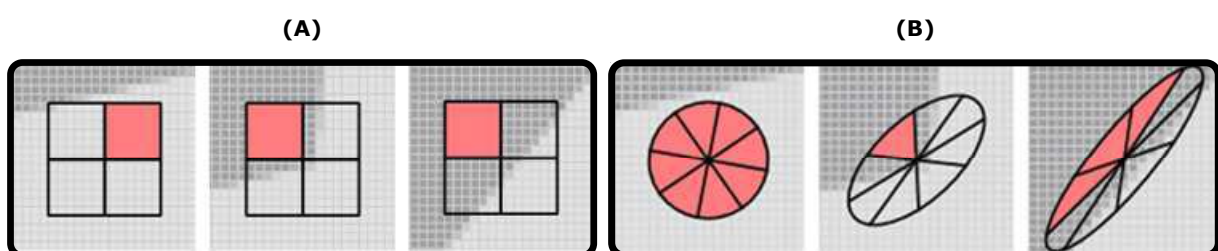
As técnicas descritas anteriormente para a redução da informação contida nas cenas retratadas fundamentaram-se em algoritmos de segmentação de imagens. Por outro lado, uma variedade de abordagens destinadas à abstração de vídeos foram propostas, com base em técnicas de suavização, em especial a suavização com preservação de bordas (*edge-preserving smoothing*).

Kyprianidis e Döllner (2008) e Kang, Lee e Chui (2009) propuseram melhorias e extensões ao arcabouço de abstração proposto por Winnemöller, Olsen e Gooch (2006). A técnica para a abstração de imagens proposta nestas pesquisas utilizava o filtro bilateral, em conjunção ao mapa de bordas obtido pela Diferença de Gaussianas. As melhorias propostas estavam relacionadas ao emprego dos elementos que compunham a imagem, obtidos após a computação do fluxo do campo de vetores da imagem, i.e., por meio da computação do Fluxo Tangente das Bordas (ETF - *Edge Tangent Flow*).

Após a obtenção de tal fluxo, era gerado o mapa de bordas a partir do uso da Diferença de Gaussianas Baseada em Fluxo (FDoG - *Flow-Based Difference Of Gaussians*), a qual favorecia a geração de contornos bem definidos. O Fluxo Tangente das Bordas foi utilizado por Kyprianidis, Kang e Döllner (2009), em conjunção com o filtro Kuwahara, um arranjo denominado filtro Kuwahara anisotrópico, destinado à renderização não fotorrealística ao estilo de cartuns (KYPRIANIDIS; KANG; DÖLLNER, 2009).

Kyprianidis, Kang e Döllner (2009) apresentaram uma generalização do filtro Kuwahara (KUWAHARA et al., 1976), um filtro de suavização que possibilita a preservação das bordas. Os autores propuseram uma adaptação do filtro que minimizava a presença de artefatos de agrupamento de regiões, fazendo adaptações no filtro Kuwahara, a fim de levar em conta a forma, a escala e a orientação da estrutura local da imagem de entrada, por meio do fluxo direcional das características da imagem. A máscara anisotrópica tinha um formato de elipse (Figura 2.11) sobre a qual suavização era aplicada, após sua subdivisão em setores triangulares.

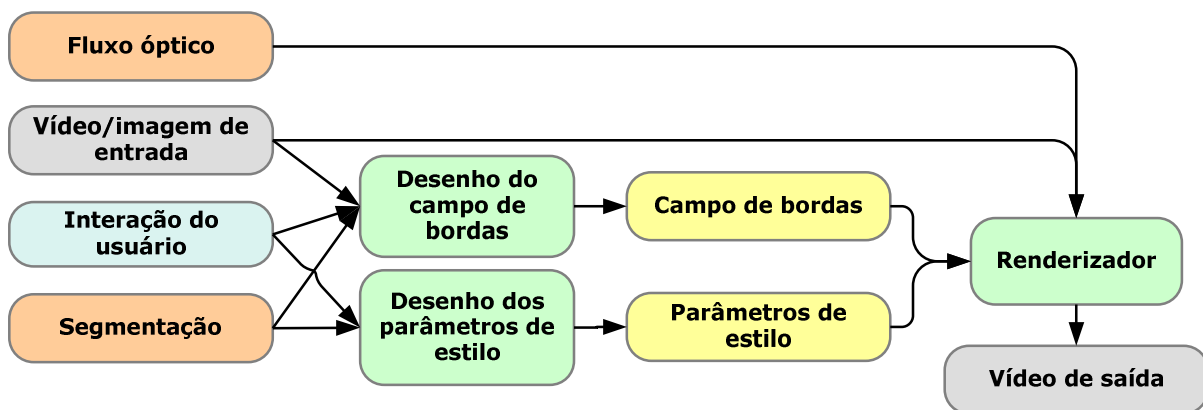
Figura 2.11 – Máscaras utilizadas no filtro Kuwahara: (A) Original, proposta por Kuwahara et al. (1976); e (B) Máscaras propostas por Kyprianidis, Kang e Döllner (2009).



A máscara elipsoidal era construída com base no gradiente local. A orientação e a média da máscara anisotrópica eram derivadas dos autovalores e autovetores do tensor de estrutura suavizada (definido por BROX et al., 2006). A resposta do filtro Kuwahara é definida por uma soma ponderada das médias locais, na qual maior peso é dado às áreas com menor desvio padrão. Kyprianidis, Kang e Döllner (2009) afirmaram que havia forte coerência temporal entre os quadros, quando a técnica era testada em vídeo, sem requerer o uso de técnicas para a estimativa do movimento em vídeo. A implementação em GPU confere à técnica proposta a capacidade de processamento de vídeos em tempo real.

Kagaya et al. (2010) ressaltaram a existência de técnicas que não empregavam o fluxo óptico para forçar a coerência temporal, tais como aquelas propostas por Klein et al. (2002), Agarwala et al (2004), Snavely et al. (2006), Wang et al. (2004). No entanto, a abordagem proposta por Kagaya et al. (2010) utilizava informações do fluxo óptico na abordagem destinada à renderização não fotorrealística. A macroarquitetura da abordagem proposta pelos autores pode ser visualizada na Figura 2.12. Atenção especial deve ser dada ao uso de técnicas de segmentação e de extração de bordas guiada pelo usuário, assim como a falta de uso de informação semântica da imagem.

Figura 2.12 – Arquitetura do sistema proposto por Kagaya et al. (2010).

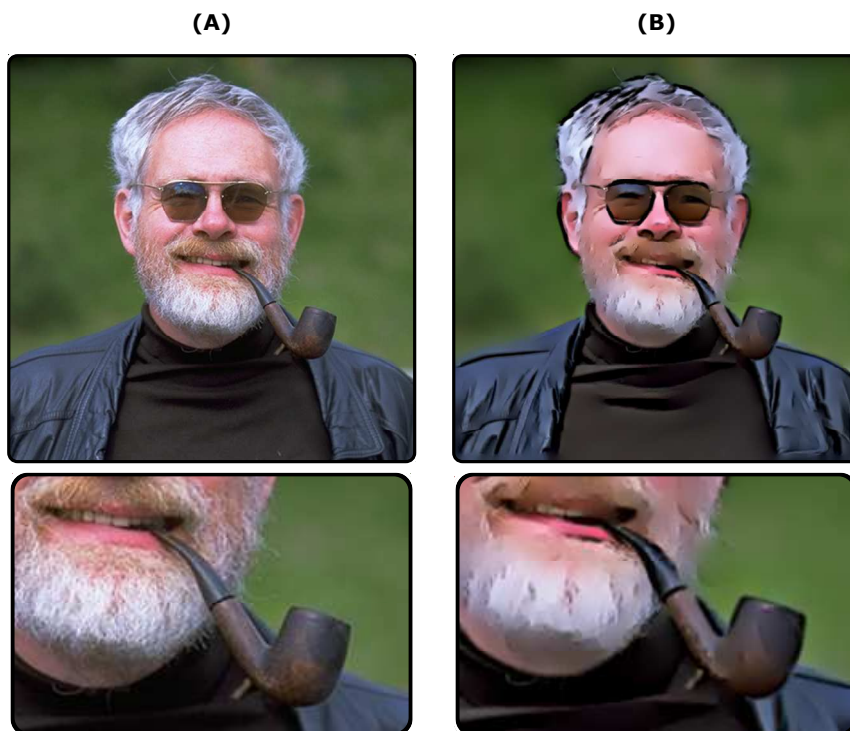


Bhat et al. (2010) apresentaram uma abordagem generalizada para a filtragem de imagens no domínio do gradiente. Uma filtragem de tal natureza tem como fundamento a construção de um campo de gradiente, i.e., a filtragem no domínio do gradiente afigura-se um tipo particular da filtragem no domínio espacial. No domínio do gradiente, há a manipulação dos valores dos pixels em conjunção à manipulação de diferenças entre pixels (gradientes de primeira ordem). A combinação destas duas características é feita por meio da solução de uma equação destinada à minimização da energia, a qual pode ser resolvida, por exemplo, a partir do método dos mínimos

quadrados¹⁴.

A abordagem de Bhat et al. (2010) também considerava uma métrica para a computação da saliência do gradiente local, a qual identifica as bordas mais proeminentes na imagem de entrada. Tal métrica beneficiava, por exemplo, a renderização não fotorrealística que se afigurou como uma aplicação possível da abordagem proposta pelos autores. Outras aplicações incluíam filtros para o aguçamento, a compressão, a pseudoiluminação e a recolorização de imagens. A abstração da imagem era feita por meio da suavização de características não salientes no domínio do gradiente, em conjunção com o exagero de características salientes da imagem, i.e., o filtro suprimia gradientes em regiões com poucas bordas e intensificava gradientes nos quais bordas salientes estivessem presentes. Um resultado da aplicação da abordagem proposta por Bhat et al. (2010) pode ser visualizada na Figura 2.13.

Figura 2.13 – Resultado da renderização obtida com a aplicação da abordagem proposta por Bhat et al. (2010): (A) Original; e (B) Renderizada.



Gastal e Oliveira (2011) observaram que uma imagem RGB podia ser interpretada como sendo uma transcrição (*manifold*) bidimensional em um espaço de cinco dimensões. Assim sendo, propuseram uma transformação isométrica de uma imagem bidimensional para dois sinais unidimensionais, na qual as distâncias entre pixels eram preservadas. Após a transformação, os autores aplicaram um filtro de

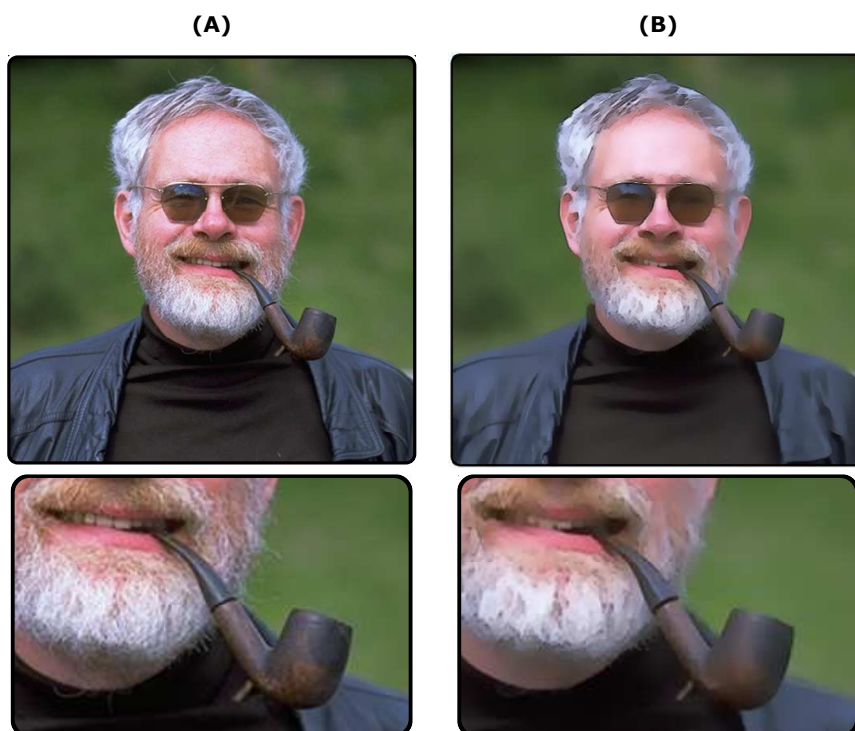
¹⁴ Método de ajuste para um conjunto de dados, realizado pela minimização da soma dos quadrados das diferenças entre o valor estimado e os valores observados.

suavização com preservação de bordas em apenas uma dimensão. Os autores afirmaram que, havendo a preservação das distâncias dos pontos da imagem no espaço unidimensional, a filtragem com preservação de bordas conservaria tais distâncias.

A abordagem proposta por Gastal e Oliveira (2011) está presente na comparação visual e na avaliação experimental realizada nesta pesquisa, devido ao fato de que os autores aplicaram a referida abordagem ao contexto da renderização não fotorrealística de imagens e vídeos. Adicionalmente, à semelhança do arcabouço proposto por Bhat et al, Gastal e Oliveira (2011) também aplicaram a abordagem proposta à pseudoiluminação, à recolorização e ao aguçamento de imagens ou quadros de vídeo.

Ling et al. (2012) relataram, após o uso da abordagem proposta por Gastal e Oliveira (2011), o aparecimento de halos e alterações tonais nas imagens testadas, os quais constituem efeitos indesejáveis introduzidos pela referida abordagem nos resultados gerados. Na Figura 4.12, o resultado corresponde a 3 iterações consecutivas da abordagem proposta por Gastal e Oliveira (2011), com o $\sigma_{espacial}$ ajustado para 10 e o σ_{radial} ajustado para 0,3, tendo sido considerada a opção NC (*normalized convolution*).

Figura 2.14 – Resultado da renderização obtida com a aplicação da abordagem proposta por Gastal e Oliveira (2011): (A) Original; e (B) Renderizada.



Winnemöller (2011) apresentou extensões do filtro DoG, a partir das quais o

método foi modificado para diminuir a sensibilidade a ruídos e visando a geração de resultados estilizados, em detrimento da geração exclusiva de um mapa de bordas. Os estilos obtidos com a aplicação da técnica incluem hachura, estilo pastel, renderizações em preto-e-branco.

A mudança de parâmetros do filtro DoG propostas por Winnemöller (2011) visaram à alteração da sensibilidade do filtro durante a detecção das bordas da imagem, a partir da utilização de uma função não binária (no caso, a tangente hiperbólica). As mudanças propostas pelo autor modificam a declividade da função tangente hiperbólica, gerando uma gama variada de efeitos estilísticos, inclusive efeitos de movimento. O autor ainda argumentou que a investigação de uma parametrização dessa natureza do filtro DoG podia beneficiar a área de NPR, posto que pesquisas posteriores poderiam empregar tal filtro em funções adicionais à geração do mapa de bordas.

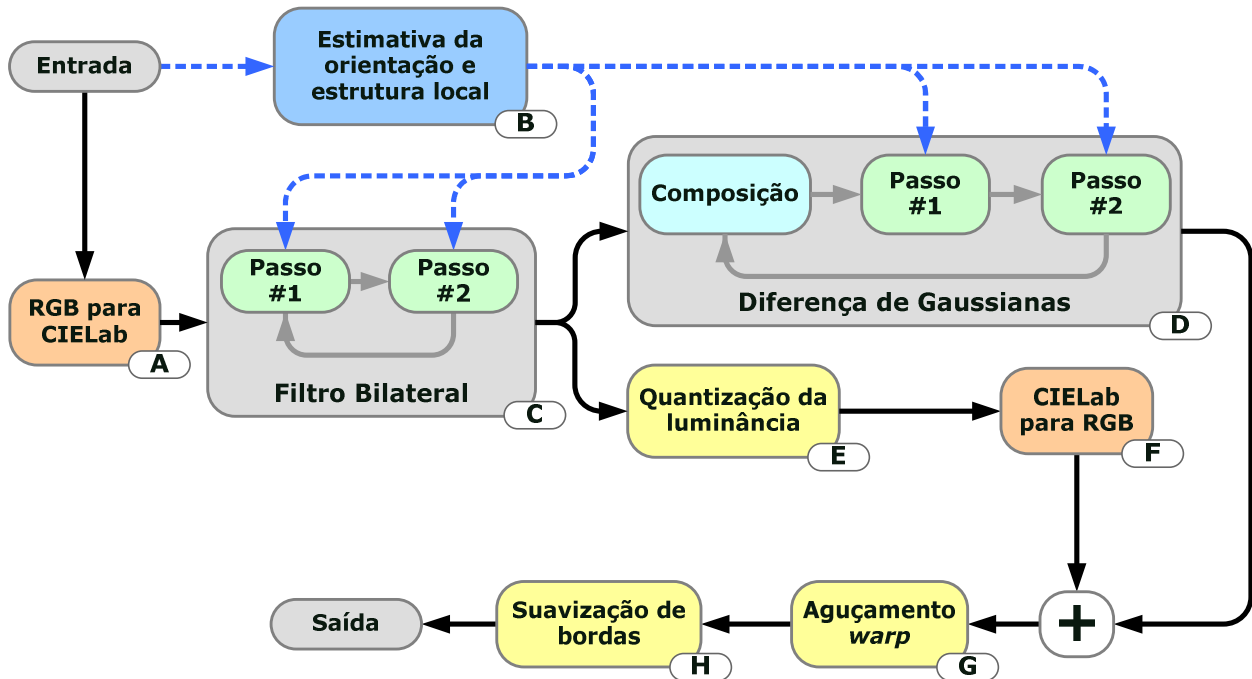
Kyprianidis et al. (2013) afirmaram que técnicas fundamentadas em filtragem de imagens no domínio espacial ofereciam uma alternativa interessante a métodos computacionalmente mais onerosos, tais como a computação do fluxo óptico ou a segmentação espaçotemporal de imagens. Adicionalmente, argumentaram que a paralelização do processo de filtragem espacial, em especial a partir do uso de unidades de processamento gráfico dedicadas (GPU), podia ser realizada de forma mais direta.

Adicionalmente, algumas técnicas de filtragem têm sido empregadas na renderização não fotorrealística de vídeos, tendo sido aplicadas quadro-a-quadro com bons resultados do ponto de vista da coerência temporal. Entretanto, tais vantagens reduzem a diversidade de estilos possíveis ao serem aplicadas técnicas de filtragem espacial.

Tal diversidade pode ser melhor explorada quando ocorre a interação com o usuário ou quando componentes de análise de alto nível do vídeo de entrada (e.g., o contexto da cena) estão presentes. A presença de um usuário facilita o refinamento no processo de estilização de oclusões de objetos, auxiliando também no refinamento de efeitos de movimento.

Nesse contexto, Kyprianidis (2013) propôs a generalização de um arcabouço para a geração de imagens renderizadas ao estilo de cartuns, composta por oito módulos. Uma visão esquemática da abordagem proposta por Kyprianidis (2013) pode ser visualizada na Figura 2.15.

Figura 2.15 – Abordagem generalizada proposta por Kyprianidis (2013) para renderização de imagens não fotorrealísticas ao estilo de cartum.



Em tal arcabouço, o processamento inicia-se com a conversão da imagem de entrada para o espaço de cor *CIELab* (módulo A). Em paralelo a essa conversão, uma representação intermediária é gerada, fundamentando-se na estimativa da orientação e na estrutura da imagem, a partir de um processo que emprega tipicamente a representação do fluxo de vetores ou ETF (módulo B).

Em seguida, a imagem é abstraída iterativamente (tipicamente, duas ou três iterações) por meio de um filtro bilateral modificado, destinado à remoção de ruído (módulo C). Após a filtragem bilateral, há a etapa de extração de bordas, na qual se emprega uma variante do filtro de Diferença de Gaussianas (DoG) (módulo D). O módulo de extração de integra a representação intermediária (módulo B) ao resultado do filtro bilateral (módulo C).

Em paralelo às operações realizadas pelo módulo D, mais iterações do filtro bilateral são realizadas, tipicamente quatro. Após a aplicação dessas iterações, a luminância da imagem é quantizada (módulo E) e, em seguida, ocorre a reconversão da imagem do espaço *CIELab* para o espaço RGB (módulo F).

A imagem gerada no módulo D é sobreposta àquela resultante do módulo F. A utilização dos módulos G e H é opcional. No módulo G, a imagem pode ser modificada a partir de um processo de realce de bordas (*warp sharp*), enquanto no módulo H as bordas podem ser suavizadas e sobrepostas à imagem final.

É importante salientar que a abordagem generalizada por Kyprianidis (2013) foi

proposta inicialmente por Winnemöller, Olsen e Gooch (2006), os quais empregaram o filtro bilateral modificado, computacionalmente mais eficiente. Na referida abordagem, a implementação computacional permitia o processamento da imagem em dimensões separadas (vertical e horizontal), a partir da utilização adicional de recursos disponíveis em placas de processamento gráficas dedicadas (GPU).

Na abordagem original de Winnemöller, Olsen e Gooch (2006), o módulo B não estava presente. Tal módulo, conforme descrito, é responsável pela computação da estimativa da orientação e da estrutura da imagem, agindo em conjunção com o filtro bilateral para produzir resultados de melhor qualidade. O passo D também foi modificado desde a proposição original feita por Winnemöller, Olsen e Gooch (2006), tendo o filtro DoG sido substituído pelo filtro detector de bordas baseado em fluxo (KANG; LEE; CHUI, 2007; KANG; LEE; CHUI, 2009; KYPRIANIDIS; DÖLLNER, 2008), o que elevou a qualidade dos esboços produzidos.

Salienta-se, neste ponto, que imagens filtradas pelo filtro bilateral costumam apresentar regiões de baixo contraste suavizadas, enquanto as regiões de alto contraste (bordas) são preservadas. No entanto, tal filtro pode apresentar limitações, e.g., quando há predominância de regiões de alto contraste na imagem a ser filtrada ou ainda quando a filtragem bilateral remove características visuais marcantes.

Adicionalmente, a aplicação iterativa da filtragem bilateral pode esmaecer as cores da imagem resultante, desfocando também as bordas. De acordo com Kyprianidis (2013), tal desbotamento pode ser atenuado por meio da sobreposição das bordas à imagem abstraída (ativando-se o módulo H). Adicionalmente, outro filtro destinado à preservação de bordas pode ser utilizado em substituição ao filtro bilateral, e.g., o filtro Kuwahara (1976), o filtro Kuwahara anisotrópico (KYPRIANIDIS; KANG; DÖLLNER, 2009) ou o filtro Kuwahara anisotrópico multiescala (KYPRIANIDIS, 2011).

Finalmente, ressalta-se que a abordagem proposta nesta pesquisa fundamentou-se no arcabouço apresentado na Figura 2.15. Em tal abordagem, há a presença de um módulo destinado a filtragem com preservação de bordas (semelhante ao módulo C), há a presença de um módulo destinado a detecção de bordas (semelhante ao módulo D), o qual realiza a filtragem por meio da utilização de redes neurais. Ressalta-se que, desde a imagem de entrada até o resultado não fotorrealístico produzido, o processamento assemelha-se àquele proposto na abordagem generalizada por Kyprianidis (2013). Detalhes e presentes nos módulos adicionais à abordagem de Kyprianidis (2013) podem ser visualizados no Capítulo 3.

A partir da exposição feita nos parágrafos anteriores, resumam-se no Quadro 2.1 as principais técnicas associadas à conversão de vídeos reais em vídeos renderizados não fotorrealisticamente ao estilo de *cartuns*.

Quadro 2.1 – Resumo das técnicas destinadas à renderização não fotorrealística de vídeos digitais.

(Legenda: ✓ = presente ; ✗ = ausente).

Técnica	Autor							
	Wang et al. (2004)	Huang (2006)	Collomosse, Rowntree e Hall (2005)	Winnemöller, Olsen e Gooch (2006)	Hong et al. (2007)	Lee et al. (2009)	Kagaya et al. (2010)	Kyprianidis, Kang e Döllner (2009)
Compensação de movimento de câmara	✗	✗	✓	✗	✗	✗	✗	✗
Redução de ruído	✗	✗	✗	✗	✓	✗	✗	✗
Separação manual fundo/tema	✓	✗	✓	✗	✓ (Auxiliada por <i>mean-shift</i>)	✗	✓ (Auxiliada por <i>Mean-shift</i>)	✗
Separação automática fundo/tema	✗	✓	✗	✗	✗	✗	✗	✗
Segmentação manual do objeto de interesse	✓	✗	✓	✗	✓	✗	✓	✗
Segmentação automática do objeto de interesse	✗	✓ (<i>Mean-shift</i> ou GMM)	✗	✗	✗	✗	✗	✗
Segmentação espacotemporal com reconstrução 3D	✓ (<i>Mean-shift</i>)	✓	✓	✗	✓ (<i>Mean-shift</i>)	✗	✓ (Segmentação por método próprio)	✗
Rastreamento de corpos não-rígidos	✓ (Semi-automático)	✓ (Simples)	✓	✗	✗	✗	✓	✗
Deteção de movimento usando fluxo óptico	✗	✗	✓	✗	✗	✓ (Análise complexa do movimento)	✓	✗
Renderizações de pinceladas artísticas	✗	✗	✓ (Temporalmente coerente)	✓ (Bilateral e quantização da luminância)	✗	✓	✓ (Pinceladas complexas)	✗
Deteção de bordas	✗	✗	✗	✓ (DoG)	✗	✓	✓	✓
Ênfase no movimento	✗	✓	✓ (Complexo)	✗	✗	✗	✗	✗
Uso de aspectos semânticos do vídeo	✗	✗	✓	✗	✗	✗	✗	✗
Suavização por fitragem anisotrópica.	✗	✗	✗	✗	✗	✗	✗	✓

É conveniente ressaltar que, no Quadro 2.1, a separação manual fundo/tema difere da segmentação manual do objeto de interesse pelo fato de algumas técnicas separarem e identificarem o que está em movimento como sendo o tema do vídeo. Outras técnicas, no entanto, deixam a cargo do usuário a seleção daquilo que será ou não renderizado, podendo ser o objeto de interesse pertencente ou não ao tema do vídeo.

Na próxima seção, é apresentado um levantamento bibliográfico de métodos destinados à avaliação subjetiva de processamento de imagens e vídeos, em especial as abordagens de avaliação de renderizações não fotorrealísticas.

2.3 Avaliação Subjetiva na área de Renderização Não Fotorrealística

Hertzmann (2010) enfatizou que há dificuldades na avaliação e na validação das teorias e dos avanços nas pesquisas no campo de NPR, tendo ressaltado que a avaliação experimental é pouco comum nas pesquisas da área e advertido que as teorias e as abordagens propostas para a renderização de imagens devem ser avaliadas a partir de dados coletados junto a usuários reais. O autor ainda afirmou que técnicas similares não deveriam ser apenas comparadas visualmente entre si, mas deveriam passar por um processo mais rigoroso de avaliação.

Adicionalmente, Hertzmann (2010) ressaltou que a avaliação experimental agrega rigor científico às pesquisas realizadas e que pesquisas na área de NPR têm adotado no processo de avaliação dos resultados uma dentre duas vertentes metodológicas mais expressivas, a saber: (i) *estudos de percepção*, a partir dos quais se busca elucidar as regras da percepção humana, ao apresentar, de forma controlada, estímulos específicos; e (ii) *estudos envolvendo usuários reais*, fundamentados na literatura da área de Interação Homem-Computador (*HCI - Human-Computer Interaction*), nos quais se avalia o desempenho dos usuários ao realizar tarefas específicas com o produto avaliado.

Collomosse e Kiprianidis (2011) corroboraram com a visão de Hertzmann (2010) no tocante à escassez de processos e dados referentes a avaliações experimentais de abordagens de NPR. Corroborando a argumentação apresentada por Hertzmann (2010), Kyprianidis et al. (2013) afirmaram que há pouca pesquisa destinada à estruturação de abordagens metodológicas para a avaliação e a validação de técnicas de NPR e que tais metodologias vêm se tornando cada vez mais importantes, tendo em vista que vários estilos tradicionais de renderização artística vêm sendo pesquisados, reproduzidos ou automatizados a partir de técnicas de NPR.

A reprodução ou automatização direcionaram as pesquisas para a proposição de novas formas de renderização ou para o aprimoramento dos métodos já consagrados na literatura da área. A redução do efeito de cintilação (*flickering*) após a segmentação de quadros de vídeo sucessivos, por exemplo, pode ser avaliada por meio de um processo quantitativo. Neste contexto, a investigação de novas formas de renderização tem evidenciado a crescente necessidade de avaliação dos resultados produzidos. Todavia, tal avaliação ocorre comumente de forma subjetiva, muitas vezes resumindo-se a comparações estéticas, com imagens apresentadas lado-a-lado, sendo muitas vezes a interpretação do resultado uma função da concordância ou não do leitor com o autor da pesquisa.

Isenberg (2013) discorreu brevemente a respeito de métodos de avaliação e validação em pesquisa científica, levantando questões acerca dos riscos inerentes à escolha de uma abordagem metodológica incorreta de validação ou da interpretação inadequada dos resultados de processos de avaliação (GREENBERG; BUXTON, 2008; HERTZMANN, 2010; COHEN, 1994; KAPTEIN; ROBERTSON, 2012). O autor também apontou publicações relacionadas a metodologias para avaliação no âmbito da pesquisa científica, as quais podem ser encontradas nas pesquisas realizadas por Kerlinger e Lee (2000) e Field e Hole (2003).

Diante da escassez de metodologias estruturadas para a avaliação experimental de técnicas de NPR, fez-se necessária a fundamentação relativa ao processo de avaliação, adotado no processo de validação da abordagem proposta nesta pesquisa. É importante salientar que a validação da abordagem proposta é um dos objetivos listados no Capítulo 1.

Apresenta-se, pois, um sumário dos principais aspectos relativos à pesquisa conduzida (Quadro 2.2). Em tal sumário são contemplados aspectos relativos à validação e avaliação de técnicas para renderização de imagens e vídeos. A apresentação de tal sumário visou a facilitar o entendimento geral de detalhes das metodologias para a validação de abordagens destinadas a renderização não fotorrealística de imagens estáticas e vídeos.

A seguir, são apresentados detalhes dos métodos adotados em pesquisas anteriores para a validação de abordagens para renderização não fotorrealística de imagens estáticas e vídeos digitais. Acrescenta-se a tal fundamentação outros métodos para a avaliação de vídeos digitais que fogem ao escopo da renderização não fotorrealística.

Quadro 2.2 – Sumário dos aspectos relativos à validação e avaliação de técnicas para renderização de imagens e vídeos.

Autor	Aspecto							
	Metodologia	Alvo investigado	Quantidade de participantes	Caracterização do participante	Tipo de instrumento	Tipo de classificação adotada	Quantidade	Aparato adicional
Schumann et al. (1996)	Qualitativa	Imagens CAD e estilizadas tradicionais	54	Arquitetos e estudantes de arquitetura	Questionário impresso	Likert de 5 pontos	-	-
Santella e DeCarlo (2004)	Qualitativa	Quatro versões de imagens abstraídas	74	Estudantes de curso de graduação	Rastreamento ocular	Mapa de calor, computação de pontos de interesse	50 imagens por participante	LCD 19", res. de 1240 x 960 pixels
Gooch, Reinhard e Gooch (2004) – Parte 1	Qualitativa	Imagens de faces de amigos	42	Estudantes de curso de graduação, pós-graduação e pesquisadores	Questionário <i>online</i>	Likert de 7 pontos (Familiaridade com a imagem)	12 imagens por participante (60 disponíveis)	Monitor a 60 cm do participante
Gooch, Reinhard e Gooch (2004) – Parte 2	Quantitativa	Imagens de faces desconhecidas	30 (3 grupos, separados por tipo de abstração)	Estudantes de de graduação, pós-graduação e pesquisadores	Questionário <i>online</i>	Tempo de reconhecimento de (5 segundos, no máximo)	12 imagens por participante (60 disponíveis)	Monitor a 60 cm do participante
Winnemöller, Olsen e Gooch (2006)	Quantitativa	Imagens de faces de celebridades	10 (5 homens)	Estudantes de graduação	Jogo da memória (digital)	Tempo de conclusão do jogo	24 imagens	-
Isenberg et al. (2006)	Qualitativa	Modelo arqueológico, botânico e médico	24	Pesquisadores de NPR, estudantes universitários e especialistas em desenho	Empilhamento livre de imagens impressas	Sem perguntas diretas. (Entrevista semiestruturada pós-teste)	30 imagens (15 estilizadas por artistas, 15 estilizadas por computador)	-
Seshadrinathan et al. (2010)	Qualitativa	10 vídeos distintos (L.I.V.E. database)	38	Estudantes de graduação	Questionários <i>online</i>	Likert de 5 pontos, preferência direta.	10 vídeos de referência, 160 distorcidos	Resolução de 1024 x 768 pixels
Mould, Mandryk e Li (2012)	Qualitativa	Imagens que evocam emoções	42	Pesquisadores de NPR, estudantes, artistas,	Questionário <i>online</i>	Likert de 5 pontos. (3 escalas para emoção e 1 escala para estética)	18 (IAPS ¹⁵)	-
Mantiuk, Tomaszewska e Mantiuk (2012)	Qualitativa	Kodak Photo CD, distorcidas com compressão JPEG2K ou máscara <i>unsharp</i>	28	Usuário principiante	Questionários <i>online</i>	Likert de 5 pontos ou preferência direta, a depender do método avaliado	10 imagens de referência (animais, faces e natureza)	Monitor NEC SpectraView 2690, res. de 1920 x 1200
Li, Mould e Davies (2013)	Quantitativa e qualitativa	Carro, gato, pássaro, pessoa, caneca, flor, prédio	30 (15 mulheres)	Artistas e leigos	Questionários <i>online</i>	Sim/Não (quantitativa) Likert de 5 pontos (qualitativa)	12	LCD 21", resol. 1440 x 900 pixels

¹⁵ *International Affective Picture System.*

Isenberg (2013) afirmou que, de forma ampla, a natureza das metodologias comumente adotadas na pesquisa científica para a avaliação de resultados recai em uma de duas vertentes, a saber: (i) *quantitativa*, quando a metodologia é focada em testes de hipóteses, na quantificação de variáveis por meio de experimentos controlados e (culminando) na análise estatística dos resultados; e (ii) *qualitativa*, quando a metodologia é focada na melhoria da compreensão do tema pesquisado, a partir de uma abordagem holística¹⁶, na qual podem ser empregadas técnicas de observação ou entrevistas (CARPENDALE, 2008).

Ambas as categorias (quantitativa e qualitativa) ou ainda sua combinação têm sido aplicadas à avaliação de abordagens de NPR. Decerto, a escolha de uma categoria específica de metodologia a adotar deverá estar alinhada aos objetivos da pesquisa conduzida e aos questionamentos que esta se propuser a responder, alinhada às hipóteses a serem testadas. Por exemplo, Isenberg et al. (2006) adotaram uma metodologia qualitativa, a qual elucidou questões referentes à opinião de participantes no tocante a ilustrações desenhadas à mão, quando confrontadas com ilustrações geradas por computador, enquanto a pesquisa quantitativa conduzida por Gooch, Reinhard e Gooch (2004) intentou elucidar diferenças na tarefa de reconhecimento de faces humanas, confrontando representações estilizadas e representações fotorrealísticas.

No contexto da avaliação e da análise de reações emocionais a representações estilizadas, Schumann et al. (1996) apresentaram um estudo sobre o uso de desenhos feitos à mão no campo da arquitetura e de representações CAD¹⁷ (STROTHOTTE et al., 1994). Os autores examinaram o efeito que a apresentação de projetos arquitetônicos, renderizados em forma de esboços, têm sobre os objetivos da comunicação (entre o arquiteto e os clientes) durante o desenvolvimento de projetos arquitetônicos.

Schumann et al. (1996) investigaram três hipóteses: (i) imagens do projeto arquitetônico, apresentadas ao estilo de esboços, são preferíveis a imagens CAD tradicionais, na fase inicial do projeto; (ii) a apresentação ao estilo de esboços obtém melhor desempenho na comunicação de aspectos afetivos e motivacionais, enquanto plotagens CAD tradicionais obtém melhor desempenho em aspectos cognitivos; e (iii) imagens apresentadas ao estilo de esboços estimulam os espectadores/clientes a

¹⁶ Conforme o Dicionário Houaiss da Língua Portuguesa (2011), o termo *holístico* refere-se a uma abordagem que busca entendimento integral dos fenômenos. Em especial, destacam-se metodologias de avaliação nos campos das Ciências Humanas e Naturais, nos quais abordagens holísticas se opõem ao procedimento analítico em que componentes são tomados isoladamente.

¹⁷ CAD – *Computer Aided Design*: Desenho Assistido por computador (DAC) ou CAD, denominação atribuída a sistemas computacionais utilizados, e.g., nas áreas de engenharia, geologia, geografia, arquitetura e design, visando facilitar a confecção de projetos e desenhos técnicos.

participarem mais ativamente da discussão durante o desenvolvimento do projeto arquitetural do que imagens CAD tradicionais.

Visando à elucidação das três hipóteses apresentadas, Schumann et al. (1996) conduziram um experimento quantitativo destinado a identificar a preferência direta e a melhoria na comunicação de idéias de projetos por meio da utilização de renderizações não fotorrealísticas. Após a verificação dos resultados, a primeira hipótese foi aceita, indicando que renderizações estilísticas podem ser usadas durante etapas iniciais em projetos arquiteturais. Tais renderizações foram usadas também durante processos de interação e modelagem que utilizam esboços como fonte primária (IGARASHI; MATSUOKA; TANAKA, 1999; SCHMIDT et al., 2007).

Na investigação da segunda hipótese, Schumann et al. (1996) solicitaram aos participantes a avaliação detalhada das impressões ou emoções causadas por três imagens distintas, adotando uma classificação advinda da área da Psicologia. Em tal classificação, os participantes classificavam cada imagem de acordo com três critérios: cognitivo, afetivo, motivacional. Após a classificação, esboços gerados com o uso de técnicas de NPR foram avaliados como significativamente maiores/melhores nos critérios afetivos e motivacionais. Imagens CAD foram classificadas como significativamente maiores/melhores em critérios cognitivos.

Ao investigar a terceira hipótese, os participantes foram questionados acerca da forma usada para comunicar alterações no projeto, na qual quatro alternativas foram exibidas, relativas à preferência direta ao comunicar aos clientes alterações no projeto a partir de: (a) descrições verbais; (b) gestos ou apontamento dos locais de alteração; (c) desenhos, em folhas de papel separadas do desenho CAD; ou (d) desenhos/rabiscos feitos diretamente na imagem apresentada.

A diferença estatisticamente significativa foi encontrada quando a alternativa (d) foi comparada, i.e., os autores registraram diferenças entre a participação ou interação do usuário quando esboços NPR e imagens CAD eram utilizados. Tal diferença revelou que os participantes estavam mais propensos a desenhar/rabiscar diretamente na imagem do esboço NPR, confirmando a terceira hipótese dos autores e apontando que o estilo de apresentação dessas renderizações pode interferir no comportamento e na interação entre o arquiteto e os clientes, nos estágios iniciais de projetos arquitetônicos.

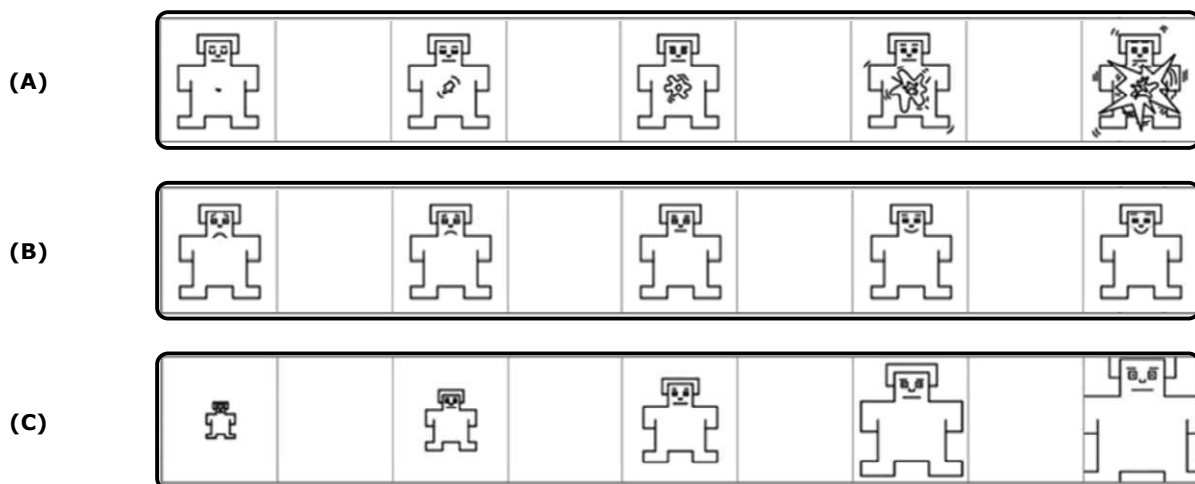
No tocante ao efeito de renderizações NPR em aspectos afetivos e motivacionais e, ainda, no tocante à utilização de NPR para guiar a atenção do espectador, Mandryk, Mould e Li (2011) e Mould, Mandryk e Li (2012) aprofundaram os estudos acerca do

efeito de renderizações estilísticas na evocação de emoções. Mais especificamente, os autores analisaram a forma pela qual a estilização de fotografias muda a resposta emocional de participantes humanos em relação a imagens apresentadas.

Ao realizar um experimento, foram utilizadas imagens do banco de imagens IAPS. Tais imagens, também empregadas em pesquisas no campo da Psicologia, retratam uma ampla gama de temas, tendo sido o banco criado especificamente para a verificação de estímulos afetivos a partir de conteúdo emocional conhecido e previamente validado.

A resposta emocional dos participantes foi sondada a partir do emprego de um conjunto de escalas comumente adotado em avaliações psicológicas, composto por três aspectos emocionais. A votação nestas escalas permitiu aos participantes analisarem e relatarem o estado emocional ao visualizar as imagens estilizadas. A escala divide-se em três dimensões: (i) valência (*valence*), que indica atração ou repulsa; (ii) grau de excitação (*arousal*), que indica excitação ou sonolência; e (iii) dominância (*dominance*), que indica o controle ou a submissão/dominância do participante a um dado sentimento. Tais escalas podem ser visualizadas na Figura 2.16.

Figura 2.16 – Escala pictórica de classificação emocional: (A) Excitação; (B) Valência; e (C) Dominância.



Fonte: Mould, Mandryk e Li (2012).

Em adição à avaliação emocional a partir da escala pictórica apresentada anteriormente, o participante deveria apresentar a opinião sobre a estética da figura apresentada. A avaliação do aspecto estético foi feita por meio da votação, em uma escala de Likert de 5 pontos, relativa à seguinte sentença: "Avalie seu grau de concordância à seguinte afirmação: A qualidade estética da imagem apresentada é agradável". Na escala apresentada foram utilizadas as seguintes âncoras semânticas: *discordo totalmente*, *discordo*, *neutro*, *concordo* e *concordo totalmente*.

Mandryk, Mould e Li (2011) e Mould, Mandryk e Li (2012) compararam a resposta emocional das imagens originais, cuja classificação prévia já havia sido estudada em outras pesquisas psicológicas e tinham a resposta emocional conhecida, com a resposta emocional dos participantes às imagens estilizadas com as seguintes técnicas: pontilhismo (SECORD, 2002); desenho de linhas e silhuetas (KANG; LEE; CHUI, 2007); efeito de pintura (HAEBERLI, 1990); efeito de pintura abstrata (ZHAO; ZHU, 2010); abstração fotográfica (ORZAN et al., 2007); desfocagem do objeto principal da cena; e desfocagem uniforme da cena.

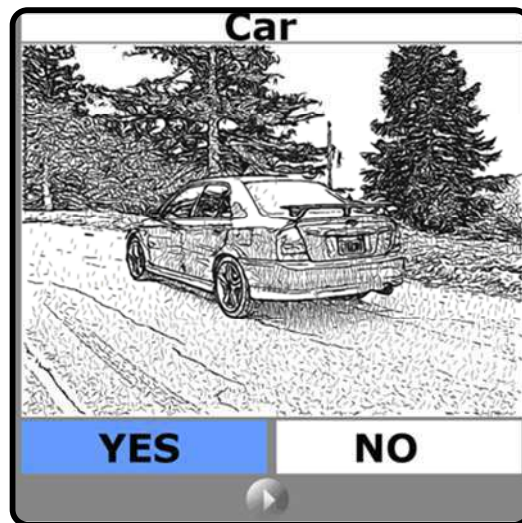
Os resultados apresentados por Mandryk, Mould e Li (2011) e Mould, Mandryk e Li (2012) foram tais que todas as técnicas de NPR consideradas no experimento de comparação mudaram de forma significativa as experiências relatadas pelos participantes. As experiências relatadas mudaram no aspecto da valência (prazer/desprazer) e excitação (energia/ativação de um sentimento), levando aos participantes a classificarem as imagens renderizadas de forma neutra, i.e., a estilização reduziu a emoção que o usuário sentiu ao realizar o teste, sem eliminá-la completamente. Os autores relataram que algumas técnicas de renderização estilística preservam melhor as emoções dos participantes, na medida em que o efeito de abstração preserva mais ou menos detalhes da cena renderizada.

Li, Mould e Davies (2013) afirmaram que processos de avaliação e validação de técnicas de NPR recorreram a medidas de similaridade adotadas na área de PDI. No entanto, os autores afirmaram não haver forma objetiva e padronizada de mensuração da qualidade de estilizações. Ademais, o processo de comparação estética, com a apresentação de imagens lado-a-lado, muitas vezes não consegue ser convincente, devido ao número reduzido de imagens retratadas ou a escolha subjetiva e restrita das imagens a serem comparadas.

Na tentativa de preencher essa lacuna, Li, Mould e Davies (2013) realizaram um estudo com participantes humanos, no qual foram investigadas duas hipóteses: (i) a escolha de um determinado tipo de imagem influencia a comparação de técnicas NPR, quando comparadas lado-a-lado; e (ii) há relação entre a qualidade estrutural e qualidade estética de imagens não fotorrealísticas (exclusivamente renderizações em escala de cinza). A investigação da primeira hipótese foi realizada por meio de um processo de verificação do tempo de resposta do participante. Após a exibição de textos e imagens renderizadas, o participante deveria responder se havia, na imagem renderizada, o objeto descrito na palavra. Se o participante cometesse um erro, este não poderia votar novamente. Um excerto desta etapa pode ser visualizado na Figura

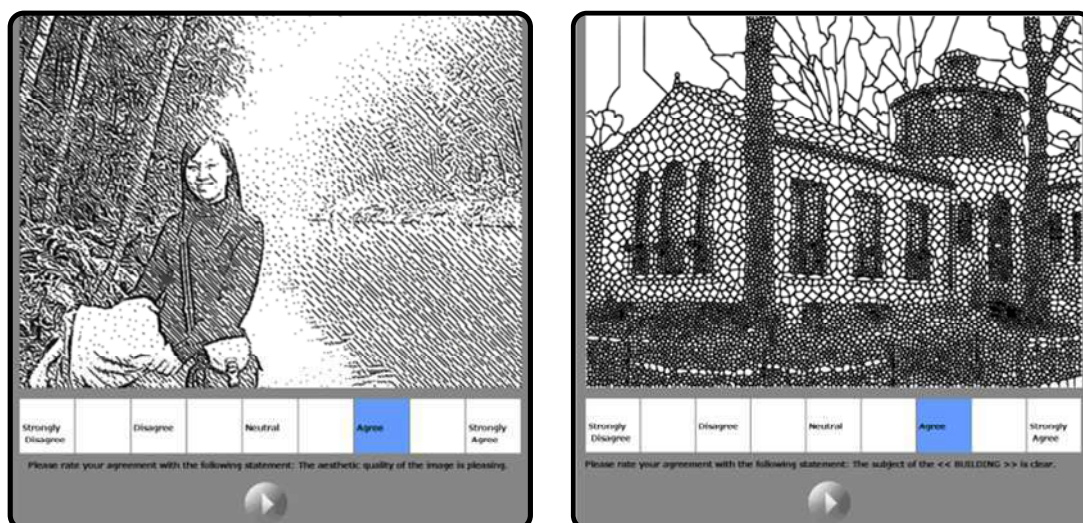
2.17, no qual um exemplo de renderização pode ser associado à palavra exibida.

Figura 2.17 – Excerto da interface computacional utilizada por Li, Mould e Davies (2013) durante a coleta de dados de tempo de resposta à renderizações NPR.



A investigação da segunda hipótese foi realizada por meio de um experimento no qual era exibida uma imagem, duas vezes, para a qual o participante votava em uma escala associada à instrução: "Avalie seu grau de concordância com a seguinte afirmação: A qualidade estética da imagem apresentada é agradável". Outra escala era apresentada, referente à qualidade estrutural, a partir da qual o participante avaliava a clareza do objeto que pertencia a uma determinada categoria, conforme a seguinte instrução: "Avalie seu grau de concordância com a seguinte afirmação: O assunto da <CATEGORIA> está apresentado de forma clara". Um excerto da interface empregada no experimento pode ser visualizado na Figura 2.18, compreendendo as dimensões estrutural e estética.

Figura 2.18 – Excerto da interface computacional utilizada por Li, Mould e Davies (2013) durante a coleta de dados de qualidade de imagens NPR: (A) Estrutural; e (B) Estética.



É importante ressaltar que os participantes receberam treinamento prévio,

votando e classificando um número pequeno de imagens, realizado na interface que seria utilizada posteriormente, para que se familiarizassem com o ambiente e com a tarefa a ser desempenhada.

Os participantes preencheram um questionário pós-teste, no qual foram registradas informações acerca de seu perfil, suas preferências artísticas, além de comentários sobre melhorias no experimento ou nas imagens renderizadas consideradas no processo de avaliação.

O estudo conduzido por Li, Mould e Davies (2013) revelou uma correlação estatisticamente significativa entre a estética e a estrutura das imagens renderizadas, i.e., o aumento da classificação estrutural coincidiu com o aumento da classificação estética. Adicionalmente, as classificações de estrutura e estética dos participantes foram influenciadas pelo conteúdo da imagem, i.e., a escolha de imagens de entrada influenciou os resultados quando as técnicas de renderização NPR consideradas foram avaliadas.

Strothotte et al. (1994) mencionaram que renderizações estilísticas poderiam guiar a atenção de participantes em um experimento. Tal menção foi estudada em profundidade por Santella e DeCarlo (2004), os quais conduziram um experimento que visava a verificar o direcionamento da atenção visual dos participantes. O objetivo dos pesquisadores era compreender a eficácia de veicular informação quando renderizações abstraídas eram utilizadas, a partir da aquisição de dados via rastreamento ocular, i.e., a hipótese a ser verificada era se a remoção proposital de detalhes podia orientar a atenção visual humana.

Para verificar tal hipótese, Santella e DeCarlo (2004) criaram quatro versões abstraídas de uma mesma imagem de entrada: (i) abstração *constante*, com uma elevada quantidade de detalhes; (ii) abstração *homogênea* em toda a imagem, resultando em uma baixa quantidade de detalhes; (iii) abstração *adaptativa 1*, na qual detalhes da imagem têm maior ou menor grau de abstração de acordo com o mapa de saliência computado, i.e., maior saliência significa menor abstração; e (iv) abstração *adaptativa 2*, na qual detalhes a serem renderizados foram baseados em pontos obtidos a partir do rastreamento ocular de outros participantes que observaram a fotografia original, i.e., pontos de fixação no rastreamento ocular significaram menor abstração.

Os resultados obtidos por Santella e DeCarlo (2004) evidenciaram que o tratamento local da abstração, a partir de informações de rastreamento ocular, tiveram maior efeito sobre o local nos quais os participantes fixavam o olhar na

imagem. Foi observado que o efeito de renderização obtido a partir da utilização da informação de saliência da imagem, obtidos por meio do rastreamento ocular, receberam maior quantidade de observações (maior incidência de observações registradas no mapa de calor do rastreamento ocular), quando comparadas a renderizações homogêneas. A renderização homogênea recebeu menor incidência de observações do que todas as demais formas de renderização, inclusive da imagem original.

Adicionalmente, Santella e DeCarlo (2004) observaram divergências significativas entre o mapa de saliências obtido por meio da aplicação de algoritmos automáticos daqueles obtidos a partir do rastreamento ocular de participantes humanos, sugerindo que a obtenção do mapa de saliência poderia ser melhorada, i.e., o mapa de saliência obtido de forma automática deveria conter menos pontos de interesse, porém pontos que fossem mais significativos.

Isenberg (2013) ressaltou que os resultados obtidos por Santella e DeCarlo (2004) forneceram evidências que validavam a hipótese mencionada, i.e., o controle da quantidade de detalhes, possível a partir da aplicação de técnicas de NPR, podia ser empregado para guiar a atenção visual de participantes humanos. Em conjunção aos resultados obtidos por Santella e DeCarlo (2004), no tocante ao uso de técnicas de renderização estilística para guiar atenção de indivíduos, investigou-se também se técnicas de NPR podiam ter efeito positivo na capacidade de indivíduos reconhecerem ou memorizarem objetos ou faces humanas.

Tal hipótese foi investigada por Gooch, Reinhard e Gooch (2004), especificamente na tarefa de reconhecimento de faces humanas. Por meio de um teste psicofisiológico, no qual foram comparadas fotografias de faces humanas às respectivas renderizações (ilustrações e caricaturas), os autores realizaram experimentos quantitativos com o intuito de medir a velocidade e precisão dos participantes em tarefas de reconhecimento facial de rostos de indivíduos conhecidos. Na primeira parte do experimento, os participantes deveriam identificar/reconhecer corretamente os colegas retratados nas imagens, que foram apresentadas de forma aleatória.

Nesta primeira etapa, a cada participante do experimento foram exibidas apenas imagens de duas naturezas, dentre as três disponíveis - fotografias, ilustrações e caricaturas. Ao analisar os resultados do experimento, Gooch, Reinhard e Gooch (2004) ressaltaram que os participantes mostraram maior velocidade de reconhecimento e identificação dos rostos presentes nas fotografias do que de

reconhecimento e identificação dos rostos presentes nas caricaturas. Entretanto, não houve diferença significativa entre as outras combinações consideradas, tendo havido uma alta taxa de reconhecimento na identificação de rostos familiares.

Na segunda etapa do experimento realizado por Gooch, Reinhard e Gooch (2004), os participantes deveriam identificar e nomear faces desconhecidas. As fotos de faces foram exibidas a cada participante, e para cada imagem apresentada havia um nome associado. Cada participante foi exposto apenas a imagens de uma única natureza, dentre aquelas disponíveis. Após a exibição das imagens ao participante, o conjunto de imagens era reorganizado e os participantes deveriam associar o nome previamente exibido às imagens de face, as quais eram exibidas uma por vez.

Após a computação dos resultados, Gooch, Reinhard e Gooch (2004) evidenciaram que o processo de associação de ilustrações aos nomes correspondentes foi duas vezes mais rápido que o processo de associação de fotografias aos nomes correspondentes. O processo de associação de caricaturas aos nomes correspondentes foi 1,5 vez mais rápido que o processo de associação de fotografias aos nomes correspondentes, mas tal diferença não se mostrou estatisticamente significativa. Tais resultados indicaram que técnicas de NPR podem ter efeito positivo na capacidade de indivíduos de reconhecerem ou memorizarem faces humanas, para o tipo de abstração e para a população específica considerada por Gooch, Reinhard e Gooch (2004).

Em outro experimento similar àquele realizado por Gooch, Reinhard e Gooch (2004), Winnemöller, Olsen e Gooch (2006) propuseram e avaliaram uma abordagem destinada à renderização de vídeos em tempo real. Ressalta-se, novamente, que tal técnica, descrita na seção 2.3, serviu de inspiração para a pesquisa conduzida, fundamentando a abordagem proposta, que é detalhada no Capítulo 3.

Winnemöller, Olsen e Gooch (2006) avaliaram a abordagem proposta com a utilização de imagens de faces, já que a abordagem proposta mostrou-se adequada tanto à geração de abstrações de imagens quanto à geração de abstrações de vídeos. Os autores verificaram a habilidade dos participantes na tarefa de reconhecimento e memorização, de faces de pessoas famosas/celebridades, confrontando fotografias e abstrações. Um diferencial de tal avaliação é o fato das imagens comparadas terem sido imagens coloridas, em contraponto às imagens em preto-e-branco presentes na pesquisa realizada por Gooch, Reinhard e Gooch (2004).

Winnemöller, Olsen e Gooch (2006) utilizaram imagens em um jogo da memória, desenvolvido de forma digital, no qual foi medido o tempo de conclusão do

jogo por cada participante. No total, os participantes concluíram o jogo da memória em menor tempo (59,95 segundos) e com menor quantidade de cliques (49,2 cliques) quando comparado ao jogo no qual foram usadas fotografias (76,13 segundos, 62,4 cliques). Os resultados corroboraram evidências elucidadas por Gooch, Reinhard e Gooch (2004), i.e., a tarefa de reconhecimento de rostos de celebridades foi significativamente mais rápida quando abstrações de imagens foram utilizadas. Os autores ressaltaram que os resultados obtidos não podiam ser generalizados para qualquer tipo de estilização NPR.

Isenberg (2013) ressaltou outra aplicação de renderizações não fotorrealísticas, denominada *visualização ilustrativa (illustrative visualization)*, a qual é subdomínio do campo da visualização interativa de dados científicos e de informações (RAUTEK et al., 2008). Em tal subdomínio, Isenberg (2013) ressaltou que a avaliação dos mecanismos de visualização ilustrativa é imprescindível, e nesta área é realizado um tipo de avaliação diferente daquelas abordadas anteriormente: seções de avaliação por especialistas (ou sessões de crítica), tipicamente empregada na avaliação de obras artísticas.

No tocante à área de visualização ilustrativa, Jackson et al. (2003) e Feliz, Laidlaw e Drury (2005) pesquisaram acerca de sessões de crítica realizadas com *designers* e educadores especialistas em ilustração, considerando a avaliação de várias técnicas de visualização, em especial técnicas de visualização de campos de vetores bidimensionais. As seções de avaliação por especialistas, enquanto estratégias de avaliação, fornecem um retorno qualitativo, por vezes rico em detalhes, que pode indicar a técnica mais adequada à melhoria na visualização de dados sob investigação. Tal retorno pode não acontecer quando uma abordagem quantitativa é adotada.

Feliz, Laidlaw e Drury (2005) relataram que seções de avaliação envolvendo especialistas foram conduzidas em tempo reduzido, quando comparadas àquelas envolvendo leigos. Laidlaw et al. (2005) e Keefe et al. (2005) analisaram as seções de avaliação envolvendo especialistas, no contexto da colaboração entre ciência e arte, realçando o papel da avaliação no fluxo de trabalho de um artista. O fluxo de trabalho de tais artistas é levado em consideração durante o processo de concepção e desenvolvimento de ferramentas computacionais destinadas à visualização estilizada.

No contexto da avaliação quantitativa e qualitativa de renderizações estilísticas, Isenberg et al. (2006) foram os primeiros a apresentar uma metodologia para a coleta de dados de participantes que comparavam, por meio de um experimento, quanto uma renderização NPR se assemelhava a imagens desenhadas manualmente por

artistas. Os autores definiram o teste de Turing para imagens NPR, i.e., o participante do experimento deve indicar se é possível distinguir (ou não) uma imagem feita por um artista de uma imagem gerada usando técnicas de NPR (SALESIN, 2002; GOOCH et al., 2010).

Isenberg et al. (2006) argumentaram que, na tentativa de investigar se as renderizações NPR passavam no teste de Turing, a adoção de medidas quantitativas fornecidas por alguns métodos mencionados anteriormente poderiam não fornecer uma resposta satisfatória. Assim sendo, uma pesquisa qualitativa foi desenvolvida para tentar elucidar três perguntas realizadas pelos pesquisadores: (i) *como os seres humanos entendiam e avaliavam imagens NPR?*; (ii) *o que os seres humanos achavam de imagens feitas a mão e imagens renderizadas por computador?*; (iii) *diferenças na avaliação de imagens feitas a mão e renderizadas por computador podiam significar uma preferência específica por um tipo de imagem?*

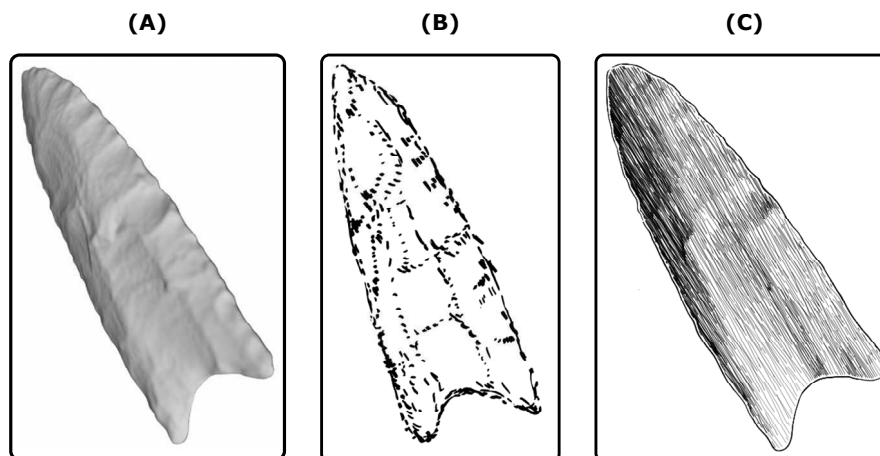
Isenberg et al. (2006) argumentam que tais questões não são facilmente respondidas apenas com o uso de métodos de avaliação quantitativa, resposta a questionários, ou ainda avaliação de desempenho em uma tarefa específica (e.g., um jogo de memória), mas requerem uma abordagem qualitativa para serem respondidas.

Para avaliar tais questões, Isenberg et al. (2006) realizaram um experimento observacional no qual os participantes não recebiam perguntas diretas a respeito do tema tratado. A ausência de perguntas diretas, de acordo com os autores, serviu para evitar algum viés no resultado do experimento. Em contrapartida, o experimento requeria do participante que ordenasse várias imagens, empilhando-as, na ordem de preferência do participante, em quantas novas pilhas fossem necessária, com qualquer número de imagens por pilha, na ordem que o participante desejasse. Mais precisamente, durante o experimento, os usuários agruparam imagens de acordo com critérios pessoais, sem critérios pré-definidos pelo pesquisador, o qual realizou, posteriormente, uma entrevista semiestruturada, a fim de evidenciar os critérios utilizados.

Três imagens, de diferentes naturezas, foram utilizadas no experimento conduzido por Isenberg et al. (2006): uma imagem de um modelo arqueológico, uma imagem de um modelo botânico e uma imagem de um modelo médico. As imagens foram desenhadas por três artistas diferentes, e foram renderizadas a partir de três técnicas de NPR diferentes. Um exemplo de renderização empregada no experimento de Isenberg et al. (2006) pode ser visualizado na Figura 2.19, na qual o modelo

arqueológico original (A) foi retratado por um artista (B) e renderizado por uma técnica de NPR (C).

Figura 2.19 – Modelo arqueológico utilizado na pesquisa realizada por Isenberg et al. (2006).



Após a realização do passo de ordenação em pilhas de imagens, cada participante deveria explicar ao pesquisador a motivação usada para gerar cada pilha. Após tal explicação, outras perguntas foram feitas acerca: (i) da preferência pessoal das imagens apresentadas; (ii) do contexto potencial de uso das imagens apresentadas; e (iii) do teste de Turing para imagens NPR, i.e., se a imagem parecia desenhada a mão ou renderizada por um computador.

Isenberg et al. (2006) analisaram os resultados da pesquisa e concluíram que participantes de perfis diferentes empilharam as imagens de forma similar e que participantes agruparam por estilo de ilustração ou pela quantidade de detalhes na imagem. No entanto, os participantes foram capazes de discernir entre imagens renderizadas por computador de imagens feitas à mão, por artistas. No entanto, essa diferença não significou preferência de um tipo de imagem a outro ou preferência direta de um tipo de renderização a outro, mas os participantes gostavam de uma determinada imagem por causa da clareza, precisão, tridimensionalidade, detalhe ou, ainda, apreciavam o aspecto artístico das ilustrações desenhadas à mão.

Na pesquisa realizada por Isenberg et al. (2006), os resultados dos experimentos evidenciaram que as abstrações geradas por técnicas de NPR foram classificadas, na maior parte das vezes, como sendo gerados por computador, o que invalidou a hipótese investigada, que era a geração de imagens que imitassem desenhos feitos à mão.

Após a realização do estudo qualitativo, Isenberg et al. (2006) forneceram recomendações e orientações para a área de NPR acerca da necessidade de conhecimento do objetivo da renderização, da necessidade de conhecimento do

público-alvo e da exploração da renderização de materiais de forma pseudorrandômica (evitando regularidades na textura do material retratado).

Diversas metodologias utilizadas na avaliação e a validação de abordagens destinadas à geração de renderizações não fotorrealísticas foram descritas, revelando uma miríade de questões que se pode querer responder a respeito das renderizações produzidas. Por exemplo, o seguinte questionamento surgiu: uma imagem renderizada ao estilo NPR pode ser indistinguível de uma imagem feita à mão por um artista humano? Nesse teste de Turing para imagens NPR, proposto por Salesin (2002), renderizações NPR ainda não obtiveram sucesso, como demonstrado por observações feitas por Isenberg et al. (2006) ou por estatísticas de síntese de texturas por Maciejewski et al. (2008).

Entretanto, ser indistinguível de uma obra humana pode não ser o objetivo central da área de NPR, i.e., pesquisas relativas a avaliação de renderizações NPR podem investigar outras questões, e.g., a motivação para o uso de renderizações de NPR ou elucidar o processo de julgamento da estética de tais renderizações (GERSTNER et al., 2012).

Os estudos mencionados anteriormente fornecem evidências dos benefícios ou efeitos de representações estilísticas obtidas usando técnicas de NPR, as quais podem ser utilizadas enquanto efeitos visuais para auxiliar discussões iniciais em projetos arquitetônicos (SCHUMANN et al., 1996), guiar o olhar (DUKE et al., 2003; HALPER et al., 2003a, 2003b) e afetar a emoção de indivíduos (MANDRYK; MOULD; LI, 2011; MOULD; MANDRYK; LI, 2012), melhorar o desempenho de indivíduos em tarefas específicas (LI; MOULD; DAVIES, 2013) e/ou afetar o modo a partir do qual indivíduos reconhecem e memorizam ilustrações de faces (GOOCH; REINHARD; GOOCH, 2004; WINNEMÖLLER; OLSEN; GOOCH, 2006).

Adicionalmente, abordagens de NPR foram avaliadas quanto à estética de imagens ao estilo de pontilhismo (MACIEJEWSKI et al., 2008), assim como quanto à estrutura e estética de vários estilos de renderizações de esboços e silhuetas (LI; MOULD; DAVIES, 2013). No entanto, Kyprianidis et al. (2013) afirmaram que nenhuma metodologia de avaliação e validação de imagens NPR consolidou-se como padrão na área.

Diante do exposto, i.e., a falta de uma metodologia consolidada para avaliação e validação de técnicas de NPR, notou-se ainda, durante a pesquisa realizada, uma escassez de métodos destinados à avaliação de abordagens de NPR especialmente relacionadas a vídeos renderizados de forma não fotorrealística, que são o foco desta

pesquisa. Ao investigar tal lacuna de pesquisa, recorreu-se a pesquisas em uma área correlata, a de padrões de testes de *codecs* de vídeos e também a pesquisas relacionadas a metodologias de avaliação da qualidade de imagens estáticas, nas quais foram realizadas comparações de métodos de filtragem de imagens. Ressalta-se, novamente, que a fundamentação teórica apresentada nesta seção serviu de base para a composição metodologia de avaliação e validação proposta e usada nesta pesquisa, a qual é apresentada no Capítulo 4.

Especificamente, a fundamentação para avaliação de vídeos apresentada a seguir, mesmo que não específica da área de NPR, conjuga-se com as metodologias de validação apresentadas anteriormente (específicas da área de NPR), formando assim o arcabouço no qual a metodologia de avaliação foi proposta.

Nesse contexto, Seshadrinathan et al. (2010) apresentaram um amplo estudo para avaliação de vídeos digitais. Ressaltam-se a seguir alguns aspectos relevantes da pesquisa realizada por Seshadrinathan et al. (2010), em especial no tocante aos materiais e métodos de pesquisa utilizados para avaliação subjetiva de vídeos, i.e., houve menor ênfase na discussão dos resultados da pesquisa realizada pelos autores.

O estudo realizado por Seshadrinathan et al. (2010) foi motivado pelo crescimento de aplicações que distorceram o vídeo de entrada, tais quais a aquisição de vídeo por sensores calibrados de forma inadequada, compressão de vídeo para transmissão, aprimoramento de vídeo degradado, cujos resultados devem ser reapresentados a um observador humano. Tal estudo evidenciou características dos métodos de avaliação, subjetivos e objetivos, para a verificação da qualidade dos vídeos reapresentados. Os autores afirmaram que a maioria das pesquisas em Computação Gráfica é validada informalmente, i.e., vários exemplos são incluídos em uma determinada publicação, passam pelo crivo e comparação com abordagens similares, ficando a cargo do leitor ou revisor a aceitação ou não das melhorias apresentadas pelos autores. Mesmo configurando um método efetivo, a diferença visual pode não oferecer subsídios suficientes para validar a pesquisa realizada, caso as melhorias apresentadas forem sutis, ou caso exemplos de imagens ou vídeos forem escolhidos de forma que as conclusões não possam ser generalizadas.

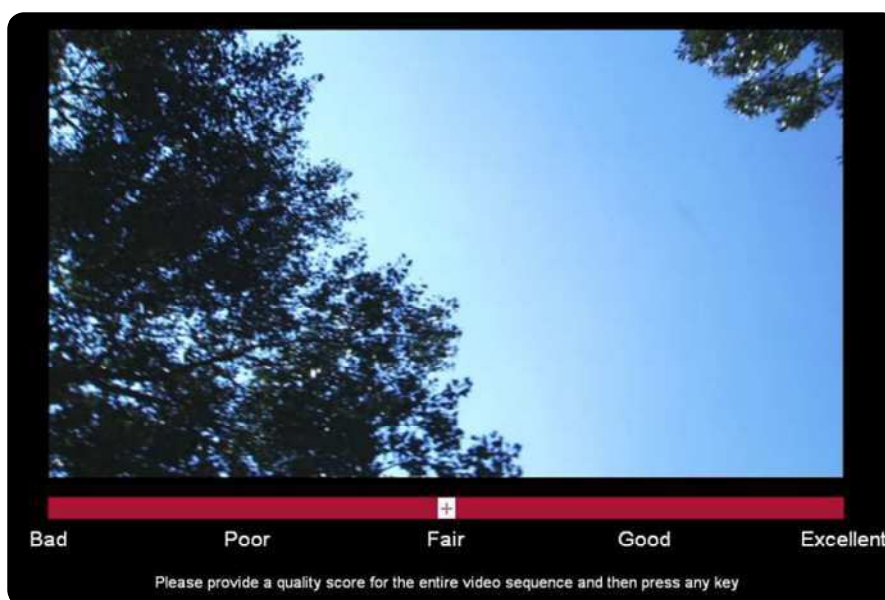
Motivados pelos fatores supramencionados, Seshadrinathan et al. (2010) afirmaram haver uma tendência crescente, nas pesquisas em computação gráfica, para a inclusão de comparações visuais dos resultados, com o uso de participantes humanos no processo de votação.

Os autores disponibilizaram, na forma de um banco de dados público, 160

vídeos, dos quais 10 são vídeos de referência e 150 são vídeos filtrados/distorcidos por quatro métodos diferentes (compressão MPEG-2, compressão H.264, sinal H.264 corrompido por transmissão em rede IP sujeita a erros, sinal H.264 corrompido por transmissão em rede sem fios sujeita a erros), que podem ser empregados em futuras pesquisas. Tal banco de dados público foi nomeado como *Laboratory for Image and Video Engineering (L.I.V.E.) Video Quality Database*.

Ao realizar os experimentos, os vídeos foram mostrados na resolução idêntica à aquisição, evitando distorções de escalonamento ou compressão. As áreas ao redor do vídeo foram mantidas na cor preta, a escala para avaliação foi mostrada na tela com o cursor centrado no meio da escala, evitando enviesar a resposta do participante. O participante poderia usar um tempo arbitrário para votar, mas não poderia modificar o voto. Um exemplo da interface utilizada pode ser visualizado na Figura 2.20.

Figura 2.20 – Excerto da interface computacional usada por Seshadrinathan et al. (2010) durante a coleta de dados a respeito da qualidade de vídeos.



A afirmação a qual o participante deveria referir-se ao realizar a marcação era: “Por favor, forneça uma classificação para a qualidade do vídeo e pressione qualquer tecla”. Antes da seção de votação, os participantes realizavam uma seção de testes para familiarização com a interface a ser utilizada na sessão de teste, na qual alguns vídeos de testes (excluídos do banco de dados público), com 10 segundos, eram avaliados, mas não incluídos na computação dos resultados estatísticos.

Seshadrinathan et al. (2010) ressaltaram ainda que a pesquisa e votação realizada com participantes humanos tornaram-se mais presente em pesquisas recentes, são mais difíceis de serem conduzidos, mais caras, demandam mais tempo, e podem induzir a erros ao serem generalizados, se a metodologia ou o teste for

conduzido de forma incorreta.

A metodologia empregada por Seshadrinathan et al. (2010) é em grande parte descrita pelo padrão ITU-R BT.500-12 (*Methodology for the Subjective Assessment of the Quality of Television Pictures*, desenvolvido desde 1974), que é uma metodologia para a avaliação subjetiva da qualidade de imagens de televisão. Em tal padrão, afirma-se que a avaliação subjetiva é empregada para que se estabeleça o desempenho de sistemas de televisão que adotam medidas que antecipam as reações dos espectadores. A caracterização de tal desempenho pode não ser possível a partir de medidas objetivas, sendo necessária a complementação com medições subjetivas.

No tocante a abordagens para avaliações subjetivas, o padrão ITU-R BT.500-12 (2012) estabelece duas categorias principais, i.e., a avaliação pode ser realizada com o uso de condições ótimas ou controladas (avaliação da qualidade) ou a avaliação pode ser realizada em ambientes não ótimos ou não controlados, nos quais a avaliação leva em consideração a manutenção da qualidade da imagem mesmo sob condições adversas, e.g., defeitos na transmissão e emissão do sinal de vídeo.

As recomendações feitas no padrão ITU-R BT.500-12 (2012) englobam vários métodos para a avaliação de estímulos visuais, dos quais se destacam: a escala de estímulo duplo de degradação (DSIS - *Double Stimulus Impairment Scale*); a escala contínua de estímulo duplo de degradação (DSCQS - *Double Stimulus Continuous Quality Scale*); e a comparação de estímulos por julgamento categórico adjetivo (SCACJ - *Stimulus Comparison Adjectival Categorical Judgement*). O padrão inclui descrições detalhadas a respeito do ambiente físico no qual os testes são realizados, metodologias de testes e indicações para processamento dos resultados.

No método DSIS, os vídeos são mostrados em pares, ao mesmo tempo, no qual o primeiro vídeo é de referência e o segundo é o vídeo degradado ou comprimido. Após a exibição, o participante, geralmente especialista em codificação de vídeo para televisão, compara o vídeo de referência com o vídeo degradado, adotando uma escala de Likert de 5 pontos, com as âncoras semânticas *muito irritante*, *irritante* e *imperceptível*.

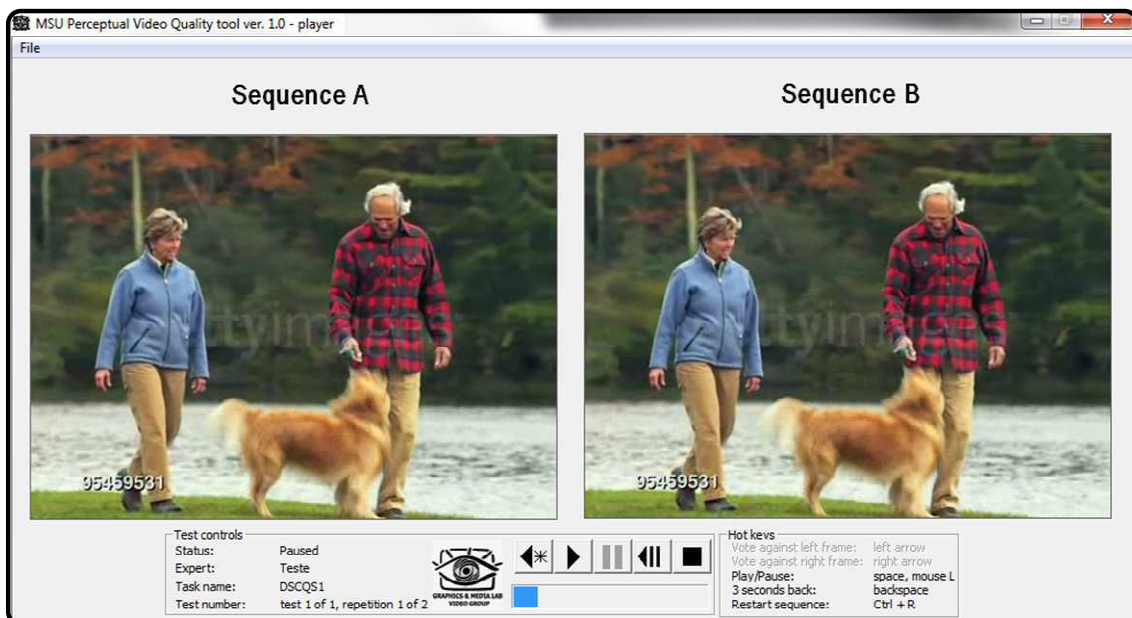
No método DSCQS, dividido em tipo 1 e tipo 2, os vídeos são exibidos lado a lado, podendo ser a exibição única ou repetida, de acordo com a conveniência do pesquisador. O participante do experimento pode alternar a execução de quaisquer dos dois vídeos exibidos, não sendo necessária a exibição dos vídeos ao mesmo tempo. Não há informação que identifique o vídeo de referência e após a exibição, o participante deve opinar sobre ambos os vídeos exibidos. No tipo 2, os vídeos são

mostrados lado a lado, mas de forma simultânea.

No método SCACJ, duas sequências são exibidas simultaneamente e após a exibição do vídeo o participante do experimento deve comparar a qualidade dos vídeos da seguinte maneira: o vídeo da esquerda é melhor, pior ou igual ao vídeo da direita.

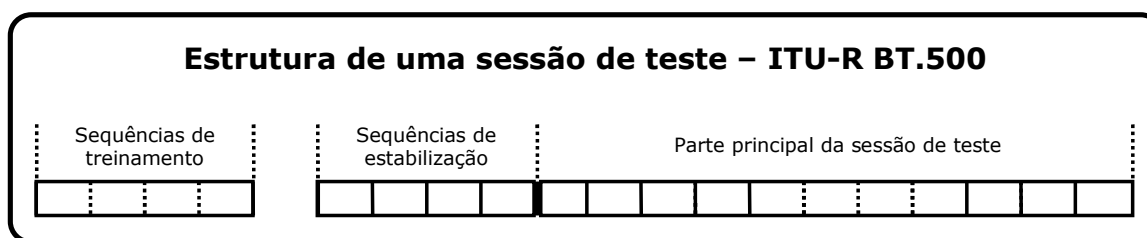
Uma implementação computacional dos três métodos sugeridos no padrão ITU-R BT.500-12 (2012), com a adição de três outros métodos, foi proposta pelo grupo MSU (GML, 2012), compondo um programa para a avaliação de percepção visual de vídeos digitais. Um excerto de tal implementação pode ser visualizado na Figura 2.21, quando usado o método DSCQS.

Figura 2.21 – Excerto da interface utilizada pelo grupo MSU para avaliação da percepção visual de vídeos.



No padrão ITU-R BT.500-12 (2012) também são ressaltadas condições gerais do ambiente no qual o experimento será realizado, tais quais o brilho e o contraste do monitor, a luminosidade da sala, a calibração correta de cores do monitor, o estabelecimento de uma ordem aleatória na apresentação dos vídeos, a alternância de períodos de descanso e atividade dos participantes, em intervalos de 30 minutos, para evitar a fadiga. Em especial, o padrão especifica que deve haver uma sessão de treinamento prévia, com um intervalo antes do início da sessão de testes. Adicionalmente, o padrão apregoa que alguns resultados obtidos apenas no início do experimento devem ser descartados, e que deve haver uma parada entre a sequência de treinamento e a sequência de testes, para que sejam respondidas eventuais perguntas dos participantes. Este fluxo de teste pode ser observado na Figura 2.22.

Figura 2.22 – Visão geral do fluxo de execução de uma sessão experimental.



De acordo com o padrão ITU-R BT.500-12 (2012), as etapas a serem seguidas durante um processo de avaliação subjetiva são: a escolha da sequência de vídeos para teste (*SRC - source*); a escolha das configurações dos sistemas que serão comparados (*HRC - hypothetical reference circuit*); a escolha da metodologia de teste dentre aquelas disponíveis, que influenciam a maneira pela qual os vídeos são apresentados e de qual maneira a opinião do especialista é coletada; a realização de testes com um número suficiente de especialistas (não se recomenda uma quantidade inferior a 15 participantes); calcular as médias de cada HRC, o que pode incluir análise estatística dos dados (com o uso de intervalos de confiança ou análise de variância).

Analisando abordagens afins àquela adotada em pesquisas na área de NPR e abordagens para a avaliação de imagens e vídeos de forma subjetiva, Mantiuk, Tomaszewska e Mantiuk (2012) compararam quatro métodos de avaliação de imagens, a partir de experimentos com humanos, relativos à classificação de algoritmos de computação gráfica, na tentativa de ajudar na confecção de testes de avaliação subjetiva. Tais testes, ao serem realizados de forma eficiente e rigorosa, devem também considerar alguns aspectos ao analisar os dados resultantes de tais experimentos, tais qual a escolha de intervalos de confiança ou escolha correta de testes estatísticos. A abordagem de comparação seguida pelos autores também foi fundamentada nas recomendações do padrão ITU-R BT.500-12 (2012).

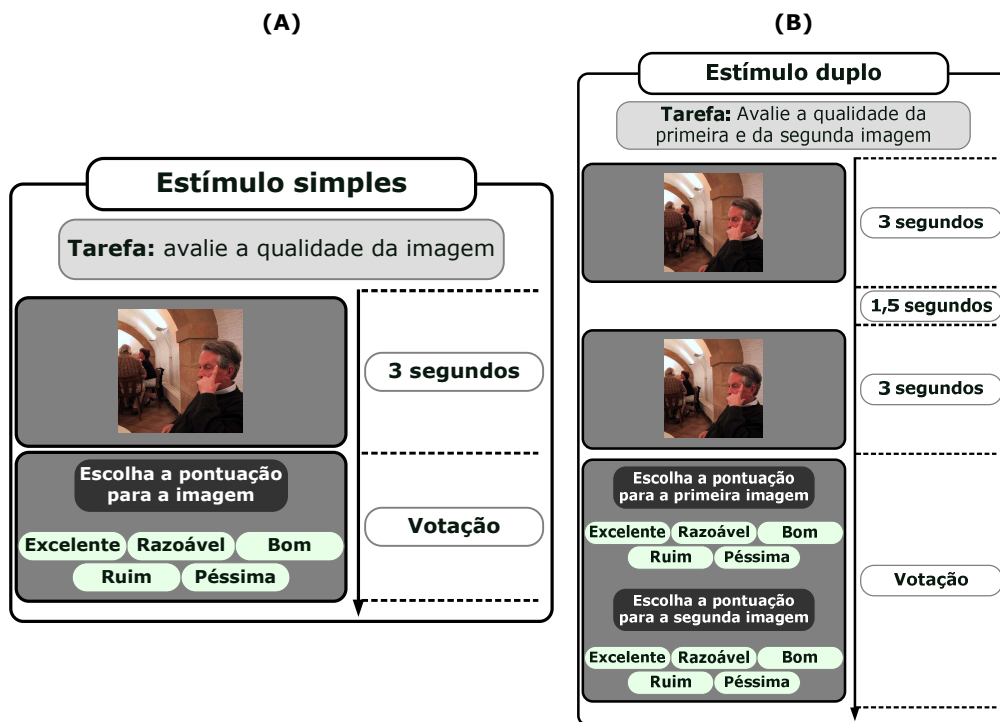
Os autores reiteraram que comparações entre um novo algoritmo de renderização em computação gráfica e métodos do estado da arte recorrem comumente a estudos com usuários, no qual um grupo de observadores classifica os resultados de vários algoritmos. Tais estudos com usuários, subjetivos, conhecidos também como avaliação subjetiva da qualidade de imagens (*subjective image quality assessment*), podem consumir muito tempo para serem concluídos e podem não garantir a obtenção de resultados conclusivos (!). Os métodos considerados por Mantiuk, Tomaszewska e Mantiuk (2012) foram:

Avaliação por estímulo *simples*: um exemplo de tal método pode ser visualizado

na Figura 2.23A. Este método é dividido em duas etapas, e.g., o participante: (i) visualiza uma imagem em um intervalo de tempo, tipicamente, 3 segundos; e (ii) pontua a qualidade da imagem a partir de uma escala numérica, comumente uma escala de Likert de 5 pontos, com âncoras de semântica diferencial (e.g., excelente, ruim, péssima).

Avaliação por estímulo *duplo*: um exemplo de tal método pode ser visualizado na Figura 2.23B. De forma similar à avaliação por estímulo simples, este método é dividido em três etapas, e.g., o participante: (i) visualiza uma imagem um intervalo de tempo, tipicamente, 3 segundos; (ii) visualiza uma tela na cor cinza, durante um intervalo de 1,5 segundos; (iii) visualiza uma segunda imagem, de forma similar a primeira; (iv) pontua a qualidade de ambas as imagens, a partir de duas escalas numéricas, similares àquelas consideradas na avaliação por estímulo simples, nas quais a qualidade das imagens deve ser auferida individualmente.

Figura 2.23 – Visão geral de dois métodos para avaliação subjetiva da qualidade de imagens: (A) Estímulo simples; e (B) Estímulo duplo.

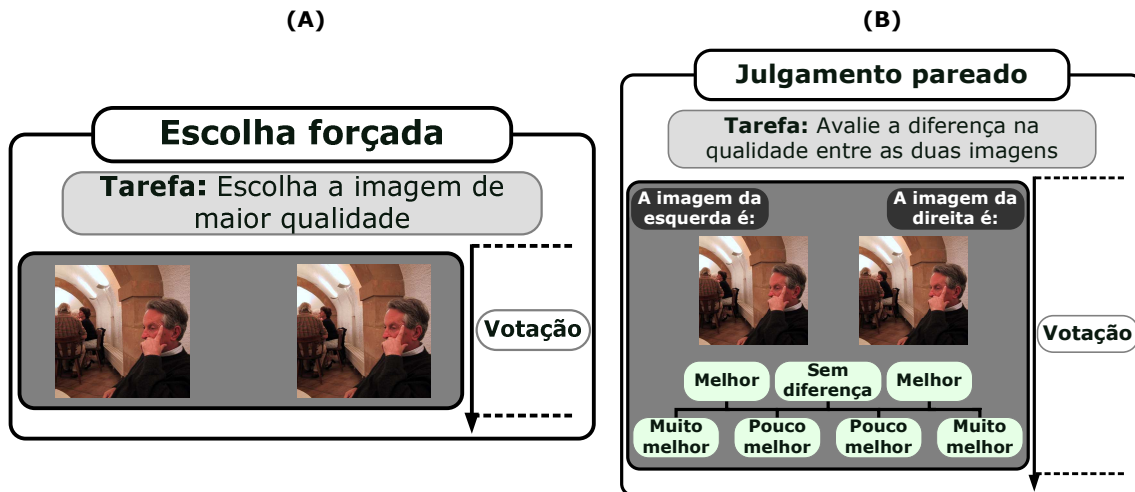


Avaliação por escolha *forçada*: um exemplo de tal método pode ser visualizado na Figura 2.24A. Neste método, o participante do experimento deve escolher deliberadamente, por comparação direta entre duas imagens, a imagem de maior qualidade.

Avaliação por *juízo pareado*: um exemplo de tal método pode ser visualizado na Figura 2.24B. Neste método, duas imagens são apresentadas ao participante. Também é apresentada ao participante uma única escala, tipicamente

uma escala de Likert de 7 pontos, na qual a comparação, direta e forçada, deve ser realizada com a atribuição de um valor na escala, a fim de diferenciar a qualidade das imagens apresentadas, e.g., a qualidade da imagem da esquerda é melhor, pior ou igual à imagem da direita.

Figura 2.24 – Visão geral de dois métodos para avaliação subjetiva da qualidade de imagens: (A) Escolha forçada; e (B) Julgamento pareado.



Fatores adicionais podem influenciar o processo de avaliação na área de NPR, quando são considerados participantes humanos: *número de participantes*, que deve ser uma amostra representativa, na qual possam ser inferidos padrões de comportamento ou perfis adequados às inferências das hipóteses de pesquisa investigadas; *conhecimendo da área específica*, tais quais artistas ou desenhistas; *quantidade adequada de imagens ou vídeos utilizados no experimento*, que deve realçar um balanceamento entre quantidade apresentada, tempo para realização do experimento, representatividade do conteúdo das imagens ou vídeos apresentados.

É importante ressaltar que métricas objetivas para avaliação podem melhorar a fiabilidade e padronização dos processos de avaliação, configurando-se menos onerosas. Entretanto, algumas características, tais como a preferência visual direta de participantes humanos, aspecto visual, cintilação visível, podem escapar ao enfoque de avaliação objetiva.

Mantiuk, Tomaszewska e Mantiuk (2012), ao relatar a comparação das quatro modalidades de métodos para a avaliação subjetiva da qualidade de imagens, ressaltaram que a escolha forçada afigurou-se a mais adequada, mais acurada e que demandou menos tempo para a conclusão das sessões de teste. Tal modalidade deve ser integrada à escolha pseudoaleatória das imagens a serem comparadas, visando à redução das comparações duplicadas entre participantes. Os autores ressaltaram que, apesar de tais elucidações, o método indicado e as discussões relacionadas são válidos

para os procedimentos experimentais, a amostra, imagens e algoritmos comparados no experimento estudado, e podem não se generalizar para quaisquer tipos de experimento. Os autores ressaltaram ainda que a comparação simples de médias pode não fornecer evidências suficientes para ressaltar diferenças entre os métodos considerados em determinado experimento. Para que tais diferenças aflorem, é comumente necessário um grande número de repetições (entre 30-60). Detalhes adicionais das formas de comparação podem ser encontrados na pesquisa de Mantiuk, Tomaszewska e Mantiuk (2012).

Finalmente, é importante frisar que a área de renderização não fotorrealística pode englobar renderizações artísticas e novos métodos para ilustração, ou ainda visualização de forma diferenciada de uma grande quantidade de dados. Conforme apresentado anteriormente, tal área desenvolveu-se nas últimas décadas, confluindo em questionamentos a respeito da validade, utilidade, adequação e aceitação da grande variedade de técnicas criadas ao longo do tempo (GATZIDIS et al., 2008). Adicionalmente, Collomosse e Kiprianidis (2011) afirmaram que ainda não há um consenso a respeito da metodologia de avaliação a ser adotada, tanto pela baixa quantidade de metodologias propostas, quanto pela abrangência restrita das metodologias existentes.

O escopo de investigação de métodos para a avaliação de técnicas de NPR englobou avaliação de imagens bidimensionais, quadros de vídeos, métodos para a avaliação de cenas tridimensionais, novos métodos para a avaliação de representações ilustrativas (*illustrative visualization*) e método para a avaliação de imagens de televisão. Especificamente, os resultados do processo de sondagem devem ser analisados, à luz das restrições ressaltadas por Hertzmann (2010), e.g., a avaliação experimental é função da amostra considerada.

Assim sendo, a pesquisa ora realizada em abstração de vídeos digitais englobou um processo de avaliação experimental, envolvendo a sondagem da opinião de universo amostral de observadores humanos a respeito da qualidade da renderização gerada após a utilização da abordagem proposta. O ponto de partida para tal avaliação foi a pesquisa realizada nos artigos de avaliação na área de NPR.

Na Seção 2.2, ressaltou-se a existência de abordagens destinadas à abstração de imagens e vídeos que se fundamentam em técnicas de suavização com preservação de bordas. Observou-se, adicionalmente, a necessidade de um estudo mais aprofundado de tais processos de filtragem, em especial sua aplicação à renderização não fotorrealística. Tal estudo é apresentado na seção seguinte.

2.4 Suavização para a Renderização Não Fotorrealística

A filtragem, enquanto operação fundamental na área de Processamento de Imagens e Visão Computacional (TOMASI; MANDUCHI, 1998), destina-se, dentre outras aplicações, à suavização de regiões, à extração de bordas ou como processo-chave para a extração de características presentes em uma imagem. A filtragem pode ser classificada com relação ao propósito (suavização, extração de bordas) ou ao domínio a qual pertence (domínio da frequência ou domínio espacial).

Com o intuito de remover ruído em imagens fotorrealísticas, têm sido propostos na literatura diversos filtros de suavização, muitos dos quais com preservação de bordas, cuja aplicação ocorre normalmente de forma iterativa, até a obtenção dos resultados desejados.

Uma variedade de abordagens para a abstração de imagens e vídeos envolvendo o emprego de filtros têm sido propostas (WINNEMÖLLER; OLSEN; GOOCH, 2006; KYPRIANIDIS; KANG; DÖLLNER, 2009; KYPRIANIDIS, 2011), muitas das quais fundamentam-se em técnicas de suavização com preservação de bordas (PERONA; MALIK, 1990; SAINT-MARC; CHEN; MEDIONI, 1991; RAMPONI, 1995). Dentre as abordagens para a abstração de imagens por suavização, com preservação de bordas, destacaram-se, durante as pesquisas, a *difusão anisotrópica* (PERONA, MALIK, 1990) e o *filtro bilateral* (TOMASI; MANDUCHI, 1998).

Tais métodos (TOMASI; MANDUCHI, 1998; COMANICIU; MEER, 2002), enquanto iterativos, podem apresentar problemas relacionados à eficiência computacional e ao critério de parada das iterações, o qual é usualmente formulado heurísticamente pelo pesquisador.

Zhao, Jin e Mao (2011) ressaltaram que as técnicas que empregam o filtro Kuwahara possuem limitações quanto à estilização direcional de elementos da imagem, nas quais há abstração das cores, mas não há abstração nas formas, curvas e contornos da imagem.

Com base nessas observações, assim como em resultados documentados na literatura (e.g., WINNEMÖLLER; OLSEN; GOOCH, 2006; KYPRIANIDIS; KANG; DÖLLNER, 2009; KANG; LEE; CHUI, 2009; KYPRIANIDIS, 2011), ressalta-se a importância da suavização com preservação de bordas como um tópico de pesquisa na área da abstração de imagens.

Diante da importância relatada anteriormente, apresenta-se, a seguir, uma breve fundamentação teórica acerca do processo de suavização por difusão anisotrópica, a qual constitui um elemento importante para o entendimento das

alterações propostas o âmbito desta pesquisa e relatadas neste documento.

De acordo com Kim (2004), a percepção visual humana de objetos do mundo real é realizada em diferentes escalas, as quais podem ser classificadas como grosseiras (*coarse-scale*) ou refinadas (*fine-scale*), tendo sido esta noção de multi-escala formalizada por Witkin (1983).

Kim (2004) afirma que a teoria do espaço de escala Gaussiano é empregada para formalizar o processamento de imagens em diferentes resoluções (escalas), cada uma das quais é obtida a partir da convolução da imagem com uma máscara Gaussiana com σ crescente. O autor afirma que a convolução de uma imagem com uma máscara Gaussiana ocorre de forma *isotrópica*¹⁸, i.e., a convolução pode ser encarada como solução para a equação diferencial que descreve a condução de calor. Tal fato indica que, no processamento de imagens em multi-escala no espaço de escala Gaussiano, haverá a distribuição do valor de cada pixel em todas as direções, ao serem obtidas imagens em escalas mais grosseiras.

Assim sendo, o espaço de escala Gaussiano também pode ser denominado espaço de escala linear ou espaço de escala isotrópico. A teoria do espaço de escala foi elaborada para sinais unidimensionais, tendo sido, posteriormente, estendida para sinais bidimensionais, i.e., imagens (KIM, 2004, *apud* VELHO; TEIRA; GONES, 2000).

O espaço de escala para sinais bidimensionais tem como base a distribuição normal bidimensional, $\mathbf{N}(\mathbf{x}_0, \mathbf{y}_0, \sigma)$, na qual $(\mathbf{x}_0, \mathbf{y}_0)$ representa a média μ e σ representa o desvio-padrão. Tal distribuição é definida pela função Gaussiana \mathbf{g} :

$$\mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{x}_0, \mathbf{y}_0, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(\mathbf{x}-\mathbf{x}_0)^2 - (\mathbf{y}-\mathbf{y}_0)^2}{2\sigma^2}\right]. \quad (2.4.1)$$

Adotando $\sigma^2=2\mathbf{t}$ e $\mu=\mathbf{0}$ a fim de se obter a notação:

$$\mathbf{G}_t(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi t} \exp\left[-\frac{\mathbf{x}^2 + \mathbf{y}^2}{4t}\right]. \quad (2.4.2)$$

Considerando $\mathbf{I}(\mathbf{x}, \mathbf{y}) : \mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ uma imagem, o espaço de escala é uma imagem 3D, $\mathbf{I}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$, a qual satisfaz a equação diferencial parcial:

$$\frac{\partial \mathbf{I}_t(\mathbf{x}, \mathbf{y})}{\partial t} = \nabla^2 \mathbf{I}_t(\mathbf{x}, \mathbf{y}) = \frac{\partial^2 \mathbf{I}_t(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}^2} + \frac{\partial^2 \mathbf{I}_t(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}^2}. \quad (2.4.3)$$

A imagem original $\mathbf{I}(\mathbf{x}, \mathbf{y}) = \mathbf{I}_0(\mathbf{x}, \mathbf{y})$ representa a condição inicial da equação. A

¹⁸ Que apresenta as mesmas propriedades em todas as direções (HOUAISS, 2001).

variável t especifica a escala da imagem correspondente. A solução da equação diferencial parcial Eq. 2.4.3 pode ser expressa como a convolução da imagem com uma função Gaussiana bidimensional (vide Eq. 2.4.2):

$$I_t(\mathbf{x}, \mathbf{y}) = G_t(\mathbf{x}, \mathbf{y}) * I(\mathbf{x}, \mathbf{y}). \quad (2.4.4)$$

A convolução apresentada na Eq. 2.4.4 pode ser separada em duas convoluções Gaussianas unidimensionais:

$$I_t(\mathbf{x}, \mathbf{y}) = G_t(\mathbf{y}) *_{(y)} \left(G_t(\mathbf{x}) *_{(x)} I(\mathbf{x}, \mathbf{y}) \right). \quad (2.4.5)$$

Tal separação permite o cálculo da convolução de forma separada para imagens, mantendo a linearidade e a invariância à translação. Para o cálculo da convolução, faz-se necessária a discretização do sinal bidimensional, sendo possível e vantajoso o cálculo prévio do espaço de escala para algumas escalas fixas, a exemplo do que é feito na biblioteca OpenCV (2006), para máscaras Gaussianas de dimensões 3x3, 5x5 e 7x7. O espaço de escala Gaussiano possui ainda as propriedades de invariância à rotação e invariância à translação. Um efeito observado após esse processamento é o deslocamento espacial progressivo das bordas e a obtenção de imagens com um aspecto cada vez mais borrado.

Kim (2004) afirmou que pesquisas subsequentes a Witkin (1983) levaram a modificações no espaço de escala linear anteriormente descrito. Perona e Malik (1990) propuseram o espaço de escala não linear, definindo assim o processo de suavização por difusão anisotrópica. Tal processo simplifica a imagem mantendo a localização espacial e a nitidez das bordas, devendo ser aplicado de forma iterativa, a fim de que sejam obtidos os resultados desejados.

O espaço de escala não linear foi obtido por Perona e Malik (1990), a partir da modificação da Eq. 2.4.3:

$$\frac{\partial I_t(\mathbf{x}, \mathbf{y})}{\partial t} = \nabla \cdot \left[\psi \left(\|\nabla I_t(\mathbf{x}, \mathbf{y})\| \right) \nabla I_t(\mathbf{x}, \mathbf{y}) \right] \quad (2.4.6)$$

na qual $\|\nabla I_t(\mathbf{x}, \mathbf{y})\|$ representa a magnitude do gradiente da I_t , ψ é denominada de *função de parada-na-aresta (edge stopping function)*, dependente da magnitude do gradiente. A função de parada-na-aresta é responsável pelo controle da intensidade da difusão em cada pixel da imagem original. Ressaltam-se três fatores que afetam significativamente os resultados obtidos por tal processo: (i) a função de parada-na-aresta escolhida; (ii) a velocidade da difusão; e (iii) o número de iterações aplicadas.

Perona e Malik (1990) discretizaram (espaçotemporalmente) a Eq. 2.4.6:

$$\mathbf{I}(\mathbf{s}, \mathbf{t} + 1) = \mathbf{I}(\mathbf{s}, \mathbf{t}) + \frac{\lambda}{|\eta_{\mathbf{s}}|} \sum_{\mathbf{p} \in \eta_{\mathbf{s}}} \psi(|\nabla \mathbf{I}_{\mathbf{s}, \mathbf{p}}(\mathbf{t})|) \nabla \mathbf{I}_{\mathbf{s}, \mathbf{p}}(\mathbf{t}), \quad (2.4.7)$$

na qual $\mathbf{I}(\mathbf{s}, \mathbf{t})$ representa uma imagem, \mathbf{s} representa a posição de um pixel, \mathbf{t} denota o número de iterações, $\mathbf{t} \geq \mathbf{0}$, $\lambda \in \mathbb{R}^+$ é uma constante que determina a velocidade da difusão (normalmente $\lambda=1$), $\eta_{\mathbf{s}}$ representa o conjunto de vizinhos espaciais do voxel \mathbf{s} , no qual são considerados quatro pixels vizinhos: norte, sul, leste e oeste, $\nabla \mathbf{I}_{\mathbf{s}, \mathbf{p}}(\mathbf{t})$ representa a magnitude do gradiente da imagem \mathbf{I} no ponto \mathbf{s} na direção (\mathbf{s}, \mathbf{p}) na iteração \mathbf{t} , i.e:

$$\nabla \mathbf{I}_{\mathbf{s}, \mathbf{p}}(\mathbf{t}) = \mathbf{I}(\mathbf{p}, \mathbf{t}) - \mathbf{I}(\mathbf{s}, \mathbf{t}), \quad \mathbf{p} \in \eta_{\mathbf{s}}. \quad (2.4.8)$$

No processo de suavização por difusão anisotrópica, em cada pixel de $\mathbf{I}(\mathbf{s}, \mathbf{t})$, a função ψ da Eq. 2.4.7 é avaliada para cada vizinho, em $\eta_{\mathbf{s}}$. Se a magnitude do gradiente $\nabla \mathbf{I}_{\mathbf{s}, \mathbf{p}}(\mathbf{t})$ for pequena, denotando uma área homogênea da imagem, a função ψ crescerá linearmente em relação à diferença dos vizinhos em $\eta_{\mathbf{s}}$, suavizando o pixel na posição \mathbf{s} . Caso contrário, o que denotará a presença de uma transição, a função ψ tenderá a zero, de modo que o valor do pixel na posição \mathbf{s} será pouco alterado.

Percebe-se a importância do cálculo da magnitude do gradiente no processo de suavização por difusão anisotrópica, o qual foi modificado na presente pesquisa. As alterações realizadas estão apresentadas no Capítulo 3.

Perona e Malik (1990) sugeriram a utilização de uma das duas funções de parada-na-aresta (ψ), apresentadas nas Eqs. 2.4.9 e 2.4.10:

$$\psi_1(\mathbf{x}) = \frac{1}{1 + \frac{\mathbf{x}^2}{2\sigma^2}} \text{ ou} \quad (2.4.9)$$

$$\psi_2(\mathbf{x}) = \exp\left[\frac{-\mathbf{x}^2}{2\sigma^2}\right] \quad (2.4.10)$$

Kim (2004) relata que Black et al. (1998) propuseram o processo de suavização por difusão anisotrópica robusta (*RAD - Robust Anisotropic Diffusion*). Os autores estabeleceram a relação entre o processo de suavização por difusão anisotrópica e a Estatística Robusta. A Estatística Robusta objetiva a atenuação do efeito de pontos fora da curva (*outliers*) no cálculo de medidas de tendência central ou dispersão (MARONNA; MARTIN; YOHAI, 2006). Adicionalmente, métodos de Estatística Robusta procuram preservar a forma, a dispersão e simetria dos dados.

Black et al. (1998) descreveram a relação entre a Estatística Robusta e o processo de suavização por difusão anisotrópica como sendo a estimativa da imagem original a partir de uma imagem corrompida com ruído (a qual contém *outliers*). Black et al. (1998) propuseram uma nova função de parada-na-aresta que produz melhores resultados do que aqueles obtidos por Perona e Malik (1990).

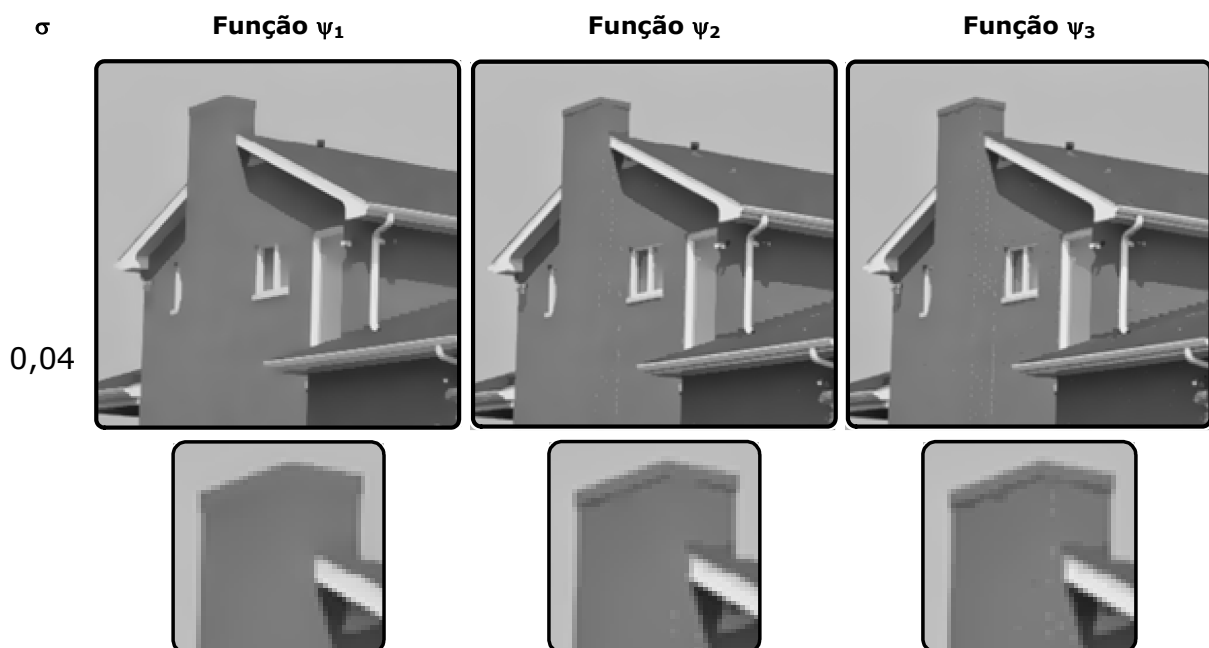
De acordo com a teoria da Estatística Robusta, Black et al. (1998) escolheram a *função de Tukey (Tukey's biweight function)* como função parada-na-aresta:

$$\psi_3(\mathbf{x}) = \begin{cases} \mathbf{x} \left[1 - \frac{\mathbf{x}^2}{\sigma^2} \right]^2, & |\mathbf{x}| \leq \sigma; \\ \mathbf{0}, & \text{caso contrário.} \end{cases} \quad (2.4.11)$$

Na Estatística Robusta, um *estimador robusto* considera o valor de apenas um dado referente a uma observação insuficiente para mudar significativamente o valor a ser estimado, i.e., um *outlier* tem pouca influência no valor a ser estimado. No contexto da suavização de imagens, a suavização anisotrópica robusta suaviza regiões homogêneas, ao mesmo tempo em que preserva as arestas (*outliers*).

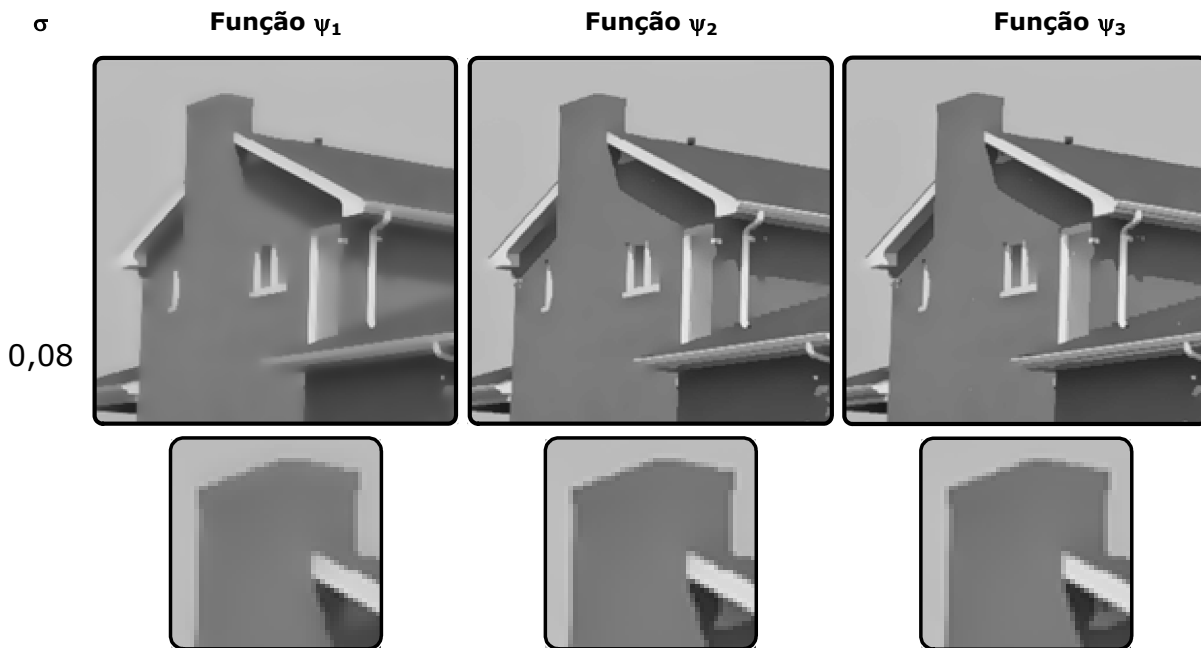
As imagens da Figura 2.25 foram obtidas após a aplicação de 50 iterações do algoritmo de difusão. As funções ψ_1 e ψ_2 são aquelas propostas por Perona e Malik (1990), enquanto a função ψ_3 é aquela proposta por Black et al. (1998). Nas Figura 2.25 e 2.26 pode-se constatar a superioridade da função parada-na-aresta proposta por Black et al. (1998) em relação àquelas propostas por Perona e Malik (1990).

Figura 2.25 – Difusão anisotrópica com diferentes funções de parada-na-aresta.



Fonte: Kim (2004)

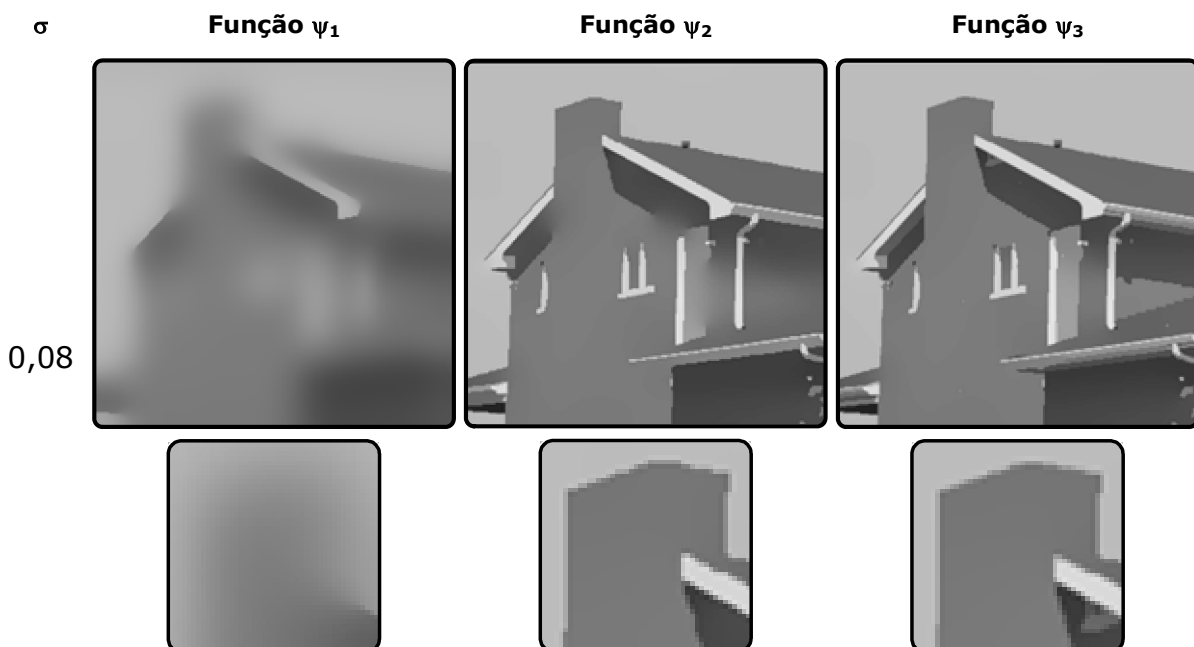
Figura 2.26 – Difusão anisotrópica com diferentes funções de parada-na-aresta.



Fonte: Kim (2004)

O comportamento da difusão anisotrópica com grande número de iterações (500 iterações) pode ser visualizado na Figura 2.27. A função de Tukey ψ_3 consegue manter as arestas mais nítidas, as funções de parada-na-aresta propostas por Perona-Malik ψ_1 e ψ_2 acabam borrando as arestas.

Figura 2.27 – Difusão anisotrópica após grande número de iterações.

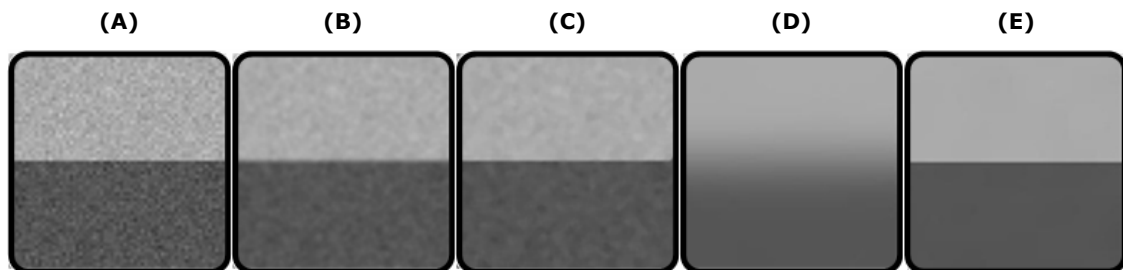


Fonte: Kim (2004)

O comportamento da suavização por difusão anisotrópica em relação a outras técnicas de suavização, pode ser visualizado na Figura 2.28, a qual reflete o comportamento visual de técnicas diferentes de suavização, tendo como entrada uma

imagem sintética. São apresentadas: uma imagem filtrada por um processo de suavização isotrópica, uma imagem filtrada espacialmente com preservação de bordas a partir da aplicação do filtro da mediana e imagens filtradas pelo processo de suavização por difusão anisotrópica.

Figura 2.28 – Comparação visual de técnicas de suavização: (A) Original; (B) Suavização Gaussiana; (C) filtro da mediana 5x5; (D) Perona-Malik; e (E) RAD.



A superioridade na preservação das bordas pode ser observada na Figura 2.28E, quando RAD é usada. A seguir, são apresentadas as considerações finais deste capítulo.

2.5 Considerações Finais do Capítulo

Diante da revisão das abordagens destinadas à conversão de vídeos reais em vídeos renderizados não fotorrealisticamente, realizada neste capítulo, assim com dos resultados produzidos por algumas técnicas presentes nessas abordagens, constata-se a possibilidade de abstração do conteúdo de imagens ou vídeos digitais por meio de técnicas de segmentação ou de suavização.

Inicialmente, foram apresentados aspectos gerais da renderização não fotorrealística de vídeos, em especial as iniciativas seminais de NPR, que focaram na imitação e reprodução de estilos de pintura para imagens estáticas, concentrando-se apenas na reprodução algorítmica do ato de desenhar e distribuir pinceladas artísticas na imagem resultante. Posteriormente, foram relatadas, de forma cronológica, as principais abordagens associadas à renderização não fotorrealística de vídeos.

Além disso, percebe-se que as técnicas automáticas revisadas têm como base filtros de suavização. Não obstante, ressalta-se que técnicas que utilizam segmentação para abstração produzem melhores resultados quando há a intervenção do usuário.

A presente pesquisa focalizou, inicialmente, nos estudos a respeito de técnicas de segmentação automática de vídeos, com ajustes para prover coerência temporal intraquadros do vídeo de entrada. Embora os resultados obtidos a partir dos algoritmos de segmentação tenham sido satisfatórios para uma gama variada de vídeos, a tarefa de integrar as regiões segmentadas e rastreá-las a partir das

informações de movimento afigurou-se, ao mesmo tempo, desafiadora e frustrante. Os desafios encontrados relacionaram-se à apresentação de cintilação ao longo do vídeo de saída e à supersegmentação de regiões dos quadros de vídeo. Tais estudos, relativos à proposição de uma abordagem baseada em segmentação espaçotemporal de vídeos, são apresentados no Apêndice A.

Por último, a pesquisa focalizou estudos relativos de filtros de suavização com preservação de bordas, visando à proposição de uma abordagem que permitisse a geração automatizada de vídeos estilizados de forma não fotorrealística, a qual é descrita no próximo capítulo.

No Capítulo 4 é apresentada uma metodologia destinada à avaliação de renderização não fotorrealística de vídeos digitais, a partir da qual a avaliação de renderizações é realizada com investigações experimentais envolvendo usuários. A proposição desta metodologia fundamentou-se na pesquisa relatada na Seção 2.3 do presente capítulo.

No Capítulo 5, são avaliadas alterações propostas para a função de parada-na-aresta apresentadas neste capítulo, bem como o uso de uma função alternativa, inspirada também na Estatística Robusta. Tais experimentos foram realizados com o intuito de observar o comportamento de diferentes funções de parada-na-aresta para o propósito da renderização não fotorrealística.

Capítulo 3

Abordagem Proposta

Conforme exposto no Capítulo 2, foram identificadas, no processo de revisão bibliográfica, duas categorias de abordagens para a conversão de vídeos em vídeos renderizados não fotorrealisticamente. A primeira delas fundamenta-se na segmentação espaçotemporal em regiões de quadros de vídeo de entrada, enquanto a segunda fundamenta-se na aplicação de filtros de suavização em quadros de vídeo de entrada, com preservação das bordas entre as regiões que compõem os quadros.

A apresentação de ambas as abordagens justificou-se pelo encaminhamento inicial da pesquisa e de seu posterior redirecionamento. Inicialmente, uma abordagem pertencente à primeira categoria supramencionada foi investigada. Embora os resultados preliminares sugerissem que tal abordagem seria promissora para atender aos objetivos de pesquisa, limitações posteriormente encontradas na resolução de problemas associados às diferentes etapas da abordagem implicaram a tomada de decisão de redirecionamento da pesquisa para outra abordagem, pertencente à segunda categoria.

A abordagem fundamentada na segmentação espaço-temporal de vídeos que foi inicialmente investigada, composta por quatro etapas, é detalhada no Apêndice A, o qual também contém os resultados mais representativos obtidos e uma discussão dos problemas encontrados.

Assim sendo, em função da análise dos resultados das referidas estratégias, é proposta e descrita uma abordagem original para a conversão de vídeos em vídeos renderizados não fotorrealisticamente ao estilo de cartuns. Este capítulo foi estruturado de modo a apresentar tal abordagem, conforme proposto nos objetivos descritos no Capítulo 1.

Três aspectos principais perfazem as contribuições originais desta pesquisa: (i) a suavização de quadros do vídeo de entrada, com preservação de bordas, por meio de um processo de suavização adaptativa por difusão anisotrópica modificada; (ii) a

aplicação de um detector de bordas neural em conjunção ao processo de suavização anisotrópica; e (iii) a combinação original dos seis módulos que compõem a abordagem proposta.

O restante deste capítulo está estruturado como segue:

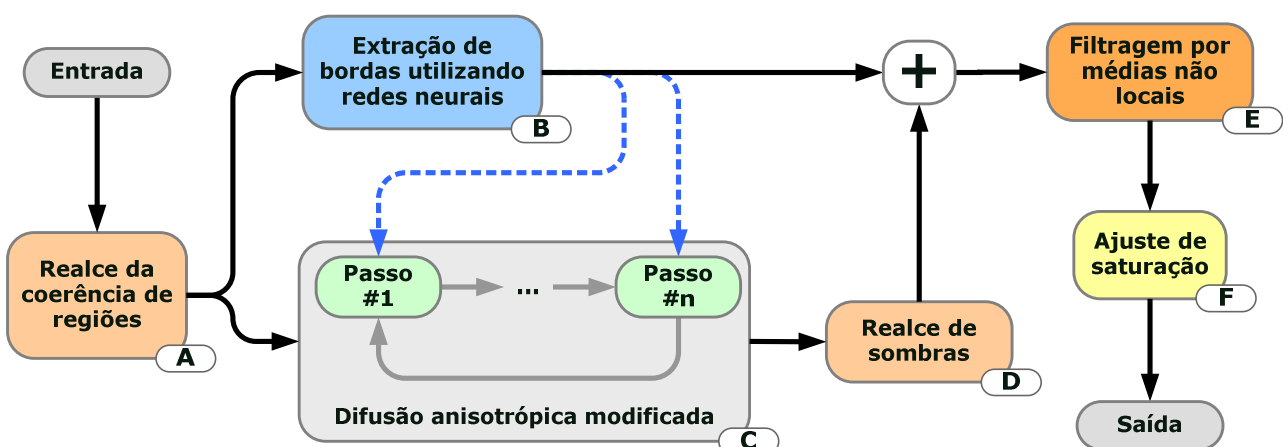
- Seção 3.1: são apresentados os detalhes dos módulos que compõem a abordagem proposta.
- Seção 3.2: são apresentadas considerações a respeito da implementação computacional da abordagem proposta, concernentes a: (i) limitações da abordagem proposta; e (ii) à comparação visual de renderizações de imagens corrompidas com ruído sintético.

3.1 Abordagem Fundamentada na Suavização por Difusão Anisotrópica

A abordagem proposta foi inspirada em características de outras abordagens concorrentes, apresentadas no Capítulo 2, em especial na generalização de um arcabouço para a geração de imagens renderizadas ao estilo de cartuns, proposta por Kyprianidis (2013) (ver Figura 2.15, p. 57). Outras fontes de inspiração para a composição desta abordagem foram os arcabouços propostos por Winnemöller, Olsen e Gooch (2006), Kyprianidis e Döllner (2008), Kang, Lee e Chui (2007) e Kyprianidis (2011; 2013).

Na Figura 3.1, é apresentada uma síntese gráfica da abordagem fundamentada na suavização de quadros por difusão anisotrópica para a renderização não fotorrealística de vídeos.

Figura 3.1 – Abordagem proposta, fundamentada na suavização por difusão anisotrópica.



Os seis módulos que compõem a abordagem proposta são descritos e discutidos a seguir.

Módulo A: a imagem de entrada é submetida a uma etapa de *pré-processamento*, com o intuito de atenuar ruído, realçar bordas, enfatizar características direcionais e adicionar um efeito artístico à imagem de entrada.

Módulo B: após a etapa de pré-processamento, a imagem resultante é submetida a um processo de *filtragem espacial*, a partir de uma máscara (*kernel*) de dimensões 5x5. Tal processo é realizado por um classificador neural tipo *Perceptron Multicamadas*, treinado para a detecção de bordas. A inclusão deste módulo à abordagem proposta justificou-se por este conferir um aspecto visual diferenciado aos resultados obtidos, em relação a abordagens concorrentes, pela presença de um módulo destinado à detecção de bordas no arcabouço generalizado por Kyprianidis (2013) e por visar à orientação da difusão anisotrópica modificada. O resultado deste módulo é um mapa de bordas.

Módulo C: a imagem de saída do módulo A é submetida a uma filtragem por um processo adaptativo de difusão anisotrópica. Tal processo de filtragem é realizado em conjunção com o mapa de bordas resultante do módulo B, o qual não é submetido ao processo de difusão anisotrópica, apenas a imagem resultante do módulo A o é. Este módulo é o cerne do processo de renderização não fotorrealística fundamentada na suavização por difusão anisotrópica. As modificações propostas no tocante ao processo de filtragem visaram a evitar a suavização dos pixels que se encontram na região de transição, ao mesmo tempo em que são suavizadas as regiões de cor similar que compõem a imagem de entrada.

Módulo D: após o processo de difusão anisotrópica, as regiões de sombra da imagem são realçadas e as bordas resultantes do módulo B são sobrepostas à imagem suavizada. A presença deste módulo visou ao aumento do realce de detalhes de escala fina da imagem, à suavização adicional da imagem resultante, visando a produzir resultados com maior abstração e ao incremento da saturação da imagem de saída.

Módulo E: a imagem combinada dos módulos B e C é filtrada, a partir da aplicação da técnica para a computação de médias não locais (*Non-Local Means - NLM*), a qual reduz o ruído e homogeneiza as regiões de cor da imagem. A presença do módulo de suavização após a sobreposição das bordas (Módulo B + Módulo D) justificou-se pelo aparecimento de artefatos espúrios, cintilação e ruído decorrente da sobreposição das bordas na imagem resultante.

Módulo F: a imagem filtrada resultante do módulo anterior é submetida a uma etapa de pós-processamento, a partir da qual a imagem de saída do módulo E é convertida para o espaço de cor HSV e a saturação (canal S) é aumentada, com o intuito de conferir um aspecto de cartum à imagem de saída. A saída desta etapa é a representação não fotorrealística da imagem de entrada.

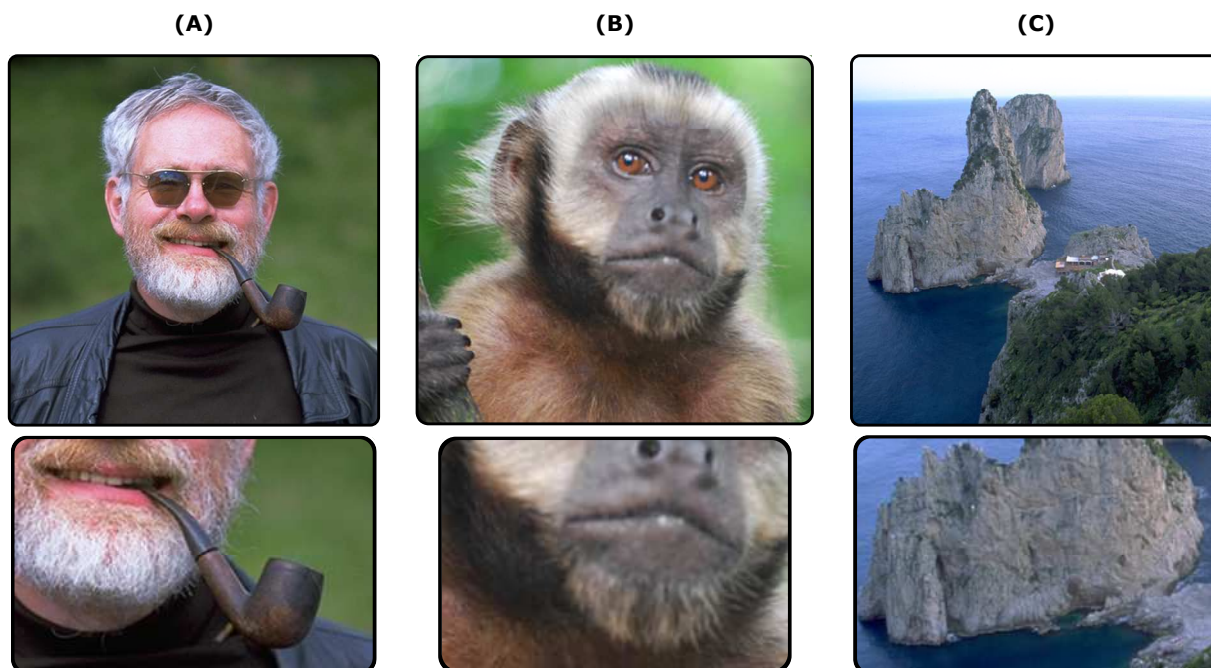
Os módulos B e C compõem o núcleo da abordagem proposta, os quais são respectivamente destinados à: (i) extração de bordas dos quadros do vídeo; e (ii) suavização por difusão anisotrópica dos quadros do vídeo. A combinação dos seis módulos representou o eixo das atividades empreendidas durante a pesquisa realizada, constituindo um dos aspectos originais relevantes da pesquisa ora apresentada.

Em especial, destacam-se como contribuições originais a combinação dos resultados do Módulo B (extração de bordas) e C (suavização por difusão anisotrópica), a alteração da função de parada-na-aresta do algoritmo de suavização anisotrópica proposto por Perona e Malik (1990), a adição de um módulo para o realce de sombras, o uso da filtragem NLM, objetivando a renderização não fotorrealística e o aumento da saturação da imagem com o propósito de conferir aspecto de cartum à imagem de saída. É importante ressaltar que a imagem gerada pelo Módulo B é combinada/ponderada à imagem suavizada no Módulo C a cada iteração do processo de suavização por difusão anisotrópica. Tal ponderação está realçada na Figura 3.1 pela linha azul pontilhada, e os detalhes concernentes a tal ponderação são apresentados na descrição do módulo C.

Nos próximos parágrafos, a apresentação dos detalhes que caracterizam cada módulo contém: (i) uma breve explanação do fundamento e do princípio de funcionamento da técnica; (ii) o resultado esperado a partir da aplicação da operação; e (iii) imagens que ilustram o resultado após o processamento das imagens apresentadas pelo módulo destacado.

Os resultados apresentados nesta seção advieram da inspeção de abordagens e técnicas para as quais foram verificados experimentalmente os melhores resultados, considerando-se cada uma das etapas da abordagem proposta. Tais experimentos são apresentados no Capítulo 5. Três imagens foram utilizadas para exemplificar a aplicação de cada módulo (Figura 3.2).

Figura 3.2 – Imagens utilizadas para exemplificar o uso da abordagem proposta: (A) Henrik Andersson; (B) Macaco-prego; e (C) Ilha de Capri.



A apresentação de tais imagens visa a facilitar a compreensão e a descrição/discussão dos resultados obtidos a partir da aplicação dos módulos que compõem a abordagem proposta. A apresentação de cada resultado contém, adicionalmente, um recorte da imagem, inserido com o propósito de evidenciar detalhes do processamento realizado. A escolha destas três imagens deve-se a sua utilização em publicações afins da literatura revisada (e.g., KYPRIANIDIS; DÖLLNER, 2008; ZHAO; ZHU, 2010; KYPRIANIDIS, 2013), assim como ao fato de conterem regiões de alto contraste em cenas de complexidade elevada, i.e., detalhes da barba na imagem de Henrik Andersson¹⁹, região com textura complexa na imagem do macaco-prego²⁰ e uma cena natural²¹, com presença de montanhas e vegetação.

Resultados adicionais gerados a partir da aplicação da abordagem proposta a imagens estáticas, utilizados no processo de avaliação da abordagem proposta (detalhado no Capítulo 4), são apresentados no Apêndice B.

Ressalta-se que a avaliação da abordagem proposta para a renderização não fotorrealística de vídeos foi realizada a partir da utilização de vídeos digitais. A proposição de uma metodologia de avaliação e validação centrada especificamente em vídeos renderizados de forma não fotorrealística constitui uma contribuição adicional da pesquisa ora documentada. Ressalta-se que esta compreendeu a aplicação da abordagem proposta a vídeos digitais (renderizados quadro-a-quadro), enquanto o uso das imagens contidas na Figura 3.2 afigurou-se adequada aos propósitos do

¹⁹ Cortesia de Philip Greenspun, disponível em www.philip.greenspun.com/images/pcd0836/henrik-andersson-56

²⁰ Disponível sob licença Creative Commons em www.flickr.com/photos/22399204@N08/3186409766

²¹ Cortesia de Philip Greenspun, disponível em www.philip.greenspun.com/images/pcd0305/capri-coast-23

presente capítulo, i.e., apresentar detalhadamente os resultados dos módulos que compõem a abordagem proposta.

Ressalta-se que, ao longo deste capítulo há menção ao tempo de execução da implementação computacional de alguns módulos. O computador utilizado para a aferição de tais tempos possui um processador Intel Core i5 (x86-64, modelo M 460), com *clock* de 2,53 GHz, quatro núcleos de processamento e 4 GB de memória principal, utilizando o sistema operacional Microsoft Windows 7 (versão 64 bits).

A seguir, cada etapa do fluxo de execução, ilustrado no diagrama em blocos da Figura 3.1, é detalhada, sendo apresentadas imagens de saída de cada módulo.

Módulo A – Realce da coerência de regiões

Kyprianidis (2013) *apud* Osher e Rudin (1990) e Weickert (1998; 1999) comentaram sobre o efeito artístico obtido por imagens filtradas utilizando-se tensores estruturais de difusão. Tal conceito fundamenta-se na computação dos autovalores e autovetores que refletem a orientação dos contornos na imagem. Tal método, denominado Filtragem com Realce da Coerência de Regiões (*Coherence-Enhancing Filtering* - CEF), faz uso de tensores estruturais de difusão para computar o gradiente da imagem.

Efetivamente, a intensidade da filtragem aplicada à imagem de entrada varia conforme a direção local de cada pixel, não sendo computada em relação à direção perpendicular aos contornos. Especificamente, a CEF utiliza-se da computação das direções normais e tangentes ao gradiente local da imagem. Os valores dos pixels computados correspondem à intensidade do gradiente, enquanto o tensor estrutural de difusão é calculado para cada pixel, a cada iteração da técnica. Tal computação iterativa esmaece e suaviza fracamente o valor do pixel quando este pertence a uma região de borda, enquanto esmaece e suaviza fortemente o valor quando o pixel pertence a regiões homogêneas. Efetivamente, a aplicação da CEF possibilita o realce das bordas e o realce de estruturas localizadas dentro das regiões delimitadas por bordas na imagem de entrada.

Esta técnica, desenvolvida por Weickert (1999), guia o processo de difusão usando um tensor estrutural de difusão suavizado, enfatizando regiões direcionais, a qual está diretamente relacionada à simplificação e estilização de imagens (WEICKERT, 1999; KANG e LEE, 2008). Mais especificamente, o processamento é executado no sentido da menor mudança da direção do tensor estrutural, aumentando a nitidez na direção ortogonal ao tensor estrutural da imagem.

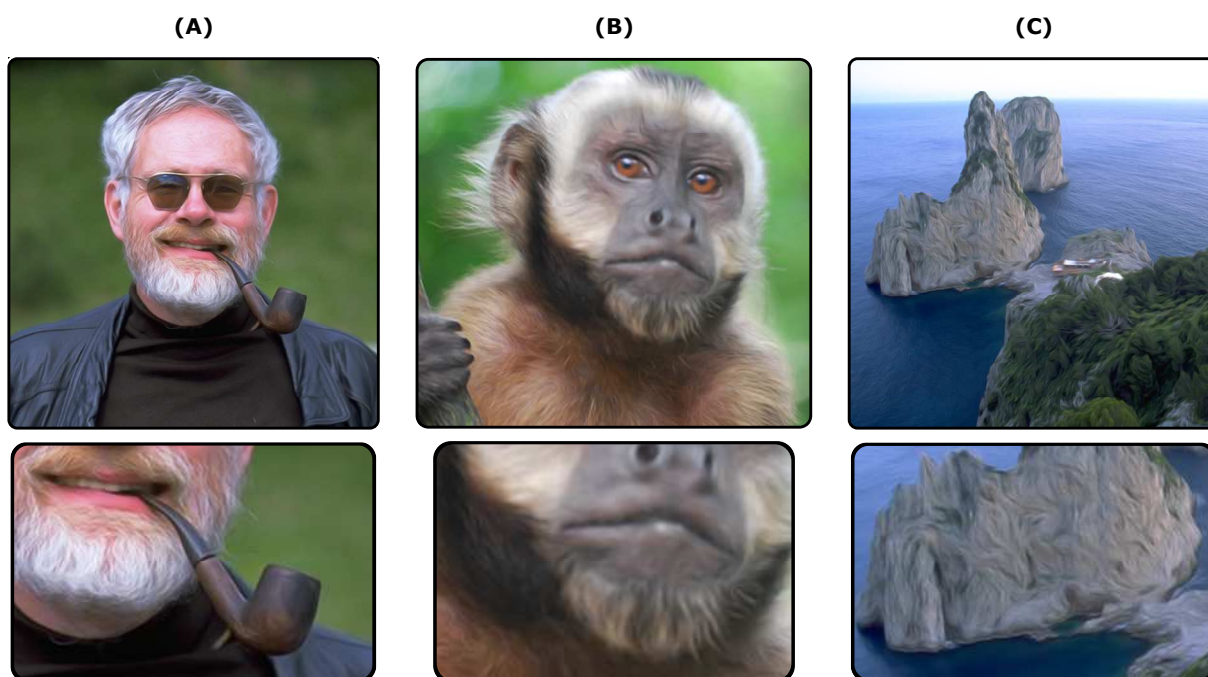
Tal método proporciona o balanceamento entre: (i) o realce de características direcionais da imagem de entrada (regiões de borda); e (ii) a suavização de regiões

homogêneas. Tal informação pode ser empregada para estilizar imagens, por meio do reforço ou exagero de características da estrutura da imagem de entrada, tais como as características direcionais e de curvaturas. De acordo com Weickert (1999), tais características são consideradas aprazíveis e harmônicas pela maioria dos humanos, tendo sido adotadas em algumas formas de arte tradicionais, e.g., em pinturas expressionistas de Van Gogh e Munch.

A filtragem CEF apresenta-se como etapa de pré-processamento, reduzindo o ruído presente na imagem de entrada e preservando as bordas. O resultado obtido a partir da aplicação de tal técnica assemelha-se à composição artística, apresentando pinceladas proeminentes, destacadas ao longo das bordas da imagem de entrada. Com efeito, a técnica preserva e realça características direcionais da imagem, enquanto aumenta o contraste, enfatizando as bordas, silhuetas e estrutura da imagem de entrada.

A inclusão deste módulo justifica-se pela incorporação de um efeito artístico às imagens resultantes do processo de estilização proposto nesta pesquisa. Tal efeito artístico se torna observável após sucessivas aplicações do filtro por um período de tempo de processamento compatível (detalhado na Seção 3.2) com os objetivos da pesquisa realizada. Tal efeito visou a melhorar o aspecto das arestas da imagem de entrada e adicionar um efeito de redemoinho para fins de abstração, enfatizando características curvilíneas presentes em regiões da imagem de entrada. Na Figura 3.3 são apresentados resultados após a aplicação da CEF.

Figura 3.3 – Resultado da aplicação da CEF para realce de características locais.



Com efeito, a revisão bibliográfica realizada permitiu verificar o emprego

específico do tensor estrutural da imagem de entrada em módulos subsequentes de outras abordagens (KANG e LEE, 2008; KYPRIANIDIS, 2013). Destaca-se, neste ponto, uma diferença importante no tocante ao uso de tal tensor na abordagem proposta nesta pesquisa: os autovetores, autovalores e o tensor que estima a estrutura local da imagem de entrada não são utilizados nos demais módulos da abordagem proposta. A imagem de entrada é processada pelo Módulo A e a imagem resultante de tal processamento é utilizada diretamente pelos módulos subsequentes, i.e., a computação do tensor de difusão é utilizada apenas na geração da imagem cujos contornos e regiões direcionais foram realçados.

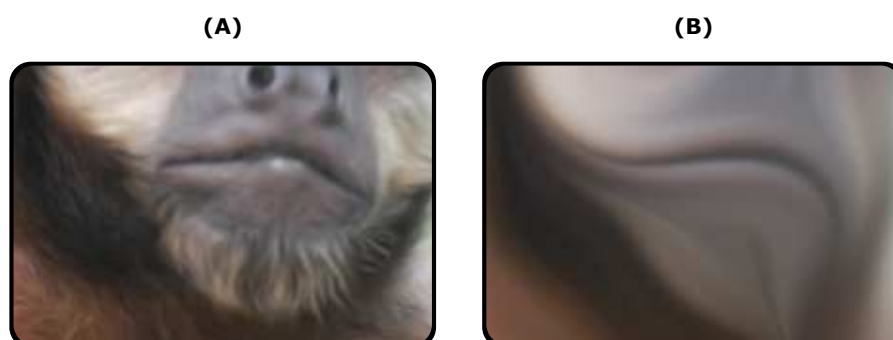
Ressalta-se uma dificuldade encontrada ao longo de toda a pesquisa ora documentada, i.e., a atribuição de valores aos vários parâmetros intrínsecos a cada módulo ou técnica pesquisada, de modo que os resultados obtidos se afigurassem adequados aos objetivos previamente definidos. A experimentação relativa a tais atribuições é detalhada no Capítulo 5, no qual são ressaltados valores empregados durante o processo de avaliação e validação da abordagem proposta.

No módulo A, ressaltam-se três parâmetros importantes: o tamanho do passo temporal ($\tau=2$), as dimensões e o tipo da máscara utilizada, i.e., máscara invariante à rotação, de dimensões 5x5, empregando a discretização padrão definida por Weickert (1999) e o valor do desvio padrão para a suavização do tensor ($\rho=5$).

Na Figura 3.4, exemplifica-se, em um excerto da imagem do macaco-prego, a aplicação da CEF utilizando-se valores adequados, $\tau=2$ e $\rho=5$ (Figura 3.4A) e inadequados, $\tau=500$ e $\rho=15$ (Figura 3.4B) aos objetivos de pesquisa.

Observa-se, no resultado apresentado da Figura 3.4B a total descaracterização do objeto retratado na imagem quando valores de parâmetros muito altos são utilizados. Ademais, verificou-se que para $\tau=500$ e $\rho=15$ a finalização do processamento deu-se após cerca de 30 minutos, utilizando o *hardware* descrito no início deste capítulo.

Figura 3.4 – Aplicação do CEF com valores de parâmetros diferenciados.



A metodologia adotada para a obtenção de tais valores está apresentada no Capítulo 4 (Seção 4.1.3, p. 126), no qual se ressaltam os critérios de seleção das abordagens a serem comparadas e configuração de seus parâmetros.

A validação experimental dos valores obtidos e as considerações sobre a parametrização para todos os módulos da abordagem proposta são apresentadas no Capítulo 5 (Seção 5.1, p. 136). Em tal capítulo, a calibração foi realizada a partir do uso de subgrupos de valores para cada parâmetro. Esta divisão em subgrupos determinou um conjunto de valores para cada parâmetro investigado.

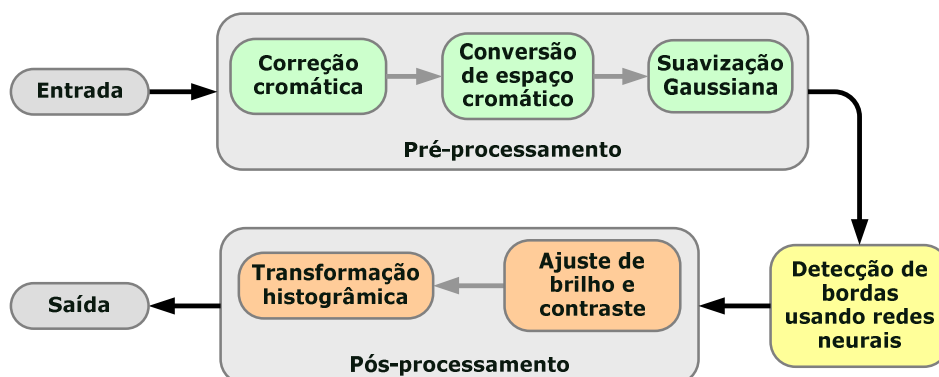
Ressalta-se que um aprimoramento da CEF foi proposto por Weickert (1999), a partir da inclusão de tensores de difusão adaptativos, os quais variavam direcionalmente o nível de suavização em cada local, visando a realçar as curvas suaves de forma diferenciada. Outras extensões foram propostas por Krissian et al. (2007), os quais produziram os tensores de difusão adaptativos nos quais estão presentes a variação da intensidade na direção dos contornos e nas direções de curvaturas principais. Detalhes adicionais acerca da CEF foram documentados por Weickert (1999), Brox et al. (2006), Tabik et al. (2006), Krissian et al. (2007), Mendrik et al. (2009) e Kroon, Slump e Maal (2010).

Módulo B – Extração de bordas utilizando redes neurais

De acordo com Gonzalez e Woods (2007), *bordas* são caracterizadas por transições abruptas nos valores de intensidade dos pixels vizinhos de uma imagem. A partir do emprego de técnicas destinadas à extração de bordas, os contornos dos elementos significativos da imagem são identificados.

Ressalta-se, no entanto, outra diferença importante na abordagem proposta nesta pesquisa, no tocante ao uso do mapa de bordas: as bordas utilizadas na abordagem proposta são geradas por meio da utilização do detector de bordas proposto pelos autores desta pesquisa (ARRUDA; QUEIROZ; GOMES, 2012), cujo diagrama em blocos é apresentado na Figura 3.5, e descrito sucintamente a seguir.

Figura 3.5 – Abordagem para detecção de bordas utilizando redes neurais.



A imagem de entrada é submetida a três etapas de pré-processamento, a saber: (i) *correção cromática* com o uso de algoritmo de constância cromática (*GrayWorld*); (ii) *conversão de espaço cromático*, do espaço RGB para o espaço YCbCr, com a utilização apenas do canal cromático Y; e (iii) *suavização Gaussiana* com $\sigma=2$. Os valores dos parâmetros utilizados no módulo B foram obtidos após a condução de um processo experimental, o qual está descrito na pesquisa realizada por Arruda, Queiroz e Gomes (2012).

Após a etapa de pré-processamento, a imagem resultante é submetida a uma filtragem espacial usando uma máscara (*kernel*) de dimensões 5x5. Tal filtragem é realizada por um classificador neural tipo *Perceptron* Multicamadas, treinado com imagens sintéticas para a detecção de bordas. 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, de acordo com Gomes (2002), é dada por:

$$c = \frac{|L_{\max} - L_{\min}|}{L_{\max} + L_{\min}} \quad (3.1)$$

na qual L_{\max} e L_{\min} são as intensidades máxima e mínima associadas a um recorte de imagem. O resultado deste submódulo é a produção de um mapa de bordas.

Após o processo de 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. O brilho e contraste da imagem resultante da etapa anterior são modificados de acordo com valores fixos, obtidos de forma experimental por Arruda, Queiroz e Gomes (2012), no qual o valor do ajuste do brilho é igual a -80 e o valor do ajuste para o contraste é igual a 80. A saída deste módulo é o mapa de bordas.

Ressalta-se a modificação da etapa de pós-processamento, presente no módulo B, visando à identificação do valor para o parâmetro que altera a resposta do detector de bordas, no tocante à quantidade e qualidade subjetiva das bordas detectadas. Visando a diminuir a proeminência e a espessura das bordas detectadas, a fim de obter um aspecto menos ruidoso da imagem e melhorar o aspecto geral da renderização final, os valores do brilho e do contraste foram ajustados, na etapa de pós-processamento, para -5 e 10, respectivamente. O efeito da modificação dos valores destes parâmetros pode ser visualizado na Figura 3.6. O detalhamento da

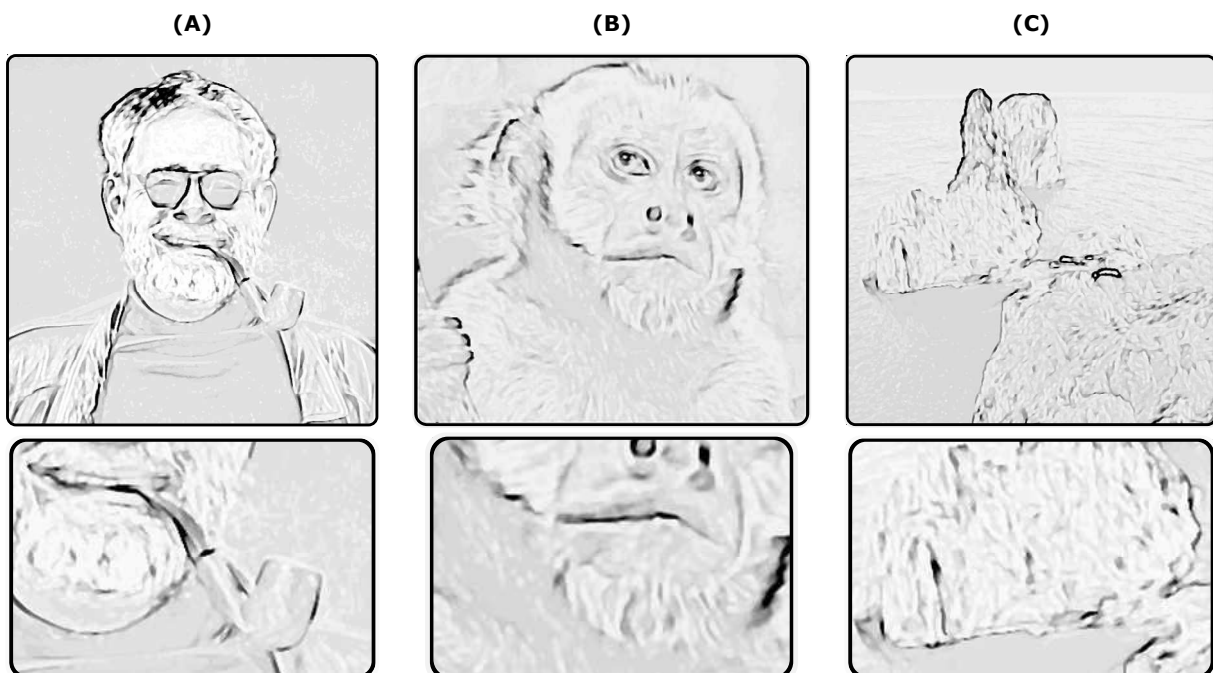
pesquisa relativa à extração de bordas utilizando redes neurais é apresentado no Apêndice N.

Figura 3.6 – Alterações realizadas no detector de bordas neural: (A) Resultado original obtido por Arruda; Queiroz; Gomes (2012); (B) Resultado após modificação de valores de parâmetros.



Após os experimentos realizados e relatados no Capítulo 5, o módulo destinado à extração de bordas gerou os resultados apresentados na Figura 3.7. Após a etapa de extração de bordas, ocorre a etapa de suavização por difusão anisotrópica, relatada a seguir.

Figura 3.7 – Resultado da aplicação da detecção de bordas que utiliza redes neurais.



Módulo C – Suavização adaptativa por difusão anisotrópica modificada

Conforme relatado no Capítulo 2, a suavização por difusão anisotrópica conserva a nitidez e o delineamento das bordas na imagem de entrada. No entanto, após a condução de experimentos envolvendo a utilização do algoritmo de suavização por difusão anisotrópica proposto por Perona e Malik (1990) e Black et al (1998),

verificou-se a necessidade de realização de duas alterações, a saber: (i) a alteração da *função de parada-na-aresta*; e (ii) a combinação dos resultados obtidos na suavização com as bordas detectadas no módulo B.

As alterações propostas tiveram como objetivo o aprimoramento do processo de renderização quando são utilizados algoritmos de suavização com preservação de bordas. A alteração visou a evitar a suavização dos pixels que se encontram na região de transição, ao mesmo tempo em que são suavizadas as regiões de cor similar que compõem o quadro do vídeo de entrada.

A alteração foi realizada na Eq. 2.4.8, que representa o cálculo da magnitude do gradiente, da seguinte forma:

$$\nabla I_{s,p}(t) = [I(p,t) - I(s,t)] + [B(p,t) - B(s,t)], \quad p \in \eta_s. \quad (3.2)$$

na qual **B** representa o mapa de bordas gerado na etapa anterior (extração de bordas), no ponto **s**, na direção **(s,p)** na iteração **t**. Tais alterações implicaram o aumento do valor da magnitude do gradiente calculado, i.e., a importância das transições para a suavização por difusão anisotrópica foi reforçada. Experimentos realizados e relatados no próximo capítulo possibilitaram, a partir dos resultados obtidos, a definição dos parâmetros *velocidade da difusão* (λ) e *número de iterações* (**t**) como 1,0 e 30, respectivamente.

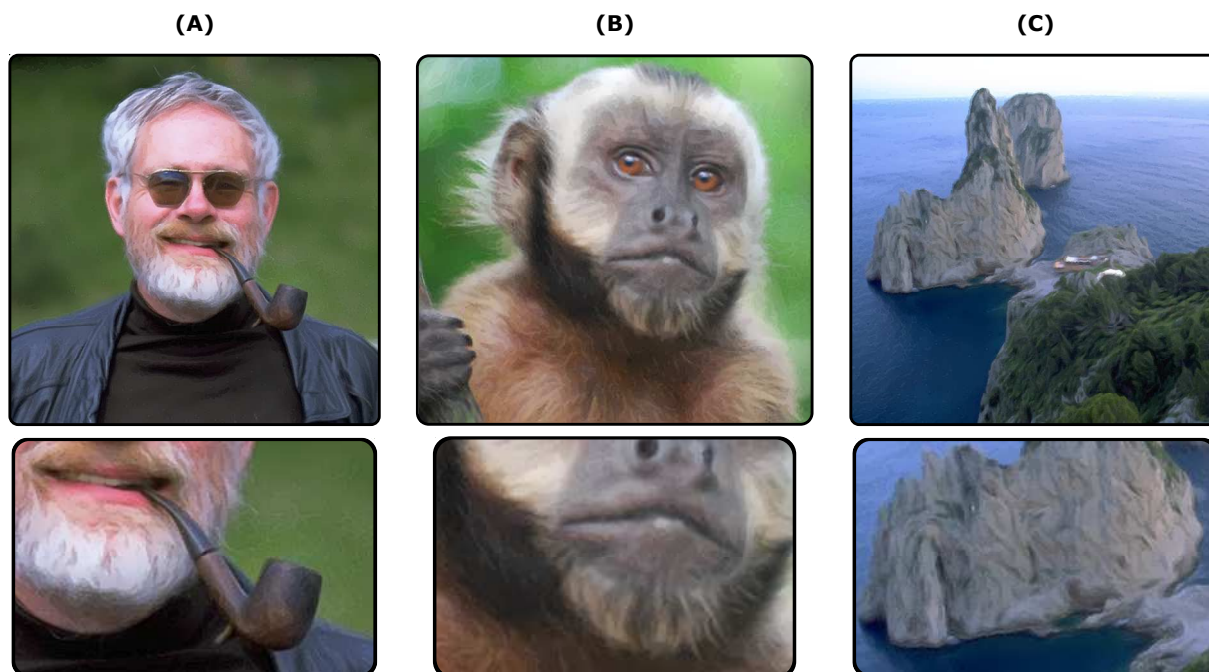
Após a condução de experimentos, observou-se que, dentre as funções de parada-na-aresta testadas, aquela que mais se adequou aos propósitos da pesquisa foi a *função de Dennis e Welsch* (1978), apresentada na Eq. 3.3. Os resultados obtidos e as funções testadas são apresentados no Capítulo 5. A razão para a escolha da função de Dennis e Welsch (1978) deveu-se à observação de um efeito mais acentuado de estilização do resultado obtido, i.e., os detalhes do quadro de vídeo são mais abstraídos, quando comparados à utilização de outras funções de parada-na-aresta, sem o comprometimento do aspecto geral da imagem renderizada.

$$\psi_4(x) = xe^{\left(\frac{-x}{\sigma}\right)^2} \quad (3.3)$$

A versão original do algoritmo de suavização anisotrópica produziu resultados visualmente inadequados para os propósitos da renderização não fotorrealística almejada, devido à suavização excessiva das imagens testadas. Ressalta-se, neste ponto, que o ajuste de sensibilidade do Módulo 1 (detecção de bordas), a combinação das bordas detectadas com a saída do Módulo 2 (suavização por difusão anisotrópica) e a modificação da função de parada-na-aresta do algoritmo de suavização por

difusão anisotrópica constituem contribuições originais da abordagem proposta para a renderização não fotorrealística de vídeos. Resultados obtidos após o processamento supradescrito são ilustrados na Figura 3.8.

Figura 3.8 – Resultado da aplicação do processo de difusão anisotrópica modificado.



Após a obtenção do quadro do vídeo abstraído, obtido por meio da combinação das etapas de detecção de bordas (Módulo B) e de suavização por difusão anisotrópica (Módulo C), seguiu-se a etapa de realce das regiões sombreadas (Módulo D); a superposição das bordas à imagem colorida; a filtragem com o uso da filtragem NLM (Módulo E) e, finalmente, a etapa de realce de cores da imagem (Módulo F). Tais módulos foram inseridos na abordagem proposta com o propósito de incorporar métodos de pós-processamento, cujo objetivo foi o de melhorar o aspecto visual do resultado do processo de renderização, visando a atender aos objetivos da pesquisa ora definidos.

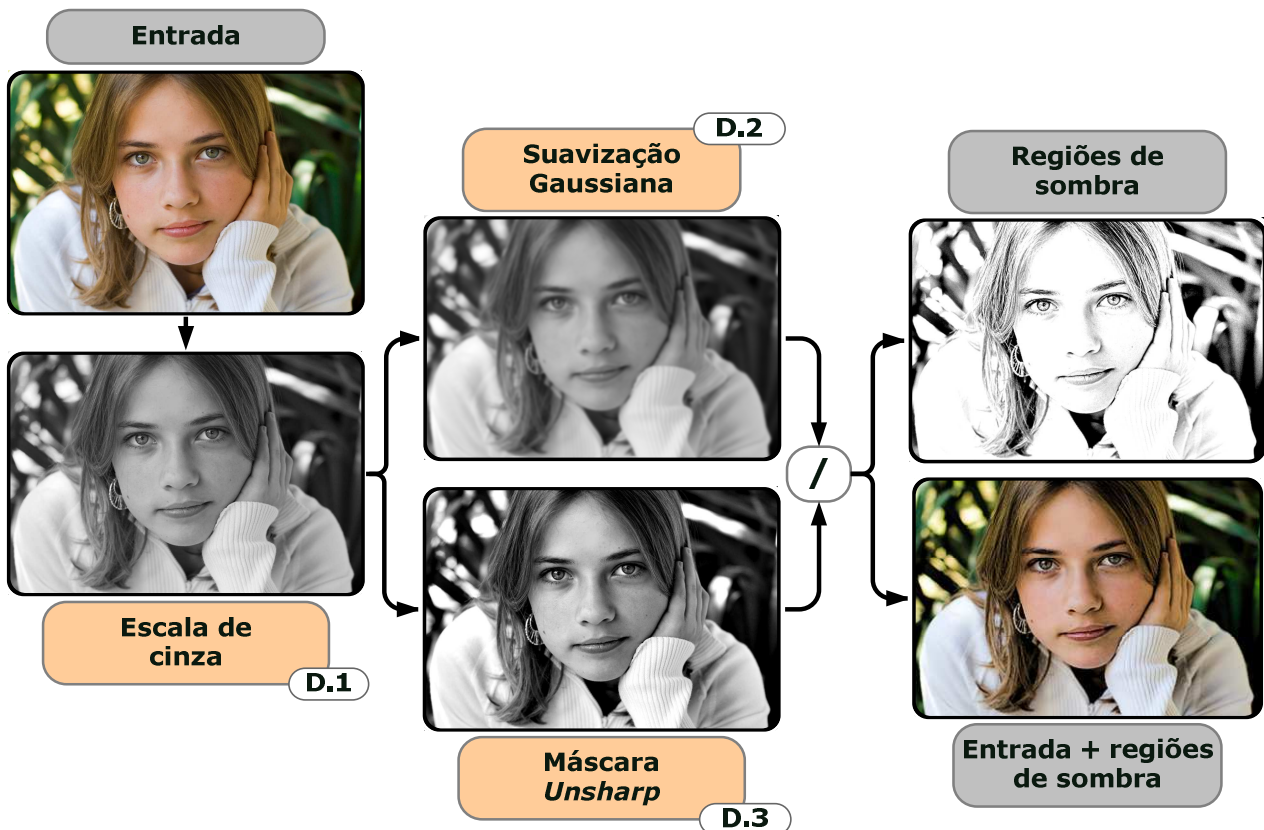
Módulo D – Realce de sombras e superposição das bordas

A etapa de realce das regiões de sombras e a superposição das bordas detectadas no módulo B foram incorporadas à abordagem proposta como forma de pós-processamento, inspiradas na abordagem generalizada por Kiprianidis (2013) e nas pesquisas realizadas por Luft, Colditz e Deussen (2006) e Pereira (2008). Esta etapa visa à melhoria do aspecto visual dos resultados obtidos, particularmente no tocante: (i) ao realce de detalhes de escala fina da imagem; (ii) à suavização adicional da imagem resultante, visando produzir resultados com maior abstração, sem comprometimento do conteúdo da imagem; e (iii) ao incremento da saturação da

imagem de saída.

Uma síntese gráfica da estratégia destinada ao realce das regiões de sombra adotada na abordagem proposta é apresentada na Figura 3.9. A estratégia se dá em três estágios, a saber: (D.1) a conversão da imagem proveniente da etapa anterior em uma imagem em escala de cinza; (D.2) a suavização Gaussiana da imagem em escala de cinza; e (D.3) a computação de uma máscara de realce do tipo *unsharp*. Após o processamento dos três estágios, ocorre a superposição da imagem obtida no estágio D.2 pela imagem obtida no estágio D.3.

Figura 3.9 – Módulo de realce de regiões de sombra.



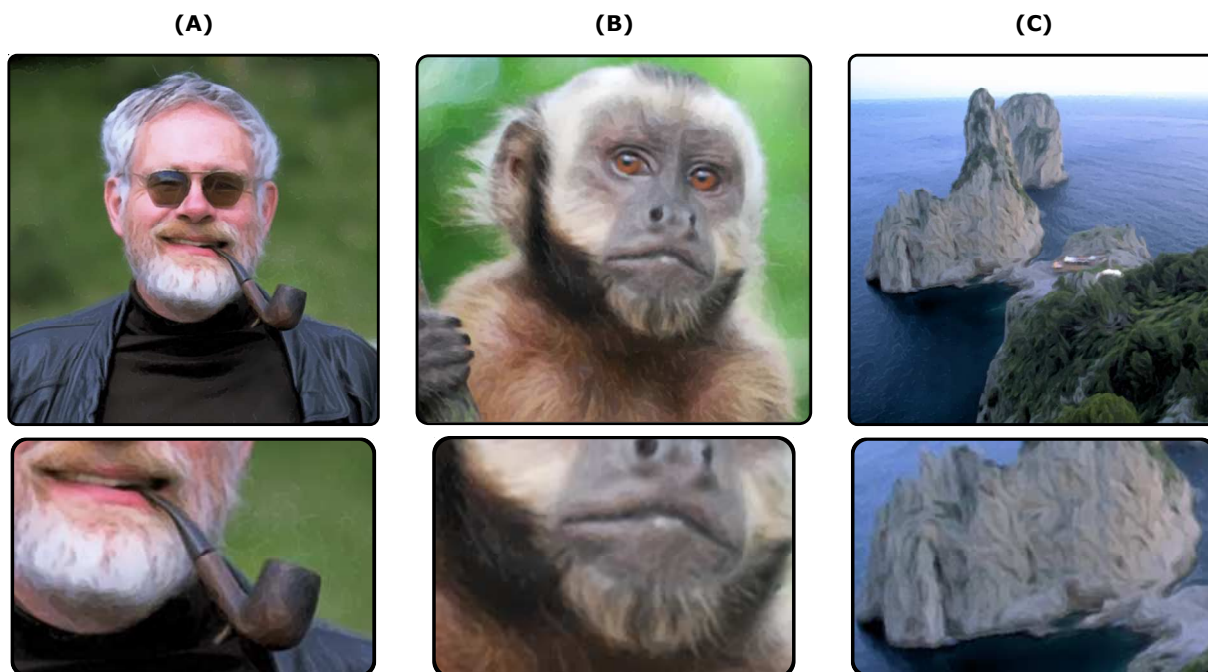
O processo de superposição e geração da imagem que contém as regiões de sombra é realizado de acordo com a Eq. 3.4.:

$$(D_3 * 256)/(D_2 + 1) \quad (3.4)$$

na qual D_3 representa o valor de um pixel da imagem resultante do módulo D.3 e D_2 representa o valor de um pixel da imagem resultante do módulo D.2. Após a etapa de realce de regiões de sombra (módulo D), pondera-se a imagem resultante com as bordas geradas pelo módulo B. A ponderação ocorre com a multiplicação dos valores dos pixels de todos os canais da imagem suavizada (módulo D) pelos valores dos pixels da imagem de borda (módulo B). Após a multiplicação de tais valores, ocorre a

divisão por 255 (valor máximo do pixel quando são utilizados 8 bits para representação computacional). Exemplos de resultados deste processamento podem ser visualizados na Figura 3.10.

Figura 3.10 – Resultado da aplicação do processo de realce de regiões de sombra.



Módulo E – Filtragem com utilização de médias não locais

A suavização por *médias não locais* (*NLM – Non Local Means*), proposta originalmente por Buades, Coll e Morel (2005), tem como propósito a remoção de ruído em imagens estáticas. A remoção de ruído ocorre a partir da substituição do valor de cada pixel por uma média ponderada não local, i.e., uma média calculada a partir de blocos da imagem. Originalmente, a técnica busca regiões, em toda a imagem, que possuam valores similares, realizando a filtragem com relação a todas as regiões buscadas.

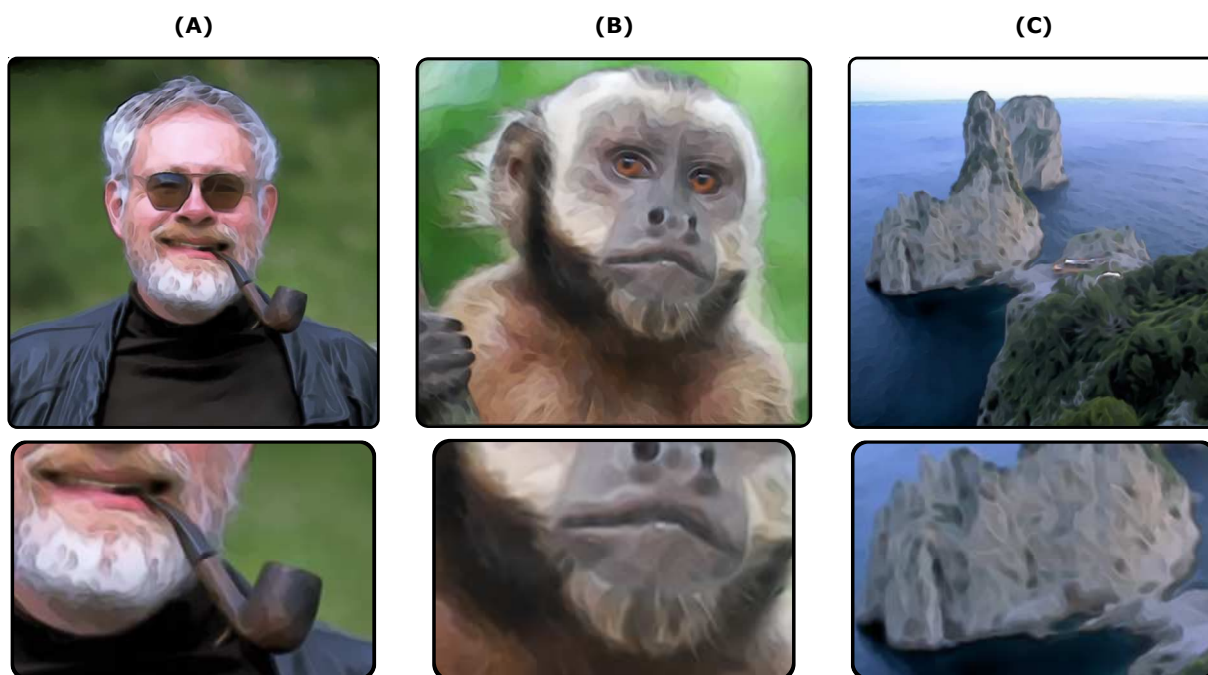
Encontra-se na literatura (GOOSSENS et al., 2010; PALMA et al., 2013) métodos distintos para: (i) buscar regiões similares; e (ii) filtrar os valores dessas regiões. Uma modificação do algoritmo NLM comumente encontrada refere-se à busca de regiões semelhantes ao redor de uma região retangular do pixel a ser filtrado, i.e., a busca não é realizada em toda a imagem de entrada. Assim sendo, a região de busca restringe-se ao entorno da vizinhança do pixel a ser filtrado.

A filtragem NLM pode ser empregada no processo de filtragem de imagens ou na reconstrução de imagens em superresolução (PROTTER et al., 2009). Destaca-se que Gambarra et al. (2012) apresentou uma implementação em *hardware* (*Field-Programmable Gate Array - FPGA*) para a filtragem NLM. Nessa implementação, tal modificação está presente, i.e., a filtragem é realizada com a utilização de dimensões

fixas das máscaras (*kernel*) utilizadas no algoritmo. Na pesquisa realizada por Gambarra et al. (2012), a dimensão da janela de busca por pixels similares foi restringida a áreas de 21x21 pixels, e as dimensões da janela de filtragem foram restringidas a 7x7 pixels. Os autores obtiveram resultados similares à técnica proposta por Buades, Coll e Morel (2005), tanto em termos de MSE quanto de percepção visual.

Fundamentado na pesquisa de Coupé et al. (2012), Dirk-Jan Kroon disponibilizou uma implementação computacional em MATLAB²² e em linguagem de programação C, a qual contém uma versão rápida da filtragem NLM. Após a realização de experimentos (ver Seção 5.1.4, p. 144), os valores para os parâmetros da técnica NLM que se afiguraram mais adequados aos propósitos de pesquisa foram: (i) dimensão da região de busca por pixels similares definida em 5x5; (ii) dimensão da janela de filtragem definida em 3x3; e (iii) força da filtragem (*filtering strength*), definida em 0,04. O efeito da aplicação da filtragem NLM ao contexto da abordagem proposta pode ser visualizado na Figura 3.11.

Figura 3.11 – Resultado da aplicação da filtragem NLM a três diferentes imagens de entrada.



No tocante às diferenças visuais obtidas mediante a aplicação da filtragem NLM, percebe-se maior abstração do resultado obtido. Em especial, constata-se a preservação das fronteiras entre regiões, assim como maior suavização de regiões da barba, dos óculos, do cachimbo, das narinas do macaco e da região de montanha, nos temas das cenas de exemplificação. Após a aplicação da filtragem NLM, procedeu-se à

²² Disponível em: <http://www.mathworks.com/matlabcentral/fileexchange/27395-fast-non-local-means-1d-2d-color-and-3d>

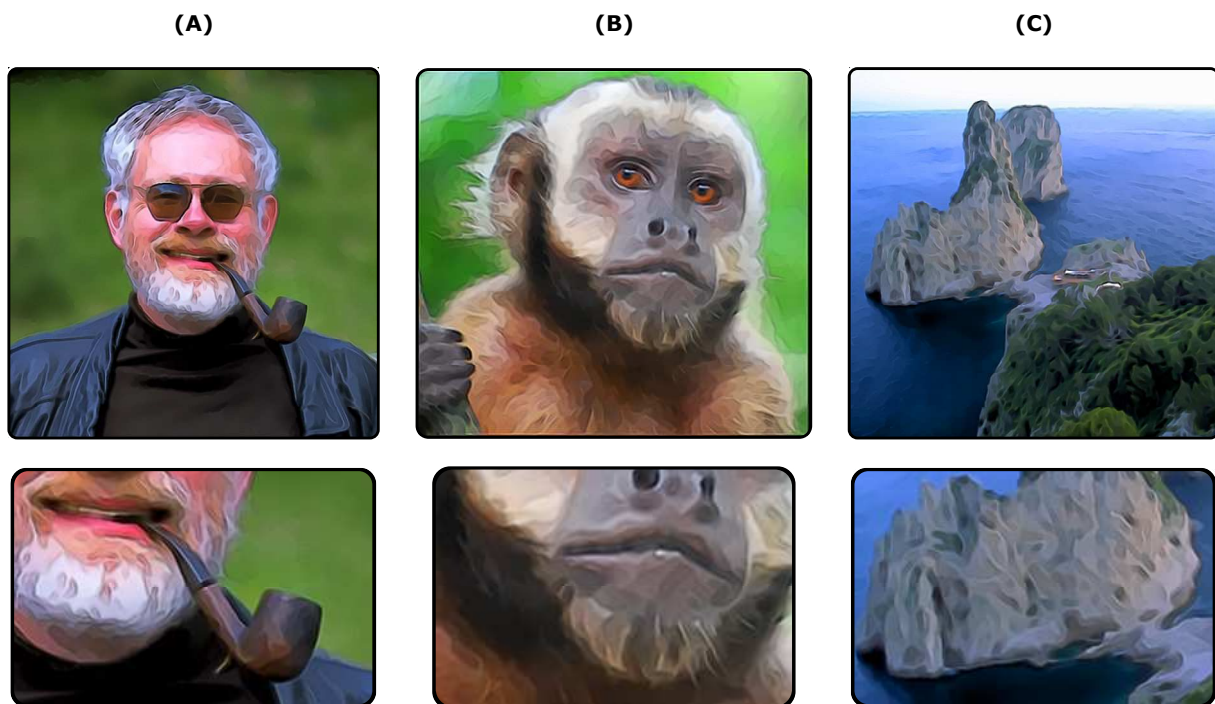
etapa de modificação da saturação da cena, a qual é aplicada a cada imagem/quadro do vídeo de entrada, conforme descrito a seguir.

Módulo F – Ajuste de saturação

De acordo com Ianeva (2003), cartuns possuem características marcantes, no tocante à paleta reduzida de cores e à saturação exagerada da cena. Tal comportamento suscitou a inclusão do módulo de ajuste de cores por meio do incremento da saturação dos quadros que compõem o vídeo abstraído. O ajuste de saturação foi aplicado visando à modificação da faixa dinâmica do canal de saturação, após a conversão da imagem para o espaço cromático HSV, com acréscimo de 30% na saturação global da imagem.

Tal operação, no contexto desta pesquisa, visou a ressaltar os detalhes da representação não fotorrealística gerada nas etapas anteriores. A etapa de transformação do canal de saturação foi incorporada à abordagem proposta e exemplos do resultado desta transformação podem ser visualizados na Figura 3.12.

Figura 3.12 – Resultado do ajuste de saturação aplicado as três diferentes imagens de entrada.



Na Figura 3.14 apresenta-se uma comparação visual entre as imagens originais e imagens abstraídas após o uso da abordagem proposta. Ressalta-se o efeito harmonioso, do ponto de vista subjetivo, da abstração obtida. Na Figura 3.15 apresenta-se uma síntese gráfica da abordagem proposta, na qual podem ser visualizadas todas as etapas do processo de renderização.

Figura 3.13 – Coleção de imagens renderizadas com o uso da abordagem proposta: (A) Imagem original; e (B) Imagem renderizada.

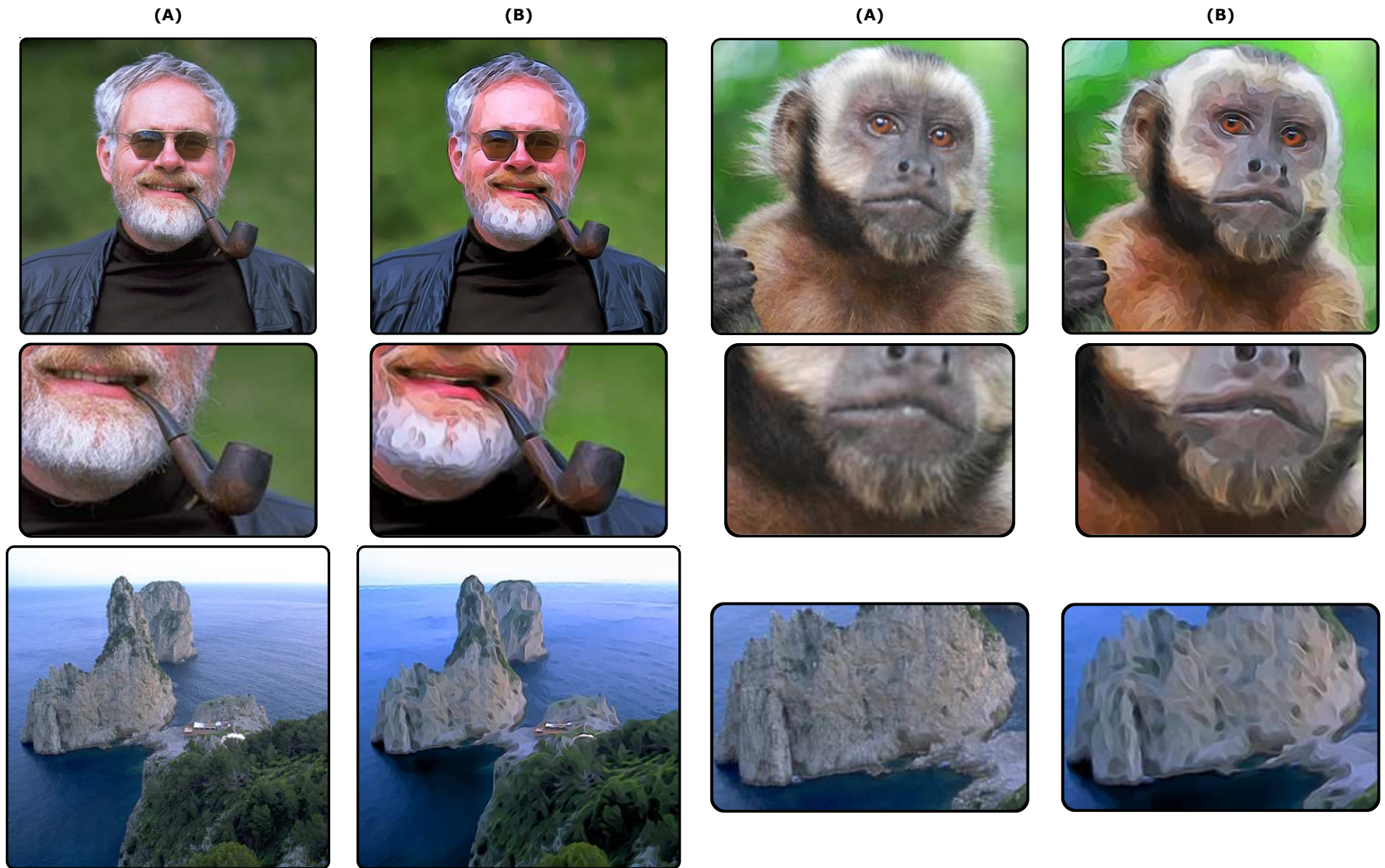
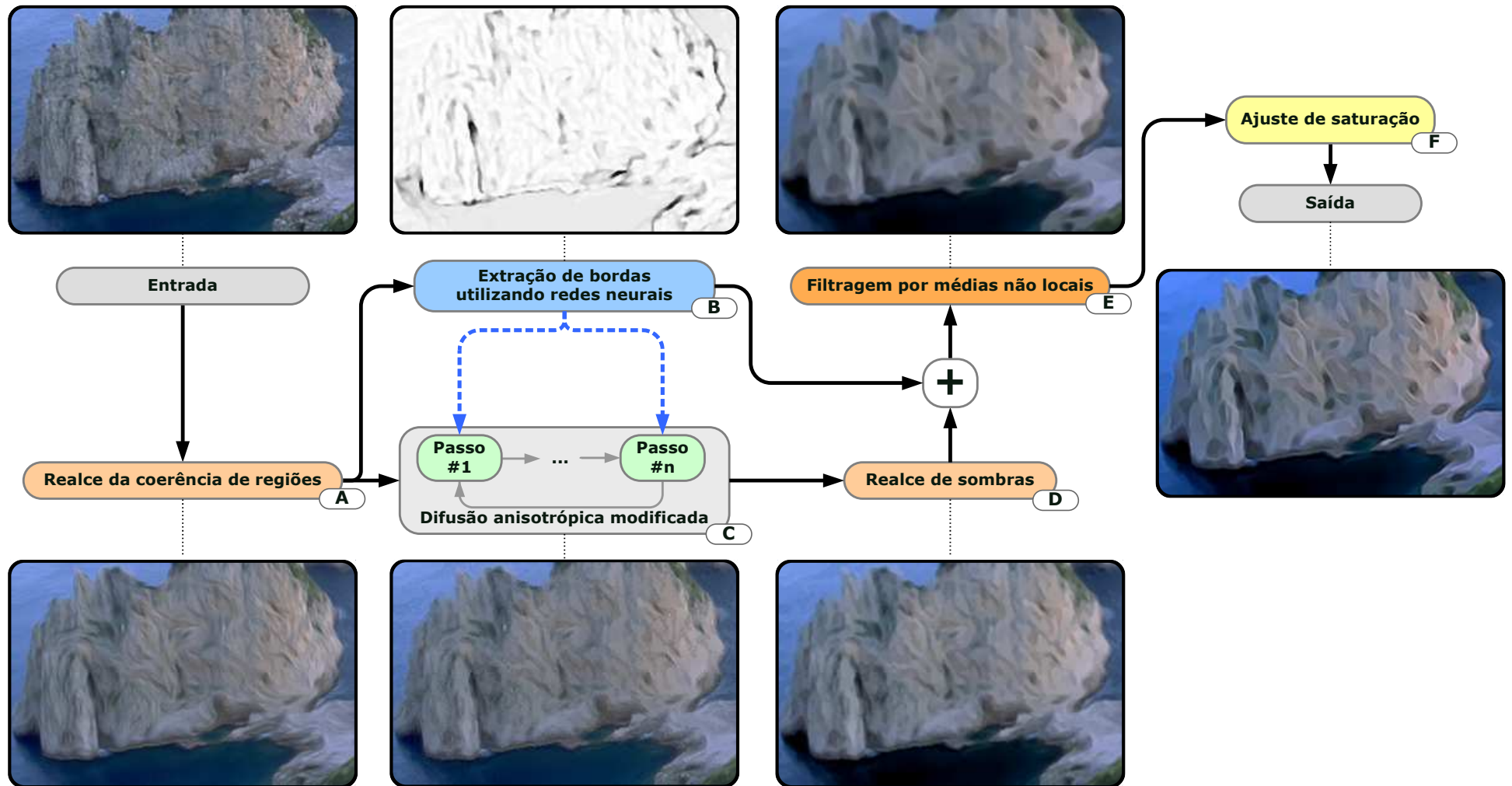


Figura 3.14 – Síntese gráfica de todas as etapas da abordagem proposta.



3.2 Limitações da Abordagem Proposta

As seguintes limitações e considerações acerca da abordagem proposta são elencadas, as quais podem servir de motivações para investigações futuras:

- Durante a realização de testes, na fase de planejamento do experimento para a avaliação da abordagem proposta (ver Capítulo 4), três vídeos que retratam cenários variados foram renderizados e se observou o aparecimento de cintilação (*flickering*) quando o valor do passo temporal τ , relativo à CEF, foi elevado para valores acima de $\tau=50$. Tal fato implicou a redução do número de iterações para a geração dos resultados, até que o efeito de cintilação não fosse observado.

- Conforme ressaltado anteriormente, Kyprianidis et al. (2013) afirmaram que, apesar das vantagens inerentes à renderização não fotorrealística fundamentada na filtragem de imagens, a possibilidade de definição de estilos ao serem aplicadas técnicas de filtragem espacial é limitada. Isto pode ser contornado mediante a interação do usuário com o sistema de renderização ou quando se empregam componentes de análise de alto nível do vídeo de entrada. Tal limitação pode ser verificada na abordagem proposta, a qual produz, atualmente, um estilo único de renderização. Tal limitação pode ser compensada se forem incorporados novos estilos ao módulo de extração de bordas, a partir da inclusão de efeitos ou de alterações do mapa de bordas gerado.

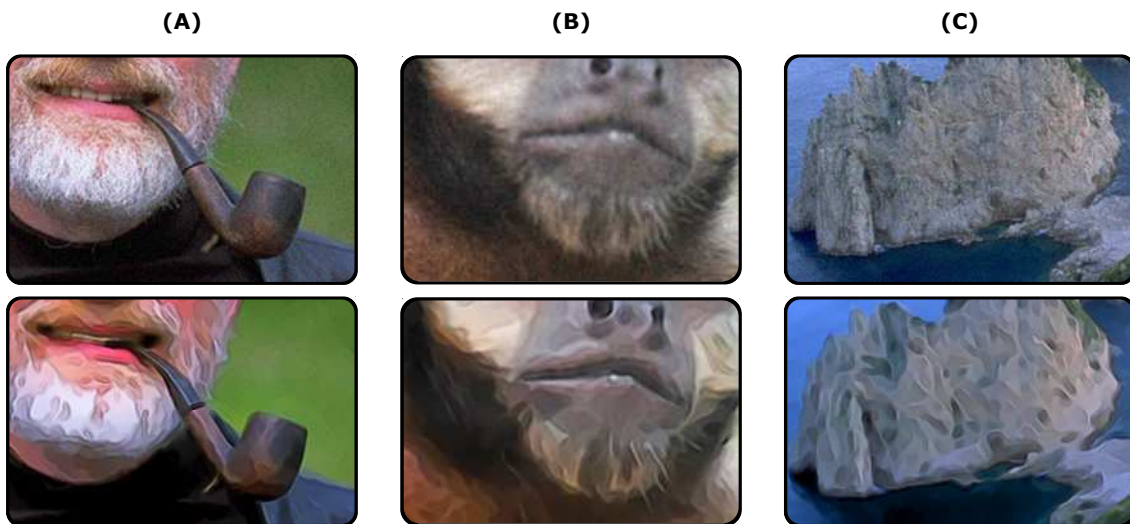
- Após a realização de experimentos com 20 imagens, verificou-se que a renderização realizada com a utilização da abordagem proposta em imagens com dimensões inferiores a 200x200 pixels não apresentou resultados satisfatórios, do ponto de vista visual. Resultados das imagens de baixa resolução apresentaram borramento acentuado e linhas de borda espessas.

- A implementação computacional da abordagem proposta foi realizada majoritariamente a partir do emprego da linguagem de programação e do ambiente MATLAB. Apenas o módulo B foi desenvolvido na linguagem de programação C++, empregando-se a biblioteca Intel OpenCV, incorporada à abordagem proposta por meio do comando *system* do MATLAB. Nos módulos A e E, operações de convolução e operações com vetores e matrizes foram compiladas para código nativo em linguagem C, a partir da utilização da função *mex* do ambiente MATLAB. Algumas considerações relativas ao tempo de execução de cada módulo da abordagem proposta são apresentadas no Apêndice M.

- A investigação do comportamento da abordagem proposta ao renderizar imagens com ruído foi realizada apenas de forma visual e subjetiva e não foi

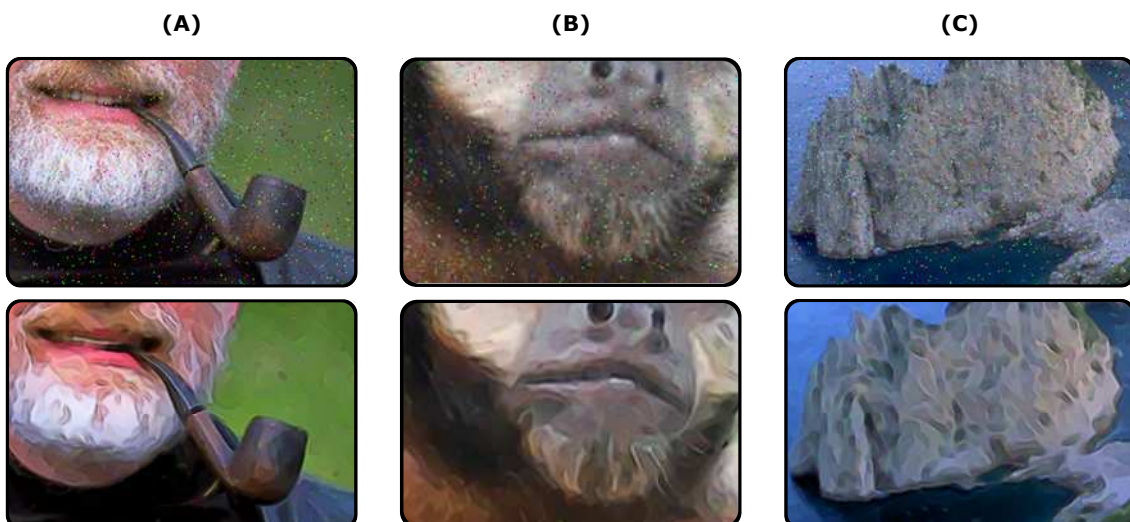
incorporada ao processo de avaliação e validação da abordagem proposta, por se tratar de um aspecto fora do escopo desta pesquisa. Dois tipos de ruído foram considerados: (i) o ruído Gaussiano, com $\sigma=0,001$; e (ii) o ruído sal e pimenta, com densidade 0,02. Na Figura 3.15 são visualizados excertos de imagens de entrada corrompidas com ruído Gaussiano e o efeito de renderização produzido, após a aplicação da abordagem proposta.

Figura 3.15 – Abordagem proposta aplicada a imagens corrompidas com ruído Gaussiano.



Na Figura 3.16 são visualizados excertos das três imagens utilizadas neste capítulo, corrompidas com ruído sal e pimenta, assim como o efeito de renderização produzido após a aplicação da abordagem proposta para a renderização não fotorrealística. Percebe-se, visualmente, a robustez da abordagem proposta ao renderizar imagens ruidosas, e.g., há a presença de poucos artefatos advindos do ruído na renderização do fundo das imagens apresentadas. Investigações futuras poderão contemplar a inclusão de imagens ruidosas no processo de avaliação.

Figura 3.16 – Abordagem proposta aplicada à imagem corrompida com ruído sal e pimenta.



3.3 Considerações Finais do Capítulo

Neste capítulo, foi descrita em detalhes a abordagem proposta para a geração de representações não fotorrealísticas de vídeos digitais ao estilo de cartuns. A abordagem preliminar, fundamentada em técnicas de segmentação espaçotemporal de vídeos e, a princípio, promissora para atender aos objetivos da pesquisa, apresentou-se inviável, em função de limitações encontradas, dentre as quais se destacam a inadequação: (i) dos resultados obtidos ao integrar resultados de algoritmos de segmentação e computação de fluxo óptico; (ii) dos parâmetros definidos para o algoritmo de segmentação de regiões escolhido; e (iii) do processo de computação do fluxo óptico nos vídeos de teste de entrada, o que implica a inviabilidade de integração do resultado ao resultado produzido pelo algoritmo de segmentação. Descrições detalhadas destas limitações podem ser encontradas no Apêndice A.

Assim, os esforços de pesquisa foram redirecionados para a proposição de uma abordagem fundamentada na suavização de quadros do vídeo de entrada por difusão anisotrópica, utilizando uma função de parada-na-aresta modificada, combinada com um detector de bordas, em quadros consecutivos do vídeo de entrada. Conforme descrito na Seção 3.1, os seis módulos que compõem a abordagem proposta estão associados a parâmetros cujos valores foram calibrados para a apresentação dos resultados que se afiguraram mais adequadas aos propósitos da abordagem proposta, no que concerne à avaliação realizada durante a pesquisa.

A concepção da abordagem proposta, fundamentada e inspirada em abordagens presentes na revisão da literatura da área (WINNEMÖLLER; OLSEN; GOOCH, 2006; KYPRIANIDIS; DÖLLNER, 2008; KANG; LEE; CHUI, 2007; KYPRIANIDIS, 2011; 2013), constitui um elemento inovador no que tange aos propósitos da renderização não fotorrealística e aos objetivos pré-definidos (ver Capítulo 1), visto que a abordagem proposta contém ou apresenta: (i) modificações na disposição dos módulos destinados à renderização não fotorrealística; (ii) modificações em técnicas/algoritmos utilizados em dois dos módulos apresentados (módulos A e B); (iii) dois módulos inexistentes em abordagens concorrentes (módulos E e F); (iv) estilo de renderização diferenciado daqueles presentes na literatura pesquisada; (v) o emprego de redes neurais como parte integrante do processo de renderização (módulo não encontrado na literatura revisada); (vi) a combinação diferenciada do módulo de detecção de bordas ao módulo utilizado para a suavização por difusão anisotrópica; (vii) redução do número de iterações no processo

de suavização por difusão anisotrópica, posto que as bordas detectadas auxiliam na obtenção do resultado final; (viii) módulo para o realce e a saturação de cores e sombras; (ix) robustez à presença de ruído, aferida de forma subjetiva, em imagens comumente utilizadas na área de NPR.

Conquanto ao conteúdo inovador relatado anteriormente, a abordagem proposta apresenta limitações, e.g., a presença do estilo singular dos resultados apresentados, a inadequação ao processamento computacional em tempo real e, ainda, a falta de uma ferramenta computacional unificada, no tocante ao emprego de uma única linguagem/ambiente de programação.

O próximo capítulo contém a descrição da metodologia destinada à avaliação de renderizações NPR, cuja proposição visou à sistematização do processo de comparação com outras abordagens destinadas à renderização não fotorrealística de vídeos digitais.

Capítulo 4

Metodologia de Avaliação de Renderizações Não Fotorrealísticas de Vídeos

Neste capítulo, é descrita a metodologia adotada no processo de avaliação de vídeos renderizados não fotorrealisticamente ao estilo de cartuns, a qual envolve a sondagem da opinião de uma amostra de observadores humanos e constitui outra contribuição original da pesquisa realizada. A motivação para sua concepção residiu na escassez, no âmbito da área de renderização não fotorrealística, de metodologias consolidadas para a validação de abordagens destinadas à geração de abstrações não fotorrealísticas, em especial aquelas destinadas à avaliação de abstrações de vídeos digitais.

A revisão bibliográfica referente à avaliação subjetiva na área de NPR encontra-se na Seção 2.3 do Capítulo 2 (p. 60), na qual foram destacados alguns aspectos concernentes ao processo de avaliação, tais como (i) a definição da quantidade de participantes do experimento, (ii) a definição do tipo de instrumento a ser utilizado no processo de avaliação e (iii) a definição do tipo de classificação a ser adotada no instrumento de avaliação escolhido, dentre outros. Destaca-se que a proposição da metodologia apresentada neste capítulo também se fundamentou em adaptações de recomendações e etapas presentes no padrão ITU-R BT.500-12 (2012), assim como em critérios definidos para a avaliação de imagens renderizadas ao estilo de cartuns, derivados das pesquisas de Ianeva (2003), Medhi, Patabandhige e Toyama (2007) e Wallraven et al. (2007).

A metodologia proposta neste capítulo é composta por cinco etapas, fundamentadas na revisão da literatura da área, a saber: (i) planejamento do experimento; (ii) elaboração do material de teste; (iii) condução do ensaio e coleta de dados; (iv) tabulação dos dados coletados; e (v) apresentação dos resultados obtidos.

Antes de direcionar a discussão para o contexto da NPR, vale a pena resgatar a

dualidade quantitativa/qualitativa dos enfoques de avaliação na pesquisa científica. Expõe-se que a adoção de um enfoque quantitativo de avaliação pode conferir (i) maior eficácia ao processo de investigação de hipóteses (objetivando sua aceitação/rejeição) e (ii) robustez em relação à mensuração de aspectos descritivos de processos investigados, assim como oferecer graus de confiança e eficácia possíveis de serem comprovados de forma diretamente objetiva. Entretanto, o enfoque quantitativo pode afigurar-se limitado, uma vez que, usualmente, há fatores não verificáveis ou explicáveis a partir da redução de eventos observados a valores de uma variável de pesquisa considerada (ROBSON et al., 2001).

Por outro lado, a adoção de um enfoque qualitativo de avaliação permite o aprofundamento do investigador na exploração de elementos afetivos, emocionais e de preferências diretas, relativos à qualidade do evento investigado. Não obstante, a adoção de um enfoque qualitativo permite a investigação de significados e explicações relativas aos dados coletados. Contudo, não se pode deixar de considerar as dificuldades inerentes a esta categoria de enfoque, e.g., a dificuldade maior na comprovação da veracidade das conclusões obtidas. É um fato que enfoques qualitativos podem desempenhar um papel importante em processos de avaliação na pesquisa científica, realçando informações com amplitude e profundidade impraticáveis de obter a partir de um enfoque quantitativo (ROBSON et al., 2001).

É conveniente ressaltar ainda que ambos os enfoques considerados afiguram-se complementares e, assim o sendo, focalizam aspectos diferentes do mesmo contexto de pesquisa (YAUCH; STEUDEL, 2003). Por exemplo, a administração de entrevistas ou de questionários estruturados pode elucidar aspectos relativos à aceitação de produtos e a preferências do usuário que não se consegue a partir da mensuração do desempenho fundamentada em métricas objetivas. Adicionalmente, informações coletadas por meio de enfoques qualitativos, na fase de planejamento do processo de avaliação, podem fundamentar a concepção de instrumentos de medição quantitativos de maior relevância à pesquisa considerada.

No tocante à dualidade qualitativo-quantitativo, Creswell (2013) afirmou que a utilização de enfoques mistos, nos quais o pesquisador mescla dados quantitativos e qualitativos, podem permitir análises mais abrangentes do problema de pesquisa considerado, havendo a integração e interpretação global das informações coletadas.

A partir da revisão da literatura da área de NPR, sumariada no Quadro 2.2 (ver

Seção 2.3, p. 62), constatou-se a prevalência de enfoques qualitativos nos processos de avaliação de abordagens NPR²³. Tal constatação reflete tão somente o fato de que processos de avaliação de renderizações NPR devem considerar contextos nos quais estão presentes fatores *psicológicos* (e.g., evocação de emoções/sentimentos, cognição), *neurofisiológicos* (e.g., fisiologia do olho, percepção cromática, atenção visual, comparação visual) e *estéticos* (e.g., apelo estético, qualidade estética), dentre outros. Além desses fatores, também é necessário considerar o propósito das renderizações a serem avaliadas, e.g., a propaganda, o *marketing* e/ou o entretenimento.

Podem ser enumeradas dificuldades adicionais, relativas à utilização de um enfoque quantitativo no tocante à avaliação de renderizações NPR, as quais se devem sobretudo: (i) à inexistência de uma metodologia de avaliação que seja consensual à comunidade de NPR; (ii) à ausência de bases de imagens de referência (*ground-truth*) de renderizações estilísticas; e (iii) ao caráter inerente e eminentemente subjetivo do processo de avaliação desta natureza.

Hertzmann (2010) e Kyprianidis et al. (2013) afirmaram que há pouca pesquisa direcionada à apresentação de metodologias estruturadas destinadas à avaliação de técnicas de NPR, acrescentando que a definição de tais metodologias tornou-se importante a partir do avanço das pesquisas na área de NPR, em especial aquelas relativas à reprodução e à automatização de estilos tradicionais de renderização artística.

A falta de sistematização no processo de avaliação de renderizações NPR, percebida em grande parte das pesquisas realizadas, a partir das descrições encontradas na literatura da área, durante a etapa de revisão bibliográfica, motivou a proposição da metodologia de avaliação descrita neste capítulo.

Decidiu-se adotar um enfoque de avaliação misto, no qual a componente qualitativa traduz-se na avaliação de aspectos subjetivos associados à utilização de renderizações NPR, e.g., a preferência visual direta pela renderização apresentada ou, ainda, a preferência visual pelas renderizações que apresentam regiões de cor uniforme, ressaltadas e delimitadas por contornos escuros. Visando a contornar a subjetividade inerente ao enfoque qualitativo, o questionamento ao respondente fundamentou-se no emprego de escalas de Likert de semântica diferencial.

²³ Veja-se, por exemplo, as descrições dos processos de avaliação apresentadas por Mould, Mandryk e Li (2012), Mantiuk, Tomaszewska e Mantiuk (2012) e Li, Mould e Davies (2013).

Posteriormente, a quantificação da opinião do participante do experimento possibilitou o processamento estatístico dos dados coletados, o qual resultou em uma síntese numérica dos aspectos de pesquisa considerados. Mediante o referido processamento, procurou-se verificar diferenças estatisticamente significativas entre as abordagens comparadas.

A metodologia de avaliação proposta também considera aspectos descritivos, mensurados por meio da administração de um questionário pós-teste, no qual aspectos relativos ao contexto de utilização de renderizações NPR podem ser elucidados. Adicionalmente, são investigados aspectos relativos à configuração de parâmetros das abordagens avaliadas, a fim de evitar vieses no processo comparativo.

O presente capítulo é composto por três seções distintas, as quais contêm, respectivamente:

- Seção 4.1: a etapa de planejamento do experimento, na qual são detalhados os critérios a serem adotados no experimento, os vídeos a serem exibidos e as abordagens a serem comparadas, além de serem detalhadas as características dos participantes do experimento;

- Seção 4.2: a etapa de elaboração do material de teste, o qual abrange a elaboração dos questionários, planejamento do processo de votação e a caracterização da sessão experimental; e

- Seção 4.3: aspectos a serem observados durante o processo de análise estatística dos resultados experimentais obtidos.

4.1 Planejamento do Experimento

Nesta seção, apresenta-se a etapa de planejamento do experimento, constituída por quatro subetapas: (i) seleção dos critérios a serem adotados no experimento; (ii) seleção dos vídeos a serem utilizados no experimento; (iii) seleção das abordagens a serem comparadas e configuração de seus parâmetros; e (iv) caracterização dos participantes do experimento.

4.1.1 Seleção dos Critérios a serem Adotados no Experimento

A seleção dos critérios a serem considerados no processo de avaliação foram definidos a partir da pesquisa bibliográfica realizada, em especial nas pesquisas de Ianeva, Vries e Rohrig (2003), Medhi, Patabandhige e Toyama (2007), Wallraven et al. (2007) e

Alencar, Queiroz e Gomes (2013), de modo a se alinhar aos objetivos de pesquisa pré-definidos.

Ianeva, Vries e Rohrig (2003) definiram cartuns como imagens compostas por uma quantidade reduzida de cores, predominantemente cores primárias, as quais são utilizadas na composição de grandes regiões uniformes, delimitadas por bordas pretas, podendo conter textos, e.g., balões de diálogo.

Medhi, Patabandhige e Toyama (2007) e Alencar, Queiroz e Gomes (2013) corroboraram as definições de Ianeva, Vries e Rohrig (2003), afirmando que: (i) cartuns devem conter regiões de cor uniforme, ressaltadas e delimitadas por contornos escuros; (ii) o tema da cena renderizada não deve apresentar falhas ou imperfeições graves, e.g., partes faltantes; (iii) o conteúdo semântico da imagem original deve ser reconhecido na imagem renderizada; (iv) as regiões de cor devem apresentar saturação elevada; (v) o fundo da cena deve ser visto de forma simplificada e com poucos detalhes; (iv) o cartum deve manter um aspecto visual agradável.

A qualificação *agradável* dada ao aspecto visual do cartum está relacionada à preferência visual direta. Wallraven et al. (2007) focalizaram a avaliação deste aspecto, tendo comparado imagens de faces estilizadas, por meio de um processo de votação que incluiu os seguintes aspectos: (i) *preferência direta*, associado à escolha da imagem que capturava melhor a essência da expressão facial; (ii) *facilidade de reconhecimento*, associado ao acréscimo/decrécimo da facilidade/velocidade de reconhecimento da face estilizada, quando comparada à versão original; e (iii) *intensidade e sinceridade* da expressão facial, associadas ao sentimento evocado pela visualização da imagem. Adicionalmente, Wallraven et al. (2007) administraram três questionários, visando a avaliar o aspecto estético, a efetividade da ilustração e a preferência subjetiva à estilização retratada.

No contexto da caracterização de renderizações ao estilo de cartuns, Zhang et al. (2011) afirmaram que vídeos renderizados ao estilo de cartuns apresentam um conteúdo retratado de forma simplificada, contêm regiões definidas e delineadas por contornos escuros, caracterizados por um estilo de cor e sombreamento bem definidos e apresentam coerência temporal intraquadros.

AlMeraj et al. (2009) corroboraram as definições e conceitos expostos anteriormente, ao verificar a qualidade subjetiva de resultados obtidos a partir da utilização de uma abordagem para a estilização NPR, a qual imitava o movimento do

braço humano ao desenhar linhas. A avaliação da qualidade da abordagem proposta pelos autores foi realizada em um experimento no qual o participante deveria: (i) indicar se a imagem renderizada foi produzida por um ser humano ou gerada automaticamente por uma aplicação computacional; (ii) classificar as imagens de acordo com a preferência direta, priorizando aquelas que se afiguravam mais atrativas, do ponto de vista subjetivo; (iii) elencar contextos de uso para as imagens classificadas na etapa (ii); (iv) apontar, dentre as imagens classificadas na etapa (ii), aquelas que poderiam ser utilizadas em livros didáticos universitários; e (v) apontar, dentre as imagens classificadas na etapa (ii), aquelas que poderiam ser utilizadas em livros didáticos infantis.

Diante do exposto, resumiam-se, a seguir, os critérios definidos após a condução da pesquisa bibliográfica relativa à renderização de vídeos digitais ao estilo de cartum:

(i) *Simplicidade*: o vídeo abstraído deve conter áreas uniformemente coloridas, delimitadas por contornos escuros;

(ii) *Preservação do conteúdo*: o tema do vídeo não deve apresentar falhas ou partes faltantes;

(iii) *Semelhança ao conteúdo original*: o conteúdo semântico do vídeo original deve ser reconhecível no vídeo abstraído;

(iv) *Nitidez*: o tema do vídeo deve ser apresentado de forma nítida no vídeo abstraído;

(v) *Coerência temporal*: o vídeo abstraído não deve apresentar cintilação (*flickering*) ou artefatos espúrios;

(vi) *Preferência visual direta*: o vídeo abstraído deve manter um aspecto visual agradável; e

(vii) *Contexto de uso*: o processo de avaliação deve evidenciar contextos de uso das renderizações apresentadas.

No processo de avaliação proposto, o participante do experimento deve atribuir um score, em uma escala de Likert de 5 pontos, com semântica diferencial (*Discordo totalmente, Discordo parcialmente, Não concordo nem discordo, Concordo parcialmente e Concordo totalmente*), referente ao grau de concordância às afirmações apresentadas. Mesmo diante da discordância registrada na literatura (e.g., COLMAN; NORRIS; PRESTON, 1997; WINTER; DODOU, 2010; DOLNICAR et al., 2011) acerca da

adoção de escalas de Likert de 5 ou de 7 pontos, a escolha de escalas de 5 pontos deu-se visando a consonância às metodologias de avaliação pesquisadas (ver Quadro 2.2, Seção 2.3, p. 62) e às recomendações do padrão ITU-R BT.500-12 (2012), para fins de comparação.

Durante a concepção da metodologia de avaliação, verificou-se que se afiguraria extenuante para o participante a votação em sete escalas diferentes, relativas aos aspectos supramencionados, assim como a votação nessas escalas para cada vídeo avaliado (totalizando a votação em 14 escalas, i.e., 7 escalas x 2 vídeos expostos lado-a-lado). Assim sendo, os critérios utilizados no processo de avaliação foram condensados em três afirmações:

- (i) O assunto do vídeo é preservado com nitidez por áreas uniformemente coloridas, delineadas por contornos escuros;
- (ii) O vídeo apresenta cintilação ou artefatos espúrios;
- (iii) O vídeo apresenta aspecto visual agradável;

Após a definição dos critérios a serem considerados pelo participante do experimento, foi realizada a seleção dos vídeos a serem utilizados no experimento.

4.1.2 Seleção dos Vídeos a serem Utilizados no Experimento

Devido à escassez de métodos destinados à avaliação de abordagens de NPR constatada durante a revisão bibliográfica (vide Seção 2.3, p. 60), decidiu-se recorrer a pesquisas na área correlata de padrões de testes de *codecs* de vídeos. Destaca-se, neste sentido, o padrão ITU-R BT.500-12 (*Methodology for the Subjective Assessment of the Quality of Television Pictures*) que versa, dentre outros aspectos, sobre testes de *codecs* de vídeo para sinais de televisão, no qual há a caracterização do sinal original dos vídeos de referência. De acordo com o padrão, os vídeos de referência são aqueles utilizados no processo de comparação aos vídeos processados após a aplicação do *codec* em teste²⁴.

O referido padrão enfatiza a observação da alta qualidade do sinal de entrada, referindo-se à ausência de defeitos visíveis neste sinal, preferivelmente armazenado de forma digital (pela conveniência no armazenamento e incremento na reprodutibilidade dos vídeos), utilizando-se um método de compressão sem perdas

²⁴ Os vídeos processados com a utilização do *codec* em teste são comumente nomeados de *impaired videos*. O vídeo obtido após tal processamento pode apresentar defeitos/deficiências ou falhas após a compressão. O termo *impairment* relaciona-se ao processamento e a percepção de defeitos/falhas nos vídeos processados.

(*lossless compression*). De acordo com o padrão, tais aspectos são cruciais para a obtenção de resultados confiáveis, ao serem realizadas análises de *codecs*.

Destaca-se outra recomendação presente no padrão ITU-R BT.500-12 (2012), alusiva à verificação da qualidade adequada do sinal de entrada do experimento a ser realizado. A qualidade adequada refere-se às dimensões espaciais do vídeo de entrada, as quais devem evidenciar diferenças entre os *codecs* a serem comparados.

Segundo o padrão ITU-R BT.500-12 (2012), o conteúdo do vídeo deve ser pertinente ao teste a ser realizado, i.e., as cenas retratadas nos vídeos devem conter cenários típicos de uso da renderização ou codificação a ser avaliada. A escolha do tipo e do conteúdo do vídeo impacta as inferências e interpretações formuladas sobre os resultados do experimento.

Diante do exposto, foram realizadas pesquisas relativas à existência de bancos de dados de vídeos a serem utilizados no processo de avaliação de técnicas de renderização não fotorrealística. Verificou-se que o vídeo apresentado na pesquisa seminal em renderização não fotorrealística de vídeos, realizada por Wang et al. (2004), intitulado *monkeybar*, esteve presente em pesquisas subsequentes, de modo que esse também foi incluído no conjunto de vídeos selecionados para o experimento realizado nesta pesquisa. O processo de definição de outros vídeos foi realizado por meio da pesquisa em bancos de dados de vídeos utilizados comumente na (i) avaliação de *codecs* de vídeos e (ii) segmentação de vídeos, duas áreas correlatas ao processo de avaliação de renderizações estilísticas, uma vez que algoritmos de segmentação têm sido utilizados em abordagens de NPR.

Na área de avaliação de *codecs*, a *Xiph.Org Foundation* destacou-se por realizar pesquisas e tutelar uma coletânea de projetos relacionados a produtos multimídia, os quais incluem a fundamentação e o planejamento de padrões de *codecs* de áudio e vídeo a serem disponibilizados no formato de implementações computacionais de código aberto e de domínio público.

Dentre as coleções de vídeo disponibilizadas pela *Xiph.Org Foundation* para a comparação de *codecs*, destaca-se a *derf's collection*, que compreende 81 vídeos, caracterizados por conteúdo, resolução espacial e formato de codificação variado, cujas sequências presentes são comumente utilizadas na avaliação de *codecs* destinados à compressão de vídeos. Os vídeos foram digitalizados a partir de uma fonte de sinal analógico, o qual foi encapsulado em arquivos no formato YUV4MPEG

(extensão de arquivo *.y4m*), codificado em um formato de compressão sem perdas. Tais vídeos estão disponíveis em servidores com acesso público, acessíveis sem custos de licença.

No contexto da avaliação de *codecs* de vídeos, o GML (*Graphics and Media Lab*) da Universidade de Moscou (*Moscow State University - MSU*) disponibiliza, anualmente, um relatório contendo um estudo comparativo entre *codecs* de compressão de vídeos, e.g., o MSU MPEG-4 AVC/H.264 *Video Codecs Comparison*. O GML disponibilizou uma aplicação de *software* para a avaliação da percepção visual de vídeos digitais. A referida metodologia de avaliação alinha-se às recomendações do padrão ITU-R BT.500-12 (2012). Adicionalmente, verificou-se que no estudo comparativo do GML (2013), foram utilizados alguns vídeos disponíveis na *derf's collection*, fato que ressalta a importância desta coleção de vídeos para a pesquisa ora documentada.

Adicionalmente, Wang e Collomosse²⁵ (2012) apresentaram uma técnica para a segmentação de vídeos de forma temporalmente coerente, no qual são verificadas melhorias quantitativas (utilizando-se a metodologia descrita por Martin et al., 2001) e qualitativas, relativas à precisão da segmentação em relação a outras técnicas que figuram no estado da arte. Tais melhorias foram evidenciadas após a realização de experimentos nos quais foram utilizados 9 vídeos, os quais possuíam características a serem consideradas no processo de segmentação, e.g., a presença de movimentos ágeis na cena ou presença de graus variados de oclusão. Os autores agruparam os vídeos de acordo com a quantidade de oclusão e a quantidade de movimento visualmente perceptíveis. A presença de tais características, distintas daquelas apresentadas nos vídeos da *derf's collection*, foram consideradas no processo de avaliação da abordagem proposta. Assim sendo, um excerto do conjunto utilizado por Wang e Collomosse (2012) foi incluído ao conjunto de vídeos a serem utilizados no processo de avaliação.

Os vídeos presentes na *derf's collection* totalizam, conjuntamente àqueles considerados na pesquisa de Wang e Collomosse (2012), um conjunto de 90 vídeos, dos quais 10 compuseram o conjunto final utilizado no experimento, após uma pré-seleção de 20. Os quadros iniciais de cada um dos vídeos do conjunto final são apresentados na Figura 4.1.

²⁵ Página do projeto disponível em: <http://personal.ee.surrey.ac.uk/Personal/Tinghuai.Wang/TMM2011.html>

Figura 4.1 – Primeiro quadro de cada vídeo utilizado no processo de avaliação.



A seguir, apresentam-se informações adicionais que justificam a seleção de cada um dos vídeos utilizados no experimento.

Cowgirl (Figura 4.1A): o tema central do vídeo é a apresentação de uma vaqueira em um cenário com cores pálidas, com tema do vídeo contrastando com o fundo da cena. Há pouco movimento do assunto do vídeo, há oclusão de parte dos objetos da cena, há trepidações e tremulações no movimento da câmera. Vídeo presente na pesquisa realizada por Wang e Collomosse (2012).

Crew (Figura 4.1B): o tema central do vídeo é a apresentação de uma equipe da NASA, com presença do assunto do vídeo com roupas com cores vibrantes e que se destacam do fundo cinza da cena. Há a presença de luz de *flashes* fotográficos, que podem ser confundidos com cintilação (*flickering*) após o processo de renderização. Há presença de muito movimento na cena. Vídeo presente na *derf's collection*.

Deadline (Figura 4.1C): o tema central do vídeo é um apresentador, o qual é pouco destacado em relação ao fundo da cena. Neste vídeo, a câmera apresenta-se fixa, há pouca movimentação em todo o decurso do vídeo. Vídeo incluído com o intuito de avaliar a coerência temporal intraquadros. Observa-se a presença de objetos coloridos por toda a cena, em contraste aos vídeos com fundo monotônico ou com poucas cores ao fundo da cena. Vídeo apresenta baixa resolução espacial. Vídeo presente na *derf's collection*.

Flower Garden (Figura 4.1D): o tema central do vídeo é a apresentação de um jardim, contendo elementos com textura natural, incluindo flores e árvores. Nota-se uma paisagem natural com pouca movimentação da cena e parte oclusa ao final do vídeo. Vídeo selecionado pela dificuldade da renderização de cenas com regiões que contém muita textura. Vídeo presente tanto na *derf's collection* quanto na pesquisa realizada por Wang e Collomosse (2012).

Ice (Figura 4.1E): o tema central do vídeo é a apresentação de uma sequência de pessoas patinando no gelo, com pouco movimento, poucas cores marcantes e apresentação do fundo monotônico. Neste vídeo, o fundo da cena e a câmera encontram-se parados, havendo pouca movimentação do tema. O vídeo foi incluído por apresentar uma cena típica com a presença de várias pessoas, em contraste a apresentação de uma cena de natureza. Objetos e assunto do vídeo apresentam poucas regiões com texturas complexas visíveis, e com as faces presentes apresentando baixa definição de elementos-chave (pouca definição de olhos, nariz, boca). Vídeo presente na *derf's collection*.

Monkeybar (Figura 4.1F): o tema central do vídeo é apresentação de duas crianças brincando em um parque, no brinquedo da barra horizontal. Este vídeo foi apresentado na pesquisa seminal destinada à renderização não fotorrealística de vídeos desenvolvida por Wang et al. (2004). Há a presença de movimento do assunto do vídeo, a câmera permanece estática, há oclusão severa de vários outros elementos, tais quais as árvores ao fundo da cena. O vídeo está presente na pesquisa realizada por Wang e Collomosse (2012).

Skateboard (Figura 4.1G): o tema central do vídeo é a apresentação de um

skatista. Vídeo de curta duração, no qual são apresentadas cores pálidas, no qual também está presente grande quantidade de movimento e pouca oclusão. Há a presença de regiões texturizadas, tais quais os arbustos. Vídeo apresentado na pesquisa realizada por Wang e Collomosse (2012).

Walkdog (Figura 4.1H): o tema central do vídeo é a apresentação de uma caminhada de duas pessoas e um cachorro. Neste vídeo, os objetos são apresentados com oclusão severa e pouco movimento dos personagens. São apresentadas regiões texturizadas, incluindo água e pelos do cachorro. O vídeo apresenta uma gama variada de cores, regiões desfocadas ao fundo da cena, com as faces presentes bem delineadas. Vídeo presente na pesquisa realizada por Wang e Collomosse (2012).

Waterfall (Figura 4.1I): o tema central do vídeo é a apresentação de uma cachoeira, no qual estão presentes regiões de texturas da natureza, pouco movimento, sem oclusão visível. Vídeo curto e com baixa resolução, com cores pouco marcantes. Apresenta desafios na renderização por apresentar água e região de floresta com muita textura. A renderização NPR deste tipo de cena pode resultar em homogeneização excessiva das regiões. Vídeo presente na *derf's collection*.

Younakim (Figura 4.1J): o tema central do vídeo é a apresentação de uma patinadora em uma competição em olimpíada. Neste vídeo, o fundo da cena apresenta-se homogêneo, havendo presença de muito movimento do tema. Pode haver dificuldade na renderização da cena devido ao movimento rápido do assunto, existindo a possibilidade de aparecimento de artefatos de cintilação (*flickering*) após a renderização. Presença de cores fortes apenas no assunto da cena, com o fundo da cena apresentando regiões que podem ser confundidos com artefatos espúrios são aspectos notórios neste vídeo. Vídeo presente na pesquisa realizada por Wang e Collomosse (2012).

Outros vídeos apresentados na pesquisa realizada por Wang e Collomosse (2012), tais como os vídeos intitulados *boy*, *dance* e *baseball*, não foram incluídos no processo de avaliação devido a alguns fatores: (i) a resolução espacial inadequada vídeos apresentados, com dimensões inferiores a 300x300 pixels; (ii) a presença de artefatos no vídeo original, tais quais a presença de ruído ou marca d'água visível; (iii) vídeo com pouco movimento dos objetos do assunto do vídeo (e.g., *galleon*, *harbour*); (iv) a presença de uma paleta de cores com pouca variabilidade e com pouca diferenciação entre o fundo e o assunto do vídeo.

Destacam-se a seguir outras características dos vídeos selecionados (e.g., a resolução espacial e a quantidade de quadros presentes no vídeo), as quais são

sumariadas na Tabela 4.1. A duração média dos vídeos selecionados é de 14,2 segundos, o que impactou a duração da sessão do experimento. A seleção de vídeos de longa duração aumentaria a fadiga do participante, enquanto a seleção de vídeos de curta duração (tipicamente, menores que 5 segundos) acarretaria a repetição exaustiva do vídeo sob análise. Além disto, a curta duração do vídeo poderia impactar a percepção de diferenças no processo comparativo dos resultados produzidos pelas abordagens consideradas.

Tabela 4.1 – Sumário das características dos vídeos selecionados.

Título do vídeo	Duração do vídeo	Dimensões (em pixels)	Quantidade de quadros	Quadros por segundo
<i>Cowgirl</i>	08 segundos	480x352	224	30
<i>Crew</i>	10 segundos	704x576	600	60
<i>Deadline</i>	45 segundos	352x288	1374	30
<i>Flower garden</i>	12 segundos	720x486	360	30
<i>Ice skating</i>	08 segundos	704x576	600	60
<i>Monkeybar</i>	13 segundos	850x634	411	30
<i>Skatingboard</i>	05 segundos	480x352	146	30
<i>Walkdog</i>	11 segundos	480x352	300	30
<i>Waterfall</i>	08 segundos	352x288	260	30
<i>Yunakim</i>	22 segundos	544x304	560	30

À etapa de definição dos vídeos a serem utilizados no processo de avaliação, sucedeu a etapa de seleção de abordagens a serem comparadas, descrita na próxima subseção.

4.1.3 Seleção das Abordagens a serem Comparadas

A classificação taxonômica proposta por Kyprianidis et al. (2013), apresentada no Capítulo 2 (vide Figura 2.1, p. 32), possibilita a delimitação do escopo de pesquisas relativas à NPR. A metodologia de avaliação apresentada neste capítulo utiliza tal delimitação no processo comparativo das abordagens, i.e., recomenda-se que o processo de comparação seja realizado entre abordagens pertencentes ao mesmo grupo taxonômico.

No tocante ao grupo taxonômico ao qual se enquadram as abordagens a serem comparadas, recomenda-se que sejam investigadas pesquisas posteriores àquela realizada por Kyprianidis et al. (2013). Recomenda-se a verificação da viabilidade na condução de um processo comparativo entre todas as abordagens listadas no grupo taxonômico considerado, de modo a propor restrições à quantidade de técnicas relacionadas. Sugere-se, ainda, que as abordagens comparadas se enquadrem no subgrupo que utiliza técnicas de processamento semelhantes. Adicionalmente, a

escolha das abordagens a serem comparadas deve considerar a semelhança visual entre resultados.

Salienta-se que a comparação de abordagens fundamentadas em princípios de renderização diferentes poderia enviesar as evidências coletadas sendo, pois, desaconselhadas, e.g., a comparação entre abordagens destinadas à renderização de mosaicos e ladrilhos e à estilização do movimento.

Após a seleção das abordagens a serem comparadas, recomenda-se a verificação dos aspectos específicos de cada uma delas. A verificação destes aspectos deve incluir: (i) a motivação para sua presença no processo de avaliação; (ii) os aspectos relativos a sua implementação computacional, e.g., a presença ou ausência de interface gráfica ou disponibilidade do código-fonte; e (iii) o processo de configuração de seus parâmetros.

A configuração de parâmetros deve ser realizada de duas formas: (i) o parâmetro é configurado de acordo com a indicação do proponente da abordagem; e (ii) o parâmetro encontra-se configurado para um valor padrão na implementação computacional disponibilizada pelo proponente da abordagem.

Após um processo de configuração preliminar, procede-se à modificação dos valores dos parâmetros cuja alteração impacta, visualmente, o resultado da estilização. Esta etapa de pré-seleção de valores deve ser fundamentada em experimentos subjetivos, realizados sobre imagens ou vídeos de referência na área de NPR. A validação das configurações de parâmetros pré-selecionados é realizada em uma etapa posterior, a partir da adoção do grupo de valores dos parâmetros mais bem avaliados experimentalmente.

Após a definição das abordagens a serem comparadas e ao processo de configuração dos parâmetros, prossegue-se à etapa de seleção da amostra populacional.

4.1.4 Caracterização dos Participantes do Experimento

Tendo conduzido um experimento comparativo de técnicas destinadas ao processamento de imagens, Mantiuk, Tomaszewska e Mantiuk (2012) teceram considerações a respeito do número de participantes e de sua caracterização. Os autores afirmaram que a opinião sobre a qualidade de imagens pode estar relacionada à representatividade da amostra considerada, uma vez que isto possibilita a investigação de padrões de comportamento e limita as inferências formuladas sobre o processo de votação. Adicionalmente, o equipamento utilizado durante a condução do

experimento e o nível de controle do ambiente (e.g., luminosidade ambiente, dimensões da tela, contraste e resolução utilizada na sessão de testes, calibração das cores da tela) pode interferir no resultado do experimento.

A observação de dois fatores no Quadro 2.2 (ver Seção 2.3, p. 62), concernentes ao tamanho e características da amostra considerada, permite constatar que não há consenso na literatura da área a esse respeito. Há experimentos realizados com amostras que variam de 10 (WINNEMÖLLER, 2006) a 74 participantes (SANTELLA; DECARLO, 2004). Adicionalmente, é possível verificar a variabilidade dos perfis dos participantes considerados, e.g., arquitetos, estudantes de arquitetura, estudantes de curso de graduação, estudantes de curso de pós-graduação, pesquisadores em geral, pesquisadores da área de NPR, artistas.

Diante desta variabilidade, frisa-se que, no âmbito da Estatística Indutiva, a delimitação do tamanho da amostra apresenta-se como fator essencial a ser observado. O tamanho da amostra deve ser considerado, no tocante à generalização das conclusões e aos aspectos práticos do processo de avaliação (e.g., dificuldade no recrutamento de participantes, custos, tempo associado). No tocante a tais restrições, considera-se que: (i) a aproximação relativa aumenta à medida que cresce o número de determinações, segundo a *lei dos grandes números* (CASELLA; BERGER, 2010); (ii) as pesquisas científicas estão inseridas em contextos nos quais a população é usualmente de tamanho considerável; e (iii) há dificuldades no recrutamento de indivíduos cuja participação esteja condicionada à remuneração ou tenham restrições de tempo na participação em experimentos.

Após a constatação destas restrições, percebe-se que determinação do tamanho apropriado de uma amostra afigura-se procedimento não trivial. Assim, a escolha da amostra deve incluir um número suficiente e representativo de casos, escolhidos aleatoriamente, que possibilite o oferecimento de segurança estatística aceitável em relação à representatividade dos dados.

A escolha do tamanho da amostra pode ser delimitada em, no mínimo, 25 ou 30 observações (HOGG; TANIS, 1993). Considera-se como *pequena* amostra aquela de dimensões inferiores a 30 observações, enquanto se considera como *grande* amostra aquela cujas dimensões excedem as referidas dimensões. Salienta-se, ainda, que o processo de inferências obtido a partir da utilização de uma pequena amostra, em populações cuja distribuição de probabilidade aproxima-se da distribuição normal, é propensa a se ajustar à distribuição t de *Student* (LEVINE; BERENSON; STEPHAN, 2000).

Há que se considerar que parte das pesquisas relatadas na área de NPR pode apresentar decisões questionáveis quanto à etapa de avaliação, no tocante à quantidade e aos perfis dos participantes, assim como ao número de imagens/vídeos utilizados.

Mesmo não havendo consenso, na área de NPR, no tocante à quantidade de participantes a serem considerados no processo de avaliação, recomenda-se implicitamente que, na realização do experimento, seja considerada uma grande amostra, conforme atestam avaliações realizadas em pesquisas recentes na área de NPR, e.g., Seshadrinathan et al. (2010), Mould, Mandryk e Li (2012) e Li, Mould e Davies (2013). Recomenda-se também o recrutamento de participantes que apresentem familiaridade com conceitos artísticos, i.e., que tenham familiaridade com o ato de desenhar e conheçam as definições de traço, linha de silhueta e estilização artística.

Após a caracterização da amostra populacional, prossegue-se à elaboração do material a ser utilizado durante os ensaios experimentais.

4.2 Elaboração do Material de Teste e Condução da Sessão Experimental

Nesta seção, são apresentados aspectos relativos à elaboração do material de teste, o qual inclui a elaboração dos instrumentos de sondagem a serem empregados pelo avaliador (e.g., questionários, entrevistas), o planejamento do processo de sondagem (e.g., votação e a caracterização da sessão experimental).

No padrão ITU-R BT.500-12 (2012), há recomendações a serem observadas durante a realização de uma sessão de testes: (i) o pesquisador deve ajustar o contraste e a resolução do monitor antes do início da sessão; (ii) as imagens ou vídeos devem ser apresentadas em uma sequência pseudoaleatória, de preferência em uma sequência diferenciada, entre sessões distintas; (iii) uma imagem ou vídeo não deve ser exibido em ocasiões sucessivas ou com o mesmo processamento; (iv) uma sessão de testes não deve ter duração superior a meia hora, incluindo o tempo relativo às explanações iniciais e tempo destinado ao treinamento para a realização do experimento; (v) sessões com duração superior a uma hora devem contemplar intervalos a cada meia hora; (vi) a sessão de testes poderá ser iniciada com uma imagem ou vídeo de treinamento, na qual são caracterizadas as abordagens a serem comparadas. Na pesquisa relatada nesta tese, a abordagem que produziu o vídeo da sessão de treinamento fundamentava-se no emprego do algoritmo *mean-shift* e na superposição do mapa de bordas produzido por filtragem DoG.

Na primeira etapa do experimento, a documentação pré-teste é apresentada ao

participante. Esta documentação contém orientações relativas à garantia de confidencialidade dos dados da sessão de votação, orientações relativas à natureza do experimento, à cessão dos dados e ao anonimato. Adicionalmente, o participante assina o Termo de Consentimento Livre e Esclarecido (TCLE), destinado a lhe informar sobre a pesquisa, de forma a lhe permitir a tomada de decisões justas e sem constrangimentos a respeito da participação no experimento, incluindo o abandono da sessão de testes a qualquer momento. A documentação relativa à esta primeira etapa é apresentada no Apêndice E.

Após a assinatura desses documentos, procede-se à segunda etapa do experimento, na qual o participante é encaminhado ao ambiente no qual a sessão de votação é conduzida. O ambiente de testes estruturado nesta pesquisa pode ser visualizado na Figura 4.2.

Figura 4.2 – Ambiente de testes.



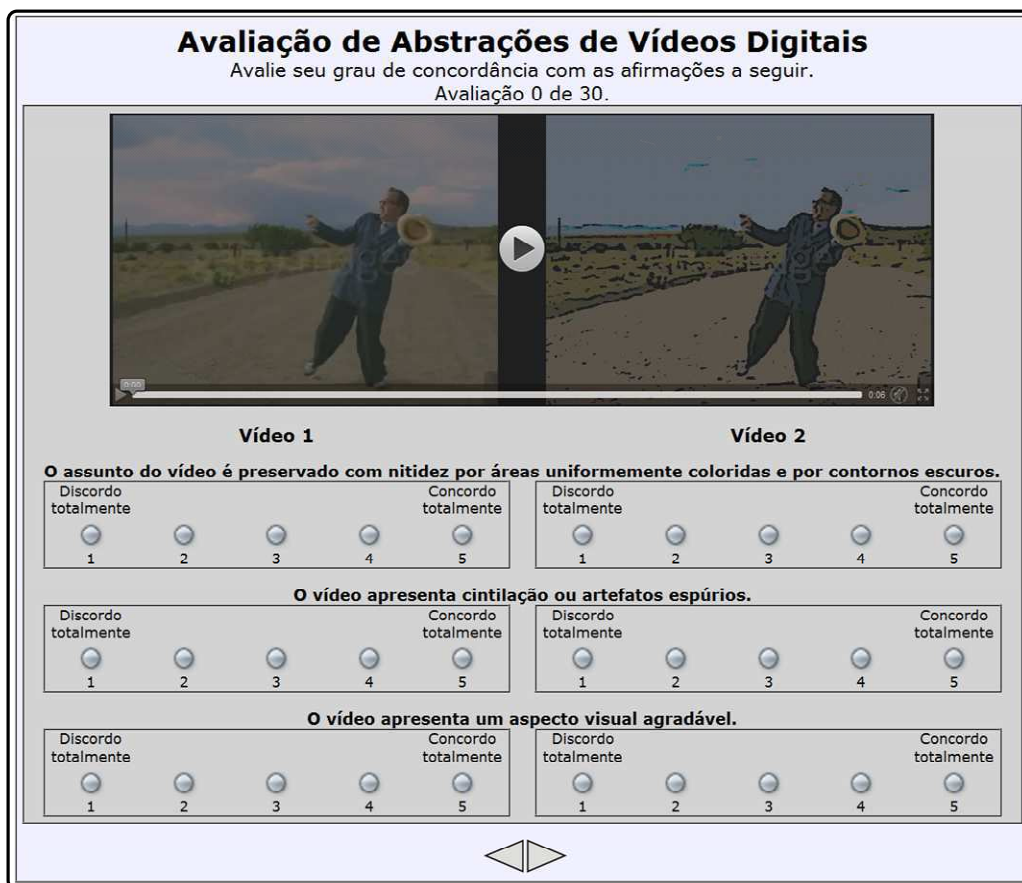
Antes do início dos testes, o participante deve ser instruído a configurar a altura da cadeira e do monitor de acordo com a preferência pessoal. Conforme destacado anteriormente, no padrão ITU-R BT.500-12 (2012) há uma recomendação relativa à realização de uma sessão de treinamento prévia (Figura 2.22 – Capítulo 2), com um intervalo que antecede o início da sessão de testes propriamente dita. Na tentativa de evitar vieses relativos ao entendimento dos conceitos apresentados aos participantes, o primeiro vídeo exibido ao participante deve ser renderizado mediante a utilização de uma abordagem diferente daquelas consideradas no processo comparativo.

O participante do experimento realiza a votação em uma interface *online* que apresenta dois vídeos, lado-a-lado, a concordância às afirmações apresentadas na

Seção 4.1.1 (p. 117), associadas a 6 escalas de Likert por comparação.

Após a exibição do vídeo de treinamento, a metodologia de votação e a interface do experimento são apresentadas ao participante. Na interface, não deve haver identificação de quaisquer elementos que possam deixá-lo a par da identificação da abordagem utilizada na renderização. Adicionalmente, deve ser ocultado o nome do vídeo e o nome da abordagem utilizada. A apresentação dos vídeos deve ser realizada de forma pseudoaleatória, sem preferência por quaisquer vídeos ou abordagens comparadas. A quantidade de vídeos apresentados e a quantidade de vezes que cada técnica é exibida devem ser balanceadas entre participantes, de forma equitativa. Um excerto de uma interface gráfica utilizada no experimento relatado nesta pesquisa é mostrado na Figura 4.3.

Figura 4.3 – Interface gráfica utilizada no experimento.



A terceira etapa do experimento consiste na administração do questionário pós-teste, relativo ao delineamento do perfil do participante, com perguntas relativas à idade, formação acadêmica e utilização de sistemas computacionais. Adicionalmente, propõe-se que o participante responda a duas perguntas de contexto livre. Na primeira pergunta, o participante é indagado sobre sugestões para o uso das renderizações dos vídeos mais bem avaliados durante a votação. Na segunda

pergunta, solicita-se ao usuário um relato contendo qualquer aspecto desagradável registrado durante a sessão de votação. O questionário pós-teste está contido no Apêndice F. Após a realização de um teste-piloto, o qual visou a validação do material elaborado, verificou-se que o tempo necessário para a realização do experimento atendia às recomendações do padrão ITU-R BT.500-12 (2012).

Após a elaboração do material de teste, deve haver a condução dos ensaios experimentais. Após a finalização de todas as sessões do experimento, prossegue-se à etapa de tabulação e análise dos resultados obtidos.

4.3 Aspectos relativos à Análise dos Resultados Experimentais

Nesta seção, são apresentados aspectos a serem observados durante o processo de análise estatística dos resultados experimentais obtidos.

O processo de avaliação, no contexto da geração de representações não fotorrealísticas, procura elucidar questões relativas à variabilidade de nuances de resultados de abordagens que visam ao apelo visual, uma vez que, como tal, é inerentemente ambíguo e subjetivo. Destaca-se que o processo de avaliação de abordagens NPR (ver Quadro 2.2, Seção 2.3, p. 62), assim como outros ensaios de avaliação, ocorre usualmente em uma única sessão por participante, fato justificado pelos custos associados ao planejamento e à condução do experimento, pela dificuldade encontrada no processo de recrutamento de participantes e pelo tempo necessário à realização do total de sessões de testes.

A proposição da metodologia de avaliação incluiu a revisão da literatura no tocante à análise inferencial dos dados coletados a partir do experimento. Mould, Mandryk e Li (2012) e Li, Mould e Davies (2013) validaram estatisticamente os resultados de seus experimentos de avaliação de NPR a partir do seguinte procedimento: (i) teste F ANOVA fator único (LEVINE; BERENSON; STEPHAN, 2000); e (ii) teste de comparação de médias. Na metodologia de avaliação proposta, as duas etapas do procedimento supradescrito são aplicadas a cada um dos três aspectos considerados. Adicionalmente, propõe-se que o processo avaliatório contemple três etapas:

(i) a avaliação intra-abordagem, i.e., a validação do conjunto de parâmetros considerado adequado a cada uma das abordagens. O questionamento a ser respondido por esta análise é: *considerando-se os três aspectos presentes na votação e o conjunto de valores dos parâmetros para uma determinada abordagem, qual foi o conjunto de valores mais bem avaliado?*

(ii) a avaliação interabordagem, i.e., a comparação visual quando consideradas abordagens diferentes. O questionamento a ser respondido por esta análise é: *considerando-se os três aspectos presentes na votação e a comparação entre as abordagens, qual foi a abordagem mais bem avaliada?*

(iii) a avaliação interabordagem quando considerado o melhor grupo de parâmetros intra-abordagem. O questionamento a ser respondido por esta análise é: *considerando-se os três aspectos presentes na votação e a comparação das abordagens com grupos de parâmetros mais bem avaliados, qual foi a abordagem mais bem avaliada?*

O teste F ANOVA fator único é utilizado por haver um único fator a ser testado: o grau de concordância do participante ao comparar duas renderizações de vídeos, considerando-se cada aspecto da avaliação. Uma vez que este teste evidencia apenas diferenças entre médias, não possibilitando comparações entre pares de grupos, utiliza-se, quando a análise estatística evidencia a possibilidade, o teste de Tukey-Kramer (LEVINE; BERENSON; STEPHAN, 2000).

A seguir, são apresentadas as considerações finais do capítulo.

4.4 Considerações Finais do Capítulo

Neste capítulo, foi apresentada uma metodologia de investigação experimental de abordagens destinadas à renderização NPR de vídeos. Após a condução da pesquisa relativa a metodologias de avaliação destas renderizações, constatou-se que:

- Na etapa de planejamento do experimento, os seguintes critérios de avaliação foram definidos: (i) o assunto do vídeo é preservado com nitidez por áreas uniformemente coloridas, delineadas por contornos escuros; (ii) o vídeo apresenta cintilação ou artefatos espúrios; (iii) o vídeo apresenta aspecto visual agradável. Ainda na etapa de planejamento do experimento, 10 vídeos foram selecionados para o conjunto final a ser utilizado no experimento, os quais podem ser categorizados de acordo com a quantidade de oclusão, quantidade de movimento visualmente perceptíveis ou presença/ausência de cores marcantes na cena. Frisa-se que a seleção de abordagens a serem comparadas, com a utilização dos critérios e dos vídeos supramencionados, deve contemplar aquelas fundamentadas em princípios de renderização semelhantes.

- A etapa de elaboração do material de testes contemplou: (i) a elaboração da documentação pré-teste (TCLE, termo de cessão de dados, termo de confidencialidade); (ii) a elaboração da interface gráfica utilizada no experimento; e

(iii) a elaboração do questionário pós-teste, relativo ao delineamento do perfil do participante do experimento.

- O processo de avaliação, realizado com usuários humanos, procura evidenciar diferenças associadas à estilização NPR produzida, quando considerados aspectos subjetivos de avaliação. As diferenças estatisticamente significativas podem ser obtidas a partir do processamento estatístico dos dados coletados, com a utilização de um teste F ANOVA fator único, seguido de um teste de comparação de médias. Estas diferenças devem considerar as avaliações intra-abordagem, interabordagem e interabordagem quando considerado o melhor grupo de parâmetros intra-abordagem.

O próximo capítulo contém a apresentação e a discussão dos experimentos realizados, os quais incluem a avaliação comparativa entre a abordagem proposta e outras abordagens concorrentes. A apresentação e discussão dos experimentos inclui os testes realizados em módulos da abordagem proposta.

Capítulo 5

Apresentação e Discussão dos Experimentos e Resultados

Neste capítulo, são apresentados os experimentos realizados visando à avaliação da abordagem proposta nesta pesquisa, destinada à conversão de vídeos em vídeos renderizados não fotorrealisticamente ao estilo de cartuns, descrita no Capítulo 3. São apresentados os resultados relativos ao processo de avaliação, o qual consistiu na verificação da capacidade da referida abordagem de atender aos objetivos da pesquisa. Adicionalmente, a avaliação objetivou a comparação dos resultados obtidos a partir da adoção da abordagem proposta e de outras abordagens concorrentes.

O presente capítulo é composto por duas seções distintas, nas quais são apresentados e discutidos:

- Seção 5.1: resultados experimentais relativos a quatro módulos que compõem a abordagem apresentada no Capítulo 3. Esta seção contém experimentos envolvendo a técnica de suavização por difusão anisotrópica, na qual são apresentados os resultados experimentais de três diferentes funções de parada-na-aresta. Adicionalmente, são apresentados resultados experimentais concernentes à técnica de filtragem com realce da coerência de regiões (*Coherence-Enhancing Filtering* - CEF) e, ainda, resultados concernentes à etapa de filtragem a partir da computação de médias não locais (*Non-Local Means* - NLN). Nesta seção, também são apresentados resultados visuais gerados a partir de valores de parâmetros adequados e inadequados aos objetivos da pesquisa. A validação de tais valores é apresentada na seção subsequente.

- Seção 5.2: os resultados obtidos no processo de avaliação de renderizações não fotorrealísticas de vídeos digitais. Os resultados das seguintes etapas são relatados e discutidos: (i) apresentação da configuração dos parâmetros das abordagens comparadas; (ii) apresentação do delineamento do perfil dos participantes do experimento; e (iii) apresentação dos resultados experimentais obtidos a partir do

esquema de votação adotado.

Conforme ressaltado no Capítulo 3, foram realizados outros experimentos relativos às abordagens de segmentação espaçotemporal de vídeos, com vistas à renderização não fotorrealística. Tais experimentos, relatados no Apêndice A, incluem técnicas destinadas à segmentação de quadros de vídeo e à computação do fluxo óptico.

Ao longo deste capítulo, menciona-se o tempo de execução associado ao processamento da aplicação de *software* referente a cada uma das técnicas consideradas. O *hardware* utilizado no processo de mensuração de tal tempo possuía um processador Intel Core i5 (x86-64, modelo M 460), com *clock* de 2,53 GHz, quatro núcleos de processamento e 4GB de memória principal, executando o sistema operacional Microsoft Windows 7 (versão 64 bits).

As três imagens utilizadas no Capítulo 3 (ver Figura 3.2, p. 95) também foram utilizadas no presente capítulo, com o intuito de ressaltar características visuais e apresentar resultados obtidos após a renderização realizada por cada abordagem pesquisada. Tais imagens estão presentes na literatura revisada, razão pela qual sua apresentação visou a facilitar o relato dos experimentos realizados. Ao final do capítulo, são apresentados *hyperlinks* relativos aos resultados de todos os experimentos realizados.

5.1 Experimentos envolvendo a Abordagem Proposta

Ao longo da pesquisa realizada foram definidos quatro aspectos cruciais a serem experimentados, visando à validação da abordagem proposta, a saber: (i) experimentos relativos à alteração da função de parada-na-aresta, para o processo suavização por difusão anisotrópica modificada; (ii) experimentos relativos à detecção de bordas; (iii) a investigação dos valores adequados à CEF aos propósitos almejados; (iv) a investigação dos valores adequados para os parâmetros da filtragem NLM.

Ressalta-se que a atribuição de valores a cada parâmetro de cada módulo da abordagem proposta foi realizada de forma experimental. Tais experimentos foram realizados por meio do processamento de 12 imagens²⁶ comumente utilizadas em pesquisas na área de NPR. Na realização de tais experimentos, utilizou-se um total de 12 imagens e, nesta seção, foi utilizado apenas um excerto da imagem da Ilha de Capri (apresentada no Capítulo 3), visando à melhor explicitação dos efeitos obtidos a partir de cada experimento realizado.

²⁶ Conjunto original de imagens apresentado em: <http://www.kyprianidis.com/p/eg2011/gallery2/>

A investigação de valores de parâmetros ocorreu de forma visual, na qual o efeito produzido após o processamento foi investigado e a variação dos valores dos parâmetros foi armazenada. Tal investigação ocorreu em três etapas, a saber: (i) a primeira etapa consistiu na atribuição de valores iniciais aos parâmetros do módulo sob experimentação; (ii) a segunda etapa consistiu na verificação dos efeitos advindos da alteração dos valores da etapa inicial; (iii) a terceira etapa consistiu na alteração de valores dos parâmetros, procedida da verificação dos resultados cuja alteração impactava significativamente o resultado obtido.

Assim sendo, a pré-seleção de valores foi realizada a partir do emprego de subgrupos de valores para cada parâmetro de cada módulo sob experimentação. A divisão em subgrupos determinou um conjunto aceitável de valores para cada parâmetro investigado em todos os módulos da abordagem proposta, os quais originaram os resultados relatados nas seções a seguir. Na Seção 5.2, apresenta-se o processo de validação dos valores obtidos nesta fase de pré-seleção.

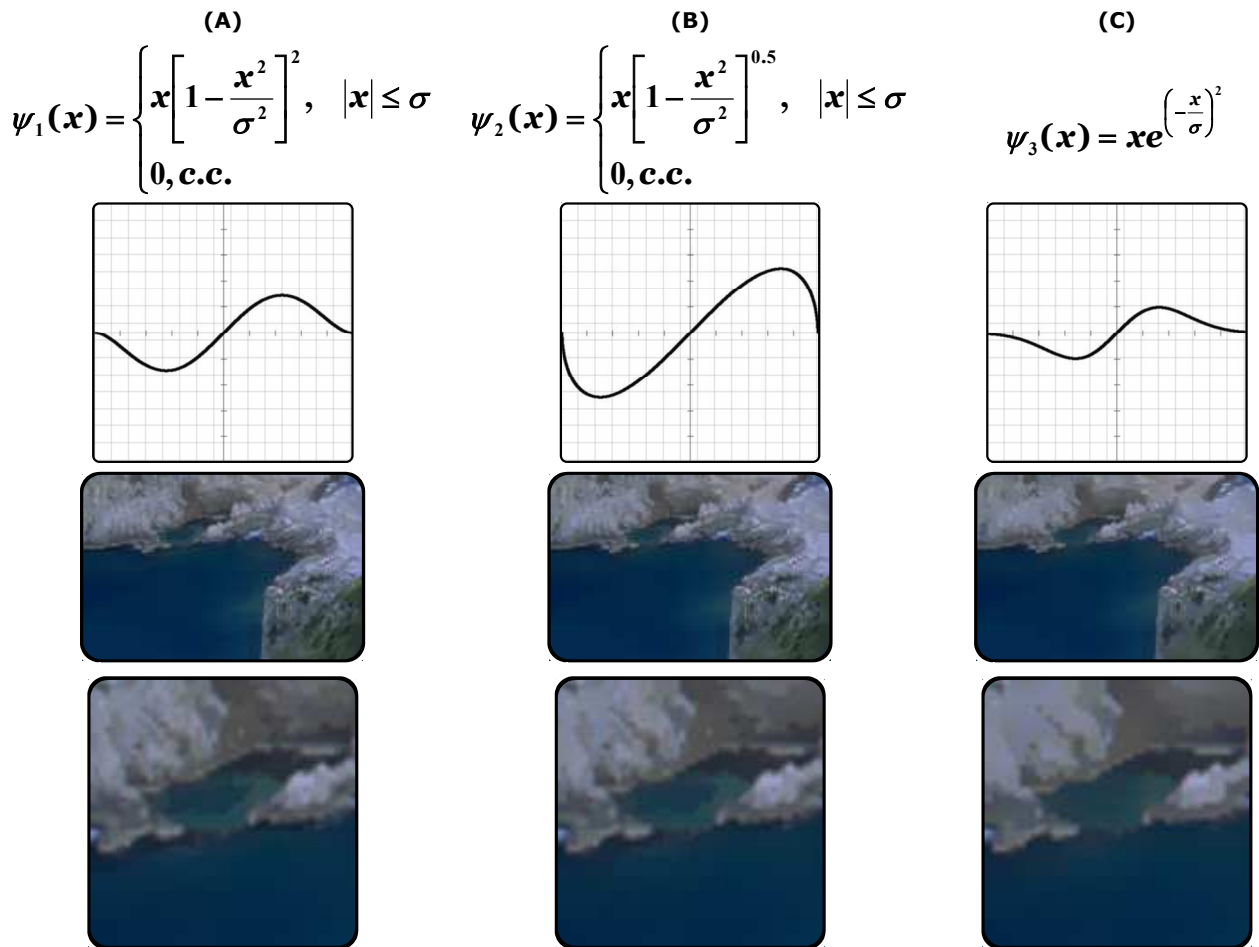
5.1.1 Experimentos relativos à Suavização por Difusão Anisotrópica

Nesta seção, são apresentados experimentos relativos à filtragem por difusão anisotrópica, sendo destacados os efeitos produzidos pela alteração da função de parada-na-aresta no processo de suavização.

É inerente ao processo de filtragem por difusão anisotrópica a pesquisa de três fatores que afetam significativamente os resultados obtidos: (i) a função de parada-na-aresta (ψ) *escolhida*; (ii) a velocidade da difusão (λ) *empregada*; e (iii) a quantidade de iterações (N) *considerada*.

A função de parada-na-aresta é responsável pelo controle da intensidade da difusão em cada pixel da imagem original. Conforme ressaltado na Seção 2.4 (Capítulo 2), Black et al. (1998) propuseram a função de Tukey (*Tukey's biweight function*) como função de parada-na-aresta para o processo de difusão anisotrópica. A função de Tukey original é apresentada na Figura 5.1A. O valor do expoente desta função foi experimentalmente alterado de 2 para 0,5, produzindo o resultado apresentado na Figura 5.1B. A partir dos experimentos realizados, investigou-se, à luz da Estatística Robusta, se outra função distinta daquela proposta por Black et al. (1998) poderia se afigurar mais adequada aos propósitos desta pesquisa, no tocante ao nível de abstração resultante de sua aplicação. Após investigar outras opções de funções de parada-na-aresta, decidiu-se experimentar a *função de Dennis e Welsch* (1978), apresentada na Figura 5.1C, como substituta para a função de Tukey.

Figura 5.1 – Difusão anisotrópica robusta com funções de parada na aresta distintas: (A) função de Tukey original; (B) função de Tukey modificada; e (C) Função de Dennis e Welsch (1978).



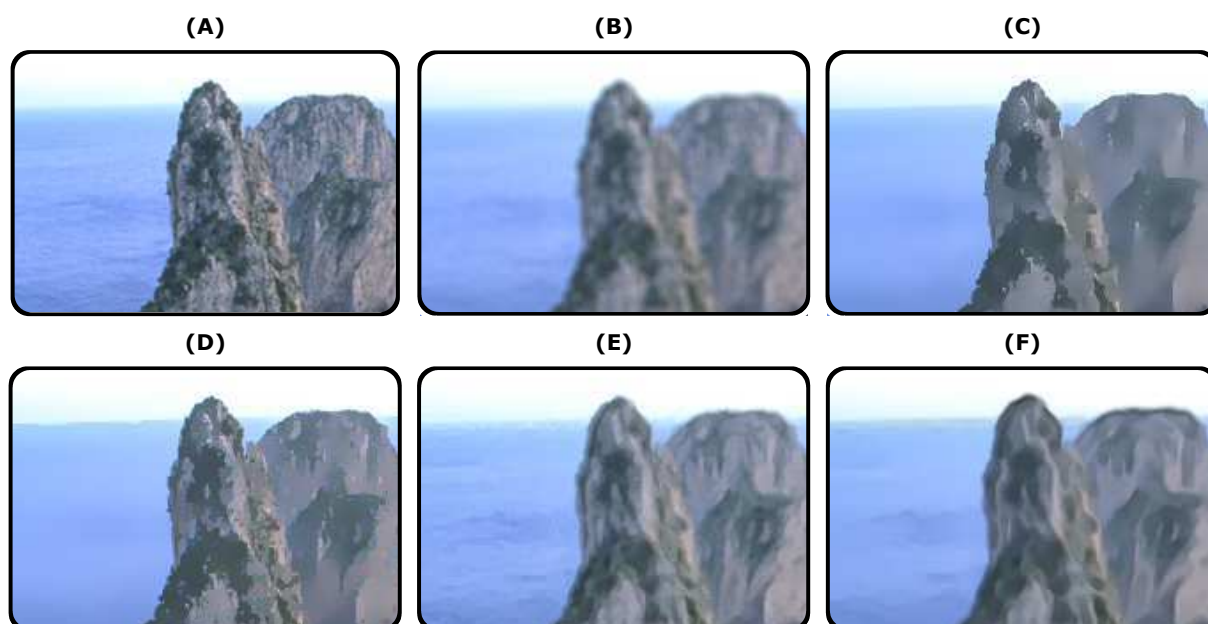
Inicialmente, experimentou-se a função de parada-na-aresta proposta por Black et al. (1998). Durante os experimentos realizados, dois parâmetros exerceram influência significativa nos resultados obtidos no processo de filtragem por difusão anisotrópica: (i) o valor para o desvio padrão (σ) a ser utilizado no processo de suavização; e (ii) a quantidade de iterações (N) a serem consideradas. Para cada uma das funções consideradas anteriormente, foram realizados experimentos com 12 imagens selecionadas dentre aquelas comumente utilizadas na área de NPR, conforme mencionado no início desta seção. Tais experimentos foram realizados com os seguintes valores dos parâmetros: $\sigma=0,025$, $N=100$, $\lambda=0,25$. O resultado da aplicação do processo de difusão à imagem da Ilha de Capri pode ser visualizado nos excertos das imagens apresentadas na Figura 5.1, na qual se encontram: (i) na primeira linha, as três funções de parada-na-aresta estudadas; (ii) na segunda linha, os gráficos das funções de parada-na-aresta consideradas; (ii) na terceira linha, um excerto da imagem renderizada, para melhor visualização do resultado obtido; (iii) na quarta linha, uma ampliação do excerto da imagem renderizada, na qual há maior aproximação e detalhamento da imagem de referência.

Verifica-se, a partir da observação dos detalhes presentes nos excertos das imagens, que o processo de suavização ocorre com preservação das bordas. Observa-se, no detalhe dos excertos das imagens apresentadas na Figura 5.1B, que a abstração produzida assemelha-se àquela produzida a partir da aplicação da função de parada-na-aresta padrão (imagens apresentadas na Figura 5.1A), embora o resultado apresente maior grau de suavização (desfocamento) das regiões e menor grau de preservação das bordas, o que pode ser observado, por exemplo, na diferença da textura da montanha e na textura do lago interno apresentado na imagem suavizada.

Constata-se, mediante a observação dos detalhes presentes nos excertos da Figura 5.1C, que a suavização produzida se assemelha àquela resultante da aplicação da função de parada-na-aresta padrão embora, desta vez, o resultado apresente maior grau de suavização. De modo geral, observa-se maior estilização do resultado quando a função de Dennis e Welsch (1978) é aplicada, e.g., os detalhes do lago interno, da textura da montanha, da maior suavização da textura da água são mais abstraídos na Figura 5.1C quando comparada à Figura 5.1A e à Figura 5.1B.

A seguir, é apresentada uma comparação visual dos resultados obtidos a partir da aplicação da abordagem proposta e de algumas técnicas afins ao processo de difusão anisotrópica. Na Figura 5.2, é apresentada uma comparação dos resultados obtidos a partir da abordagem proposta e da aplicação das seguintes técnicas: (i) *filtragem Gaussiana*; (ii) *filtragem bilateral* (TOMASI; MANDUCHI, 1998); (iii) *difusão anisotrópica robusta* (BLACK et al., 1998).

Figura 5.2 – Comparação de resultados: (A) Original; (B) Suavização Gaussiana; (C) Suavização por filtro Bilateral; (D) Difusão anisotrópica robusta; (E) Resultado da difusão anisotrópica utilizada na abordagem proposta; (F) Resultado da utilização dos seis módulos da abordagem proposta.



O maior grau de abstração associado à aplicação da função de Dennis e Welsch (1978), conjugado à preferência de tal função evidenciada no processo de avaliação descrito na Seção 5.2, corroborou a escolha desta função para a composição da abordagem proposta.

Vale salientar que a abordagem proposta contém módulos que enriquecem o processo de suavização, os quais não constam nas técnicas comparadas. Entretanto, a comparação visual apresentada nesta seção foi motivada pela pesquisa realizada, em especial, na comparação apresentada por Gastal e Oliveira (2011) e por Zhao, Jin e Mao (2011), nos quais as técnicas de suavização apresentadas na Figura 5.2 são comparadas a abordagens propostas pelos autores. Adicionalmente, é relevante salientar que se notou a necessidade de comparação com técnicas de referência (suavização isotrópica e suavização anisotrópica).

No tocante ao filtro Gaussiano, a importância de sua presença na comparação visual foi evidenciada pela comparação visual apresentada por Zhao, Jin e Mao (2011). A filtragem Gaussiana se relaciona à abordagem proposta no que diz respeito à relação existente entre a difusão isotrópica e a difusão anisotrópica, apresentada no Capítulo 2. A máscara Gaussiana aplicada à imagem apresentada na Figura 5.2B tinha dimensões de 5x5 pixels, com $\sigma=11$.

É importante ressaltar características intrínsecas à *filtragem bilateral* (TOMASI; MANDUCHI, 1998), uma técnica semelhante à *filtragem por difusão anisotrópica* (PERONA; MALIK, 1990). Trata-se também de um processo de suavização de imagens com preservação de arestas, fundamentado na utilização da distância espacial e da distância radial entre os pixels, combinadas por meio de duas funções de ponderação, a fim de determinar o peso atribuído a cada pixel da vizinhança, e.g., ao se atribuir pesos de baixa magnitude aos pixels com valores muito diferentes do valor do pixel central, o processo de filtragem evita que um pixel de borda seja fortemente influenciado por pixels vizinhos, preservando as transições entre regiões.

Nos testes realizados, foram utilizadas três iterações consecutivas do filtro bilateral, tendo sido o diâmetro da vizinhança de cada pixel configurado para 30, o valor da distância espacial para 80 e o valor da distância radial entre cores para 50. Adicionalmente, para efeitos de comparação, foi incluída a difusão anisotrópica robusta, proposta por Black et al. (1998).

Observa-se, na Figura 5.2, que os resultados obtidos apresentam diferenças visíveis, em maior ou menor grau, na renderização da textura da montanha, no tema retratado. Tais diferenças suscitaram a investigação de estratégias que possibilitaram

o melhoramento da abordagem proposta, e.g., a melhor calibração do detector de bordas ou a verificação do número de iterações do algoritmo de difusão anisotrópica considerado. A suavização Gaussiana não confere um aspecto estilizado ao objeto retratado. A aplicação iterativa da filtragem bilateral e da filtragem por difusão anisotrópica confere um esmaecimento às cores da imagem, resultante do processo de filtragem e desfocamento.

Considerando-se a análise visual das abordagens comparadas, alguns aspectos da abordagem proposta podem ser destacados: (i) maior grau de estilização obtido; (ii) maior definição das fronteiras entre regiões; (iii) homogeneização cromática intrarregional. Ressalta-se que parte das texturas do tema retratado é melhor renderizada na abordagem proposta.

A seguir, são apresentados experimentos relativos à filtragem com realce da coerência de regiões (CEF).

5.1.2 Experimentos relativos à CEF

A aplicação da CEF às imagens de entrada da abordagem proposta, visando à abstração das imagens, possibilitou o realce das bordas e a modificação de estruturas curvilíneas localizadas no interior das regiões delimitadas por bordas. Tal técnica foi aplicada com o intuito de realçar aspectos direcionais da imagem, a fim de simplificá-la e estilizá-la, como ocorreu em pesquisas anteriores (WEICKERT, 1999; KANG; LEE, 2008).

Foram realizados experimentos relativos à atribuição de valores aos seguintes parâmetros da CEF: (i) o tamanho do passo temporal (τ) a ser considerado; (ii) as dimensões da máscara e o tipo da máscara utilizada, i.e., máscara invariante à rotação (definida por Weickert, 1999), de dimensão 5x5 pixels; e (iii) o valor do desvio padrão para a suavização do tensor (ρ).

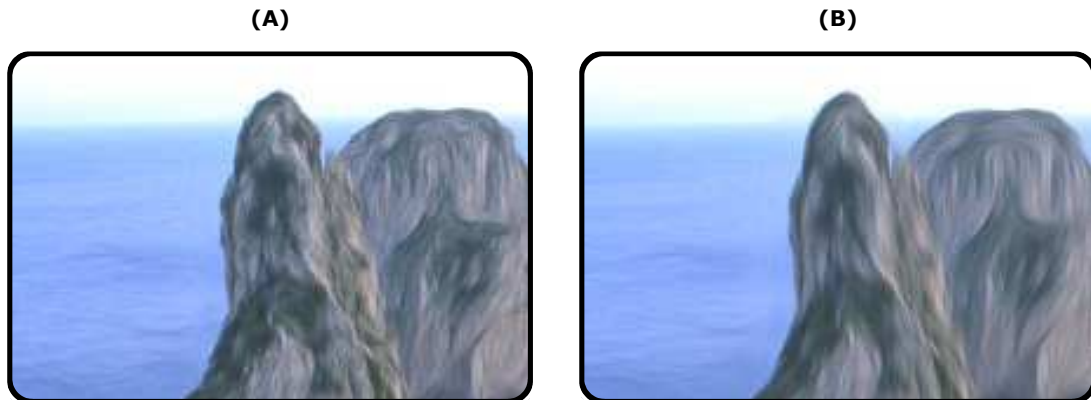
Fundamentado na pesquisa realizada por Coupé et al. (2012), a implementação computacional desenvolvida por Dirk-Jan Kroon²⁷ contemplou a implementação de uma máscara de filtragem invariante à rotação. De acordo com o autor da implementação computacional, verificava-se um desfocamento das bordas, dependendo do ângulo a estas associado. Adicionalmente, verificou-se que, ao se incrementar o passo temporal ($\tau \geq 100$), quando da filtragem de quadros consecutivos de vídeo, ocorria o efeito de cintilação (*flickering*).

Na Figura 5.3, exemplifica-se, em um excerto da imagem da Ilha de Capri, a

²⁷ Disponível em: <http://www.mathworks.com/matlabcentral/fileexchange/27395-fast-non-local-means-1d-2d-color-and-3d>

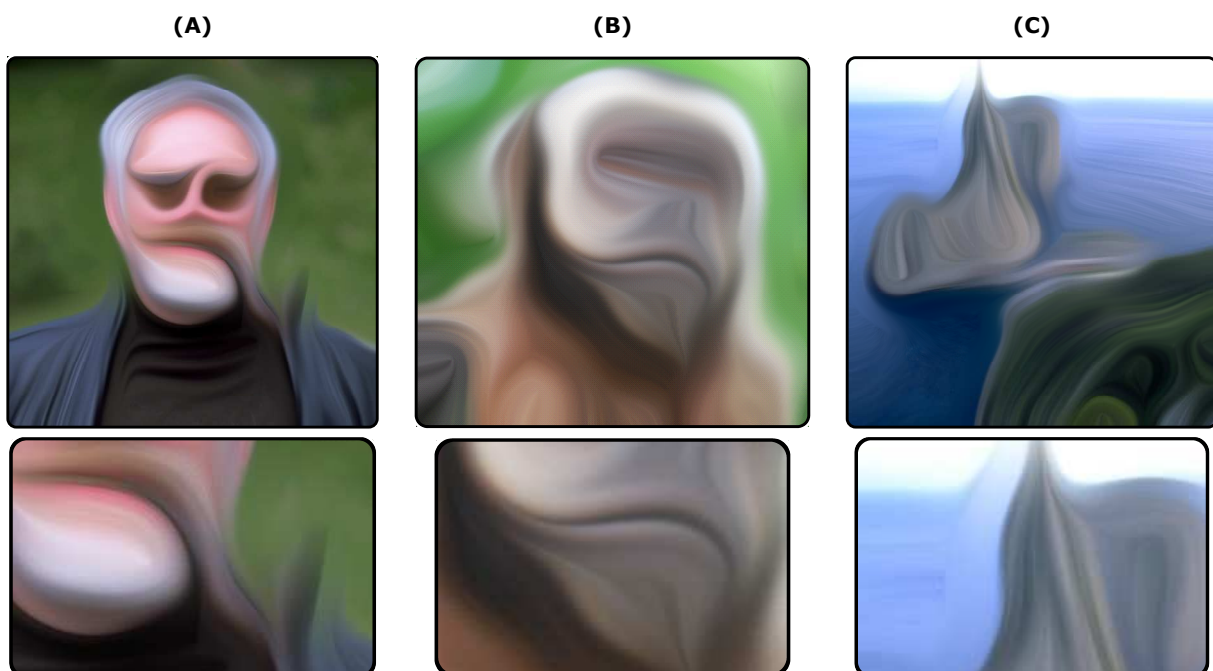
aplicação da CEF, utilizando-se os seguintes valores: $\tau=2$ e $\rho=5$ (Figura 5.3A) e $\tau=6$ e $\rho=7$ (Figura 5.3B). Observa-se, no resultado apresentado na Figura 3.4B, o efeito acentuado resultante da aplicação da CEF à imagem. A renderização da Figura 5.3A foi realizada em 4 segundos, enquanto a renderização da Figura 5.3B foi realizada em 14 segundos, ao ser utilizado o mesmo *hardware* descrito no início deste capítulo.

Figura 5.3 – Exemplo de filtragem utilizando a CEF: (A) $\tau = 2$ e $\rho = 5$; e (B) $\tau = 6$ e $\rho = 7$.



Em um experimento posterior, aplicou-se a CEF utilizando-se valores dos parâmetros τ e ρ inadequados aos objetivos de pesquisa, i.e., $\tau=400$ e $\rho=15$ (Figura 5.4). Verificou-se que, para tais valores, a finalização do processamento deu-se após cerca de 25 minutos. O resultado de tal experimento pode ser observado na Figura 5.4, na qual se percebe a total descaracterização dos objetos retratados nas imagens.

Figura 5.4 – Adoção de valores inadequados para os parâmetros τ e ρ da CEF.



Os experimentos permitiram elucidar a influência dos principais parâmetros a serem considerados na aplicação do efeito estilístico apresentado. Foi escolhido o

processo com menor tempo de execução e menor efeito estilístico visível, mas perceptível. Diante do exposto, i.e., diante da diferenciação dos resultados obtidos após o experimento com valores diversos dos parâmetros da CEF, optou-se pela pré-seleção dos valores $\tau=2$ e $\rho=5$, a serem utilizados na abordagem proposta, no módulo que utiliza a CEF. Adicionalmente, a validação de tais valores ocorreu no processo de votação, apresentado na Seção 5.2.

A seguir, são apresentados experimentos relativos ao processo de sobreposição de bordas à imagem filtrada por difusão anisotrópica.

5.1.3 Experimentos relativos à Sobreposição de Bordas

A combinação do processo de filtragem por difusão anisotrópica com a ponderação por um detector de bordas neural suscitou a investigação de valores relativos aos parâmetros que afetam significativamente os resultados obtidos. Tal combinação provê, dentro do processo de suavização, a redução do número de iterações do algoritmo de suavização por difusão anisotrópica.

Assim sendo, ao se reduzir o número de iterações ($N \leq 30$, em contraposição ao valor $N=100$, inicialmente considerado), o aspecto visual não apresentou alterações significativas. Entretanto, tal redução implicou a redução do tempo necessário para o processamento dos quadros do vídeo e, conseqüentemente, a aceleração do processo de abstração. No âmbito da pesquisa relatada nesta tese, foi utilizada a abordagem para a geração de mapa de bordas utilizando redes neurais, proposta por Arruda, Queiroz e Gomes (2012). Os valores para os parâmetros dos módulos que compõem tal abordagem estão descritos na referida pesquisa.

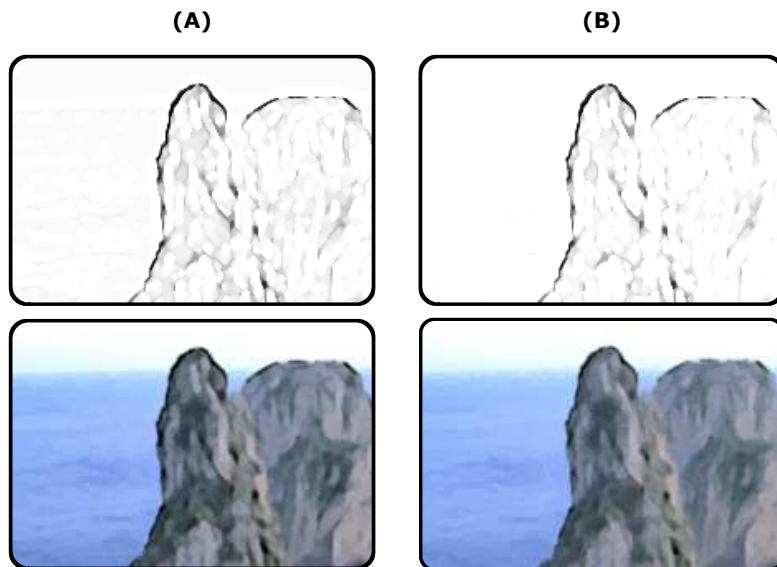
Nesta seção, apresenta-se apenas o experimento relativo à ponderação do mapa de bordas a ser sobreposto à renderização obtida após o processo de suavização por difusão anisotrópica adotada. Os experimentos relatados nesta seção visaram a elucidar a influência visual da alteração do valor de brilho do mapa de bordas sobreposto à imagem renderizada.

Ressalta-se que tal sobreposição ocorre após a etapa de realce de regiões de sombra, presente no módulo D da abordagem proposta. A sobreposição consiste na ponderação entre a imagem suavizada e as bordas geradas pelo detector de bordas neural. Antes de tal ponderação, há o ajuste de brilho do mapa de bordas, o qual visa à melhoria do aspecto final da renderização, tal qual a menor presença de ruído. O ganho do valor de brilho (**B**) foi realizado considerando-se os valores 10 e 50. A ponderação do mapa de bordas com brilho ajustado e da imagem suavizada por

difusão anisotrópica ocorre a partir da multiplicação dos valores dos pixels de todos os canais da imagem suavizada pelos valores dos pixels da imagem de borda. Após essa multiplicação, ocorre uma normalização (divisão por 255), quando utilizados 8 bits para a representação computacional do valor de um pixel.

Na Figura 5.5, é apresentado o efeito da sobreposição das bordas à imagem suavizada, utilizando-se os valores dos parâmetros supramencionados.

Figura 5.5 – Comparação de ponderação de bordas com diferentes valores de brilho: (A) $B=10$; e (B) $B=50$.



Observa-se o efeito de estilização da imagem mais proeminente, do ponto de vista de maior realce das regiões de fronteira entre a montanha e o mar, no excerto da cena retratada, quando as bordas modificadas com valores de brilho menores são combinadas à imagem gerada a partir do processo de difusão anisotrópica. Tal efeito contrapõe-se ao menor efeito estilístico apresentado na Figura 5.5B, e.g., o efeito observado na transição da montanha para o céu.

Os experimentos realizados permitiram elucidar a influência da sobreposição das bordas detectadas à imagem filtrada a partir do processo de difusão anisotrópica. Foi escolhido o valor de ajuste de brilho $B=50$, o qual apresentou efeito da sobreposição das bordas menos proeminentes, porém visualmente perceptível. A validação da escolha de tal valor ocorreu no processo de votação, apresentado na Seção 5.2.

A seguir, são apresentados experimentos relativos à filtragem NLM.

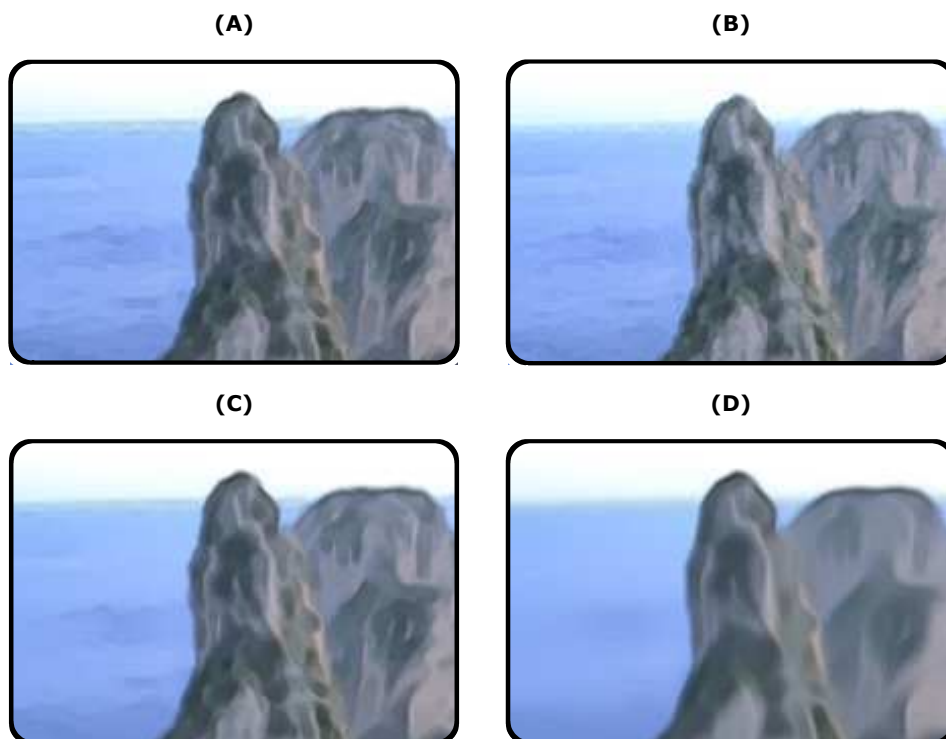
5.1.4 Experimentos relativos à Filtragem NLM

A suavização por *médias não locais* (NLM – *Non Local Means*), proposta por Buades, Coll e Morel (2005), tem como propósito a remoção de ruído em imagens estáticas

(para detalhamento, ver Seção 3.1, p. 105). Nesta pesquisa foi utilizada a filtragem NLM na qual a busca de regiões semelhantes é realizada ao redor de uma região retangular do pixel a ser filtrado, i.e., a região ao redor do pixel é comparada a blocos de pixels de sua vizinhança. Dois parâmetros afiguraram-se importantes durante a realização de experimentos relativos à filtragem NLM: (i) as dimensões, em pixels, do raio do bloco a ser comparado (R_b); e (ii) o raio de pesquisa dos pixels da vizinhança (R_v).

Na Figura 5.6, é apresentado o resultado da filtragem NLM e seu efeito sobre as regiões do excerto da imagem da Ilha de Capri, utilizando-se os seguintes valores dos parâmetros: (A) imagem após a sobreposição das bordas, sem a realização do processo de filtragem NLM; (B) $R_b=3$, $R_v=3$, filtragem realizada em 1,4 segundos; (C) $R_b=3$, $R_v=5$, filtragem realizada em 3,4 segundos; e (D) $R_b=7$, $R_v=9$, filtragem realizada em 60 segundos.

Figura 5.6 – Resultado da filtragem NLM para conjuntos de valores diferenciados: (A) Imagem de entrada; (B) $R_b=3$, $R_v=3$; (C) $R_b=3$, $R_v=5$; e (D) $R_b=7$, $R_v=9$.



Na implementação computacional da filtragem NLM considerada, ocorre a divisão da imagem em sub-blocos (tipicamente, 150 sub-blocos), visando ao emprego eficiente de memória e ao emprego de cache de CPU. Entretanto, a alteração da quantidade de sub-blocos não implicou na redução do tempo de execução, quando considerados os mesmos parâmetros para R_b e R_v .

Observa-se que os resultados obtidos apresentam diferenças visíveis, em maior

ou menor grau, na renderização da textura da montanha, no tema retratado. A aplicação da filtragem NLM com os valores que produziram os resultados exibidos nas Figura 5.6C e D conferiu um aspecto desfocado a toda a imagem, com perda das regiões antes texturizadas (visíveis na Figura 5.6A). Considerando-se (i) a análise visual dos resultados obtidos, (ii) o grau de estilização obtido, (iii) o grau de homogeneização cromática intrarregional e (iv) o tempo de processamento necessário à aplicação da filtragem, adotaram-se os valores $R_b=3$ e $R_v=5$ (Figura 5.6C) na abordagem proposta.

A seguir, são apresentados os experimentos relativos ao processo de avaliação comparativa entre abordagens destinadas à renderização não fotorrealística de vídeos digitais.

5.2 Resultados da Avaliação de Renderizações de Vídeos

O processo de avaliação de abordagens de renderizações fotorrealísticas compôs-se de cinco etapas distintas, evidenciadas no Capítulo 4. A etapa de planejamento do experimento está apresentada na Seção 4.1 (p. 117), enquanto a etapa de elaboração do material de teste está apresentada na Seção 4.2 (p. 129). Nesta seção são apresentados os resultados obtidos nas seguintes etapas: (i) configuração dos parâmetros das abordagens comparadas; (ii) delineamento do perfil dos participantes do experimento; e (iii) resultados da avaliação comparativa entre abordagens.

5.2.1 Configuração dos Parâmetros das Abordagens Comparadas

A classificação taxonômica proposta por Kyprianidis et al. (2013), apresentada no Capítulo 2 (ver Figura 2.1, p. 32), possibilitou a delimitação do escopo desta pesquisa à renderização não fotorrealística de vídeos a partir de um processo de difusão anisotrópica, assim como possibilitou a delimitação do processo comparativo das abordagens àquelas pertencentes ao mesmo grupo taxonômico.

No tocante ao grupo taxonômico ao qual se enquadram as abordagens fundamentadas na filtragem espacial, Kyprianidis et al. (2013) destacaram 21 pesquisas realizadas no período compreendido entre os anos 2004 a 2011. Esse grupo taxonômico pode ser expandido pela adição de outras pesquisas mais recentes (WINNEMÖLLER, 2011; MOULD, 2012). Verificou-se a inviabilidade da condução de um processo comparativo entre a abordagem proposta e todas as demais listadas no grupo taxonômico supramencionado, de modo que se decidiu restringir as abordagens consideradas no experimento àquelas enquadradas no subgrupo que fundamenta o processo de renderização na difusão anisotrópica, i.e., a restrição da quantidade de

abordagens comparadas justificou-se por tais abordagens não utilizarem o processo de difusão anisotrópica como técnica central do processo de estilização.

Destaca-se a exclusão do experimento comparativo ora relatado:

(i) da abordagem fundamentada no domínio do gradiente, proposta por Bhat et al. (2010);

(ii) das abordagens destinadas à produção de renderizações compostas exclusivamente por bordas e silhuetas, propostas por Orzan et al. (2007) e por Winnemöller (2011);

(iii) da abordagem fundamentada em estatísticas locais da imagem, proposta por Papari, Petrov e Campisi (2007); e

(iv) das abordagens para a renderização de regiões a partir de filtragem morfológica, propostas por Bousseau et al. (2007) e Criminisi et al. (2010).

Diante do exposto, e fundamentado na metodologia descrita no Capítulo 4 (Seção 4.1.3, p. 126), quatro abordagens foram selecionadas: (i) abstração em tempo real com o emprego do filtro bilateral e bordas obtidas por uma filtragem DoG (WINNEMÖLLER; OLSEN; GOOCH, 2006); (ii) abstração por difusão anisotrópica fundamentada no filtro Kuwahara (KYPRIANIDIS; KANG; DÖLLNER, 2009); (iii) abstração por transformação isométrica bidimensional em dois sinais unidimensionais (GASTAL; OLIVEIRA, 2011); e (iv) abstração em tempo real com o emprego do filtro bilateral fundamentado em fluxo (ETF) e bordas obtidas por uma filtragem FDoG (KYPRIANIDIS; DÖLLNER, 2008).

É relevante destacar que os proponentes das abordagens comparadas nesta pesquisa mencionaram, na literatura revisada, a aplicabilidade das referidas abordagens à produção de vídeos estilizados, por meio do processamento quadro-a-quadro do vídeo de entrada.

Após a seleção das abordagens a serem comparadas, procurou-se verificar os aspectos específicos de cada uma delas. A seguir, apresenta-se, para cada abordagem considerada: (i) a motivação para sua presença no processo de avaliação; (ii) os aspectos relativos a sua implementação computacional, e.g., a presença ou ausência de interface gráfica ou disponibilidade do código-fonte; e (iii) o processo de configuração de seus parâmetros.

Após um processo de configuração preliminar, procedeu-se à modificação dos valores dos parâmetros cuja alteração impactou, visualmente, o resultado da estilização. A etapa de pré-seleção foi fundamentada em experimentos realizados

sobre duas imagens, Henrik Andersson e Ilha de Capri, ambas apresentadas no Capítulo 3 (ver Figura 3.2, p. 95). A validação dos valores pré-selecionados foi realizada em uma etapa posterior, e está relatada na Seção 5.2.3 (p. 155).

A primeira abordagem selecionada foi aquela destinada à abstração em tempo real a partir da utilização do filtro bilateral (TOMASI; MANDUCHI, 1998), proposta por Winnemöller, Olsen e Gooch (2006) e, posteriormente, incorporada ao aplicativo *ToonPAINT* (apresentado no Capítulo 1). Kyprianidis (2013) a generalizou sob a forma de um arcabouço para a geração de imagens renderizadas ao estilo de cartuns (ver Figura 2.15), o qual serviu de inspiração para a abordagem aqui proposta, razão da presença da abordagem proposta por Winnemöller, Olsen e Gooch (2006) no processo de avaliação realizado nesta pesquisa.

Durante a realização desta pesquisa, apesar de não ter sido encontrada uma implementação computacional produzida e disponibilizada diretamente por Winnemöller, Olsen e Gooch (2006), encontrou-se uma versão codificada por Douglas Lanman²⁸, no MATLAB²⁹, a qual foi modificada durante a pesquisa realizada, com o intuito de viabilizar a renderização quadro-a-quadro de vídeos digitais. A modificação do código-fonte incluiu a possibilidade da atribuição de valores diferenciados a cada parâmetro, visando à automatização do processo de geração das renderizações de todos os quadros dos vídeos selecionados.

A abordagem proposta por Winnemöller, Olsen e Gooch (2006) contém três módulos principais (ver Figura 2.7, p. 45): (i) módulo destinado a filtragem com filtro bilateral; (ii) módulo destinado à detecção de bordas por meio da diferença de Gaussianas (DoG); (iii) módulo destinado à redução de cores mediante a pseudoquantização do canal de luminância.

A filtragem bilateral utiliza a distância espacial e radial entre os pixels, combinadas por meio de duas funções de ponderação, a fim de determinar a contribuição de cada pixel adjacente ao pixel a ser filtrado. O módulo destinado a detecção de bordas fundamenta-se na filtragem da imagem por diferença de Gaussianas (DoG). Nele, podem ser definidas a cor/espessura da borda, assim como a sensibilidade do detector. No módulo destinado à pseudoquantização do canal de luminância, define-se a quantidade de cores presentes na renderização. Uma descrição detalhada de cada parâmetro pode ser encontrada no Apêndice D.

Os parâmetros que exercem maior influência no resultado produzido pela

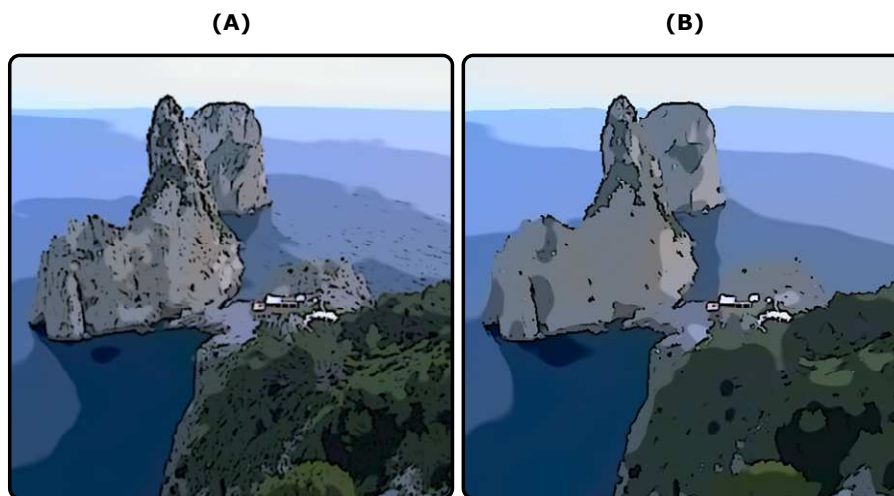
²⁸ Página pessoal disponível em: alumni.media.mit.edu/~dlanman/

²⁹ Código disponível em: www.mathworks.com/matlabcentral/fileexchange/12191-bilateral-filtering

abordagem proposta por Winnemöller, Olsen e Gooch (2006) são: (i) as dimensões da vizinhança do pixel a ser filtrado (ω), (ii) a contribuição da distância radial ao processo de suavização (σ_{rad}). Assim sendo, tais parâmetros foram os utilizados no processo de geração do conjunto de vídeos utilizados no processo de avaliação. Ressalta-se que a parametrização da abordagem proposta por Winnemöller, Olsen e Gooch (2006), a partir do uso de subgrupos de valores para os parâmetros descritos anteriormente, determinou o conjunto aceitável de valores para cada parâmetro investigado.

Na Figura 5.7, exemplifica-se, a partir da renderização da imagem da Ilha de Capri, a aplicação da abordagem proposta por Winnemöller, Olsen e Gooch (2006) utilizando-se valores adequados, $\omega=2$ e $\sigma_{rad}=5$ (Figura 5.7A) e inadequados, $\omega=10$ e $\sigma_{rad}=10$ (Figura 5.7B) ao propósito da renderização não fotorrealística. Observa-se, no resultado apresentado da Figura 5.7B a descaracterização do objeto retratado na imagem. Ressalta-se que uma imagem apresenta-se descaracterizada, com a utilização de valores inadequados, quando não há a observância de algum dos critérios apresentados na Seção 4.2.1, e.g., o assunto da cena original não pode ser reconhecido na renderização.

Figura 5.7 – Aplicação da abordagem proposta por Winnemöller, Olsen e Gooch (2006) com valores de parâmetros diferenciados: (A) $\omega=2$ e $\sigma_{rad}=5$; e (B) $\omega=10$ e $\sigma_{rad}=10$.



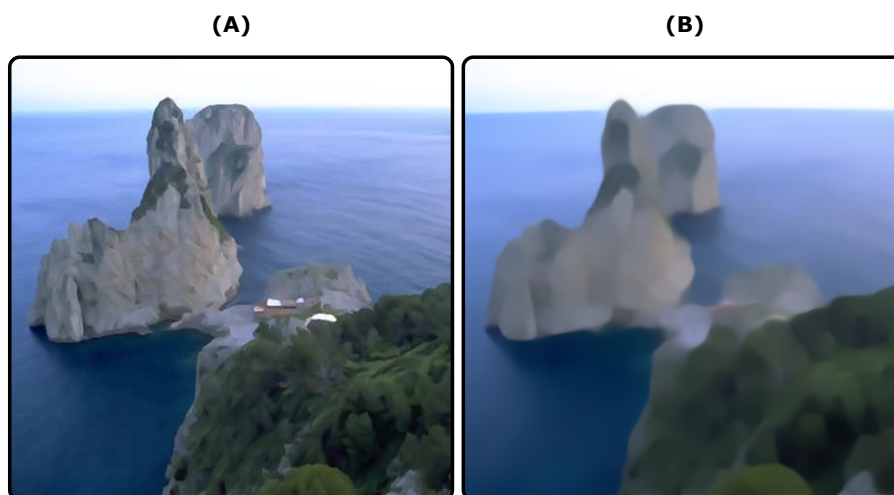
A segunda abordagem selecionada utiliza a difusão anisotrópica fundamentada no filtro Kuwahara, proposta por Kyprianidis, Kang e Döllner (2009). Conforme descrito no Capítulo 2 (ver Seção 2.2, p. 52), a abordagem proposta por Kyprianidis, Kang e Döllner (2009) fundamenta-se na computação da forma, escala e orientação da estrutura local, por meio da computação do fluxo direcional das características da imagem. A máscara anisotrópica sobre a qual suavização é aplicada, após sua subdivisão em setores triangulares, apresenta um formato elipsoidal, sendo construída com base no gradiente local.

Vale salientar que, ao contrário das demais abordagens utilizadas no processo de avaliação, a pesquisa desenvolvida por Kyprianidis, Kang e Döllner (2009) não contemplou módulos destinados à sobreposição de bordas ou redução da paleta de cores. Entretanto, sua presença no experimento justificou-se pela utilização da difusão anisotrópica no processo de estilização de imagens e vídeos.

Os parâmetros que exercem maior influência no resultado produzido pela abordagem proposta por Kyprianidis, Kang e Döllner (2009) são: (i) o valor para o fator de ponderação dos setores homogêneos da máscara (α); (ii) a quantidade de setores na qual a máscara em formato de elipse é subdividida (N); (iii) o raio da máscara em formato de elipse, em pixels (σ_r). Os subgrupos formados pelos valores destes parâmetros determinaram um conjunto aceitável de valores para cada parâmetro investigado.

Na Figura 5.8, exemplifica-se a aplicação da abordagem proposta por Kyprianidis, Kang e Döllner (2009) utilizando-se valores adequados, $\alpha=2$; $N=8$ e $\sigma_r=8$ (Figura 5.8A) e inadequados, $\alpha=10$; $N=8$ e $\sigma_r=18$ (Figura 5.8B) ao propósito da renderização não fotorrealística. Observa-se, no resultado apresentado da Figura 5.8B a descaracterização do objeto retratado na imagem, e.g., a excessiva homogeneização das regiões que compõe a porção de água da imagem.

Figura 5.8 – Aplicação da abordagem proposta por Kyprianidis, Kang e Döllner (2009) com valores de parâmetros diferenciados: (A) $\alpha=2$; $N=8$ e $\sigma_r=8$; e (B) $\alpha=10$; $N=8$ e $\sigma_r=18$.



Kyprianidis, Kang e Döllner (2009) disponibilizaram uma implementação computacional³⁰ da abordagem apresentada, que compreende uma interface gráfica na qual são detalhados todos os parâmetros supradescritos.

A terceira abordagem selecionada foi proposta por Gastal e Oliveira (2011), na

³⁰ Código disponível em: <https://code.google.com/p/gpuakf/>

qual a renderização não fotorrealística é realizada por meio de uma transcrição (*manifold*) da imagem bidimensional de entrada em dois sinais unidimensionais, os quais são submetidos a um processo de suavização com preservação de bordas. Havendo a preservação das distâncias dos pontos da imagem no espaço unidimensional, estas serão conservadas após o processo de filtragem.

A abordagem proposta por Gastal e Oliveira (2011) contemplou dois módulos principais: (i) módulo destinado à filtragem anisotrópica; e (ii) módulo de detecção e superposição das bordas à imagem filtrada. A imagem gradiente é utilizada como mapa de bordas com espessura unitária (*um pixel*), sendo superposta à imagem de saída. A inclusão desta abordagem ao experimento foi motivada pela utilização, pelos autores, de um processo de filtragem com preservação de bordas distinto daqueles caracterizados anteriormente.

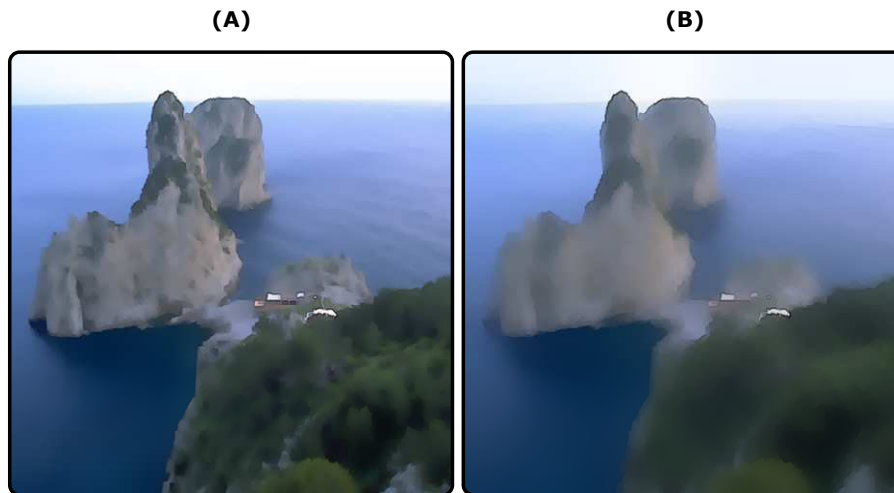
De acordo com Gastal e Oliveira (2011), o processo de filtragem é realizado em duas etapas: (i) há o processamento das linhas que compõe a imagem de entrada; e (ii) há o processamento das colunas que compõe a imagem de entrada. A abordagem é iterativa e o número de iterações a ser aplicada em cada etapa dependerá da geometria e do conteúdo da imagem de entrada. Os autores observaram o aparecimento de artefatos nos resultados produzidos, tais como barras horizontais (*stripes*), quando os valores dos parâmetros da abordagem (σ_{radial} e $\sigma_{espacial}$) são aumentados. Tais artefatos não são percebidos se, a cada iteração do algoritmo, o valor de cada uma das variâncias (σ_{radial} e $\sigma_{espacial}$) for configurado para a metade do valor correspondente na iteração anterior. Adicionalmente, de acordo com os autores, o número típico de 3 iterações afigura-se necessário para que a filtragem obtenha resultados satisfatórios.

Os autores também afirmaram que a filtragem realizada a partir da convolução normalizada (NC) é a mais adequada aos propósitos da estilização e abstração de imagens, pois regiões similares são suavizadas, enquanto suas bordas são preservadas e acentuadas. A utilização do modo NC resulta em maior resposta da filtragem quando o sinal considerado está localizado nas adjacências dos pixels de bordas.

Na Figura 5.9, exemplifica-se a aplicação da abordagem proposta por Gastal e Oliveira (2011) utilizando-se valores adequados, $\sigma_{radial}=25$; $\sigma_{espacial}=0,4$ e $N=3$ (Figura 5.9A) e inadequados, $\sigma_{radial}=60$; $\sigma_{espacial}=1,0$ e $N=21$ (Figura 5.9B) ao propósito da renderização não fotorrealística. Observa-se, no resultado apresentado da Figura 5.9B a descaracterização do objeto retratado na imagem, e.g., o desfocamento das regiões

de floresta.

Figura 5.9 – Aplicação da abordagem proposta por Gastal e Oliveira (2011) com valores de parâmetros diferenciados: (A) $\sigma_{\text{radial}}=25$; $\sigma_{\text{espacial}}=0,4$ e $N=3$; e (B) $\sigma_{\text{radial}}=60$; $\sigma_{\text{espacial}}=1,0$ e $N=21$.



Gastal e Oliveira (2011) disponibilizaram uma implementação computacional³¹ da abordagem apresentada, a qual compreende uma interface gráfica que detalha todos os parâmetros descritos, codificada via *framework* Qt. Adicionalmente, os autores disponibilizaram o código-fonte da abordagem na versão MATLAB. Para os propósitos do experimento, o referido código foi modificado, com o intuito de incluir um módulo destinado à computação do mapa de bordas, utilizado no processo de estilização.

A quarta abordagem selecionada foi proposta por Kyprianidis e Döllner (2008), na qual a estilização de imagens é realizada em tempo real a partir do emprego do filtro bilateral, fundamentado na computação do fluxo de características da imagem. O filtro bilateral utilizado na abordagem proposta pelos autores apresentava melhorias em relação ao filtro bilateral original (TOMASI; MANDUCHI, 1998), relacionadas ao emprego dos elementos componentes da imagem, obtidos após a computação do fluxo do campo de vetores da imagem, i.e., por meio da computação do Fluxo Tangente das Bordas (ETF - *Edge Tangent Flow*).

De acordo com os autores, após a computação do campo de vetores da imagem, gera-se um mapa de bordas, a partir do uso da Diferença de Gaussianas Baseada em Fluxo (FDoG - *Flow-Based Difference Of Gaussians*), a qual favorece a geração de contornos bem definidos. Em seguida, o mapa de bordas gerado é sobreposto à imagem filtrada mediante a utilização do filtro bilateral.

A abordagem proposta por Kyprianidis e Döllner (2008) assemelha-se àquela proposta por Winnemöller, Olsen e Gooch (2006) e contém cinco módulos principais:

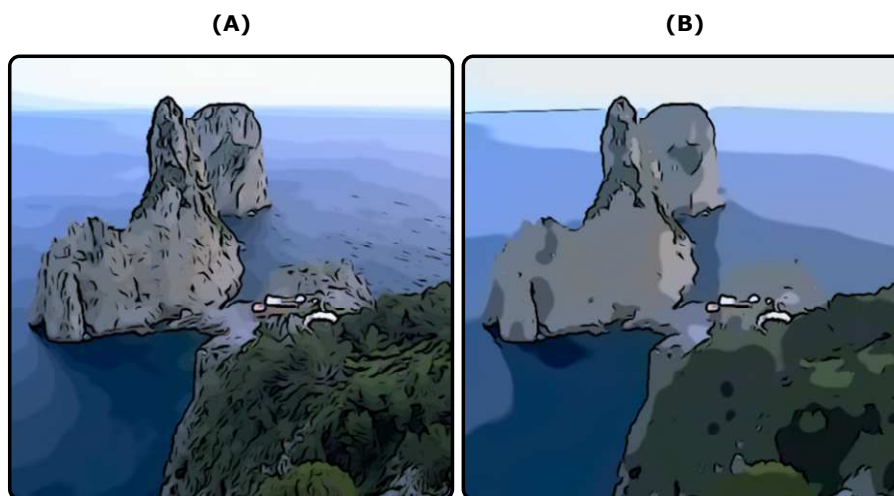
³¹ Código disponível em: <http://inf.ufrgs.br/~eslgastal/DomainTransform/>

(i) módulo destinado à extração do ETF; (ii) módulo destinado à filtragem com filtro bilateral fundamentado no ETF; (iii) módulo destinado a detecção de bordas FDoG; (iv) módulo destinado a redução de cores com a pseudoquantização do canal de luminância; e (v) módulo destinado a suavização da imagem.

Os parâmetros que exercem maior influência sobre o resultado produzido pela abordagem proposta por Kyprianidis e Döllner (2008) são: (i) a quantidade de iterações do filtro bilateral após a pseudoquantização das cores (n_a); (ii) a quantidade de iterações do filtro bilateral após a aplicação do filtro de detecção de bordas (n_e); e (iii) a quantidade de cores a serem reduzidas no processo de pseudoquantização (n_{bins}).

Na Figura 5.10, exemplifica-se a aplicação da abordagem proposta por Kyprianidis e Döllner (2008) utilizando-se valores adequados, $n_a=4$; $n_e=1$ e $n_{bins}=20$ (Figura 5.10A) e inadequados, $n_a=10$; $n_e=10$ e $n_{bins}=7$ (Figura 5.10B) ao propósito da renderização não fotorrealística. Observa-se, no resultado apresentado da Figura 5.10B, o vazamento intraregional de regiões de floresta e regiões de rochas.

Figura 5.10 – Aplicação da abordagem proposta por Kyprianidis e Döllner (2008) com valores de parâmetros diferenciados: (A) $n_a=4$; $n_e=1$ e $n_{bins}=20$; e (B) $n_a=10$; $n_e=10$ e $n_{bins}=7$.



Kyprianidis e Döllner (2008) disponibilizaram uma implementação computacional³² de sua abordagem, compreendendo uma interface gráfica que detalha todos os parâmetros supradescritos, codificada via *framework* Qt. A implementação disponibilizada pelos autores faz uso de recursos de GPU, o que confere à técnica proposta a capacidade de processamento de imagens e vídeos em tempo real.

Após seleção das abordagens a serem comparadas e a configuração cuidadosa

³² Código disponível em: <https://code.google.com/p/flowabs/>

dos parâmetros de cada abordagem, procedeu-se o recrutamento dos participantes do experimento e à realização dos ensaios experimentais. Os dados colhidos são apresentados nas seções seguintes.

5.2.2 Delineamento do Perfil dos Participantes do Experimento

No experimento relatado neste capítulo, fundamentado na metodologia de avaliação descrita no Capítulo 4 (Seção 4.1.4, p. 127), considerou-se uma amostra de 30 participantes, os quais foram recrutados em função de sua familiaridade com conceitos artísticos, em especial com o ato de desenhar e com conceitos relativos às definições de traço, linha de silhueta e estilização artística.

Tal familiaridade afigurou-se importante por se considerar que, em pesquisas na área de NPR, pretende-se elucidar princípios e conceitos encontrados ao serem concebidas peças de arte tradicionais, tais como pintura, ilustração, desenho gráfico ou, ainda, a investigação de conceitos relativos à abstração da realidade, fazendo uso de renderizações expressivas e estilizadas.

Assim sendo, do grupo de 30 participantes (17 do sexo masculino e 13 do sexo feminino), foram recrutados 28 alunos regularmente matriculados no curso de graduação em Desenho Industrial da Universidade Federal de Campina Grande e 2 alunos do curso de graduação em Arte e Mídia da mesma universidade.

A formação do curso de Desenho Industrial contempla a especialidade em desenho/*design* de produtos. A especificidade na área de desenho de produto concerne à criação, concepção e especificação de *objetos* que serão produzidos de forma industrial e padronizada. Entretanto, apresentam-se na grade curricular do curso disciplinas específicas destinadas à concepção de projetos de peças gráficas, desenho por meio de observação e desenho à mão livre (*croqui*) ou à elucidação de técnicas de renderização volumétrica. Conclui-se que a formação dos alunos neste curso atendeu aos requisitos relativos à familiaridade do participante com conceitos da área de NPR.

Outras características relativas ao perfil dos usuários do experimento destinado à comparação de abordagens são apresentadas na Tabela 5.1. O participante médio do experimento é destro, possui ensino superior incompleto, está localizado na faixa etária entre 18 e 23 anos, utiliza sistemas computacionais diariamente. Percebe-se o balanceamento da quantidade de participantes do sexo masculino e feminino, e o equilíbrio de participantes que utilizam óculos e que não os utilizam.

Tabela 5.1 – Perfil dos participantes do experimento realizado nesta pesquisa.

Característica	Quantidade	Porcentagem
Ensino superior incompleto	28	93%
Ensino superior completo	2	07%
Sexo masculino	17	57%
Sexo feminino	13	43%
Destro	27	90%
Canhoto	3	10%
Utiliza óculos	16	53%
Não utiliza óculos	14	47%
Faixa etária 18-23	16	53%
Faixa etária 24-29	11	37%
Faixa etária 30-35	3	10%
Conhecimento básico em informática	3	10%
Conhecimento intermediário em informática	24	80%
Conhecimento avançado em informática	3	10%

Após a realização de cada sessão do experimento, foi solicitado a cada participante que respondesse, por meio do preenchimento do questionário pós-teste, perguntas relativas ao contexto de utilização das renderizações apresentadas. Os contextos apontados pelos participantes sugerem a utilização: na produção de vídeo infantil; na apresentação de vídeo com a aplicação de efeito aquarela; na confecção de propaganda de um produto; na estilização de vídeos pessoais; na adição de um efeito de intervenção em um curta-metragem de animação; na apresentação de um vídeo publicitário, em uma apresentação de vídeo com efeito de esmaecimento, no qual o vídeo passa do resultado mais abstrato (utilizando uma técnica de estilização apresentada) à cena real; na renderização de vídeos de sonhos e/ou de vídeos lúdicos; na renderização de fotos a serem inseridas em redes sociais, na renderização de excertos em vinhetas; e na renderização de vídeos de viagens.

Após a realização de cada sessão do experimento, foi solicitado que cada participante apresentasse sugestões de melhorias para o processo de votação. A seguir apresentam-se os aspectos relatados: dificuldades na diferenciação entre vídeos (relativo à etapa de comparação intra-abordagens); dificuldade na votação, quando considerados vídeos curtos (e.g. *skating board*); dificuldade na avaliação de vídeos, quando não há a presença de alvos humanos; dificuldade na votação quando são apresentados vídeos com conteúdo repetido; dificuldade na avaliação de vídeos nos quais as bordas e contornos apresentaram-se pronunciados; maior dificuldade em avaliar vídeos que apresentam movimento acentuado da câmera.

Após a finalização de todas as sessões do experimento, cuja duração totalizou 3 meses, procedeu-se à tabulação e análise dos resultados obtidos.

5.2.3 Resultados da Avaliação Comparativa entre Abordagens

Nesta seção, são apresentados e discutidos os resultados obtidos após a realização

dos ensaios experimentais. Destaca-se que, durante o processo de seleção das abordagens a serem comparadas à abordagem proposta, a configuração dos parâmetros de cada abordagem foi realizada de forma cuidadosa, i.e., os parâmetros pré-selecionados não contemplaram configurações de parâmetros inadequadas.

O processo ideal de avaliação da abordagem proposta, comparada às quatro abordagens concorrentes, deveria contemplar a configuração de todos os parâmetros para todas as abordagens. Verificou-se que o tempo total de uma sessão de testes ultrapassaria 20 horas. A configuração dos parâmetros e a redução da quantidade de vídeos a serem votados por cada participante, viabilizou a realização da seção de testes em um tempo médio de meia hora.

Visando a facilitar a descrição das comparações efetuadas, considerando-se todos os valores dos parâmetros descritos anteriormente, sumaria-se na Tabela 5.2 os diversos grupos de parâmetros considerados e suas respectivas configurações. Adicionalmente, o nome do autor de cada abordagem foi substituído pelo nome da abordagem, a saber: (i) *Transformada de Domínio*, proposta por Gastal e Oliveira (2011); (ii) *Kuwahara Anisotrópico*, proposta por Kyprianidis, Kang e Döllner (2009); (iii) *Flow Bilateral+FDoG*, proposta por Kyprianidis e Döllner (2008); e (iv) *Bilateral+DoG*, proposta por Winnemöller, Olsen e Gooch (2006).

Tabela 5.2 – Grupos de parâmetros para cada abordagem comparada.

Transformada de Domínio				Kuwahara Anisotrópico			
$\sigma_{radial}=25$		$\sigma_{radial}=40$		$\alpha=1$		$\alpha=2$	
$\sigma_{espacial}=0,3$		$\sigma_{espacial}=0,5$		$N=3$		$N=8$	
$N=3$	$N=6$	$N=3$	$N=6$	$\sigma_r=6$	$\sigma_r=8$	$\sigma_r=6$	$\sigma_r=8$
G1	G2	G3	G4	G1	G2	G3	G4

Flow Bilateral+FDoG				Bilateral+DoG			
$n_a=4$		$n_a=6$		$\omega=3 \times 3$		$\omega=5 \times 5$	
$N_{bins}=8$	$N_{bins}=12$	$N_{bins}=8$	$N_{bins}=12$	$\sigma_{rad}=0,1$	$\sigma_{rad}=0,3$	$\sigma_{rad}=0,1$	$\sigma_{rad}=0,3$
G1	G2	G3	G4	G1	G2	G3	G4

Abordagem proposta			
$\rho_{cef}=5$		$\rho_{cef}=7$	
$\psi_{bordas}=\text{Welsch}$		$\psi_{bordas}=\text{Tukey}$	
$B=10$	$B=50$	$B=10$	$B=50$
G1	G2	G3	G4

Salienta-se que o processo de avaliação fundamentou-se na medida de desempenho visual, i.e., procurou-se elucidar a qualidade percebida de vídeos por um ser humano ao analisar os elementos associados à: (i) preservação com nitidez, por áreas uniformemente coloridas, do tema central; (ii) apresentação de cintilação ou

artefatos espúrios; (iii) apresentação de aspecto visual agradável.

Houve o balanceamento igualitário na quantidade de vídeos a serem votados, tendo sido gerados 344 vídeos para cada abordagem comparada, que contemplavam todas as variações de valores de parâmetros. No total, foram gerados 1720 vídeos (344 vídeos/abordagem x 5 abordagens), os quais foram votados por 30 participantes. Cada participante votou 30 vezes, comparando os vídeos par-a-par (totalizando 60 vídeos comparados), os quais foram distribuídos de forma equitativa, com cada abordagem apresentada em quantidade idêntica a cada participante.

A mudança do vídeo a ser avaliado poderia ser realizada apenas após a marcação de todas as 6 escalas apresentadas. Nenhum vídeo continha som. No experimento realizado, o tempo médio de votação foi de 29:11 minutos (\pm 9:50 minutos), com o tempo máximo de 1 hora e mínimo de 19 minutos. Do total de 60 vídeos, 18 (30%) compuseram a categoria intra-abordagem e 42 (70%) compuseram a categoria interabordagem.

Fundamentado na abordagem descrita no Capítulo 4 (ver Seção 4.3, p. 132), apresenta-se a seguir os resultados das seguintes avaliações: (i) intra-abordagem, na qual se avalia o conjunto de parâmetros considerado adequado a cada uma das abordagens comparadas; (ii) interabordagem, na qual são relatadas comparações quando consideradas abordagens diferentes; e (iii) interabordagem quando considerado o melhor grupo de parâmetros intra-abordagem.

Deve-se ressaltar que o teste F ANOVA fator único permitiu verificar se as médias das amostras coletadas eram significativamente diferentes, com um grau de confiança igual a 95%.

A fim de elucidar a presença de diferenças estatisticamente significativas entre os grupos de parâmetros intra-abordagem, quando consideradas as cinco abordagens comparadas, foram conduzidos quinze testes F ANOVA fator único (3 testes para cada aspecto, 5 abordagens comparadas). Todos os testes F ANOVA fator único são apresentados no Apêndice G.

Após análise dos testes F ANOVA fator único, constatou-se que, usando-se $\alpha=0,05$, todos os valores F são *maiores* do que F_{crit} , quando considerados os aspectos 1 e 3, o que significa que **há diferenças** estatisticamente significativas entre as médias destes aspectos, para todas as abordagens consideradas. Na Tabela 5.3, apresentam-se sumariados os valores F e F_{crit} obtidos nos testes F ANOVA realizados.

Tabela 5.3 – Sumário dos testes F ANOVA, considerando avaliações intra-abordagens.

Abordagem	Aspecto	Valor F	Valor F _{crit}
Transformada de Domínio	Aspecto 1	3,334	2,682
	Aspecto 2	10,504	2,682
	Aspecto 3	5,038	2,682
Kuwahara Anisotrópico	Aspecto 1	5,680	2,682
	Aspecto 2	1,004	2,682
	Aspecto 3	3,307	2,682
Bilateral+DoG	Aspecto 1	3,334	2,682
	Aspecto 2	8,248	2,682
	Aspecto 3	7,094	2,682
Flow Bilateral+FDoG	Aspecto 1	4,095	2,682
	Aspecto 2	1,954	2,682
	Aspecto 3	4,231	2,682
Abordagem proposta	Aspecto 1	3,406	2,682
	Aspecto 2	1,100	2,682
	Aspecto 3	3,213	2,682

Os valores destacados em negrito na Tabela 4.1 evidenciam a não diferenciação dos grupos de parâmetros das abordagens Kuwahara Anisotrópico, Flow Bilateral+FDoG e da Abordagem Proposta, quando considerado o Aspecto 2. A ausência de diferença estatisticamente significativa evidencia que os conjuntos de parâmetros escolhidos para a votação não produziu renderizações de vídeos com a presença de cintilação ou artefatos espúrios.

Após a realização dos testes F ANOVA fator único, procedeu-se a comparação de médias de pares de grupos de parâmetros, por meio da utilização do teste de Tukey-Kramer. Um sumário dos resultados destes testes, com um intervalo de confiança de 95% para os grupos de parâmetros considerados, é apresentado na Tabela 5.4. Outros dados relativos ao teste realizado são apresentados no Apêndice H.

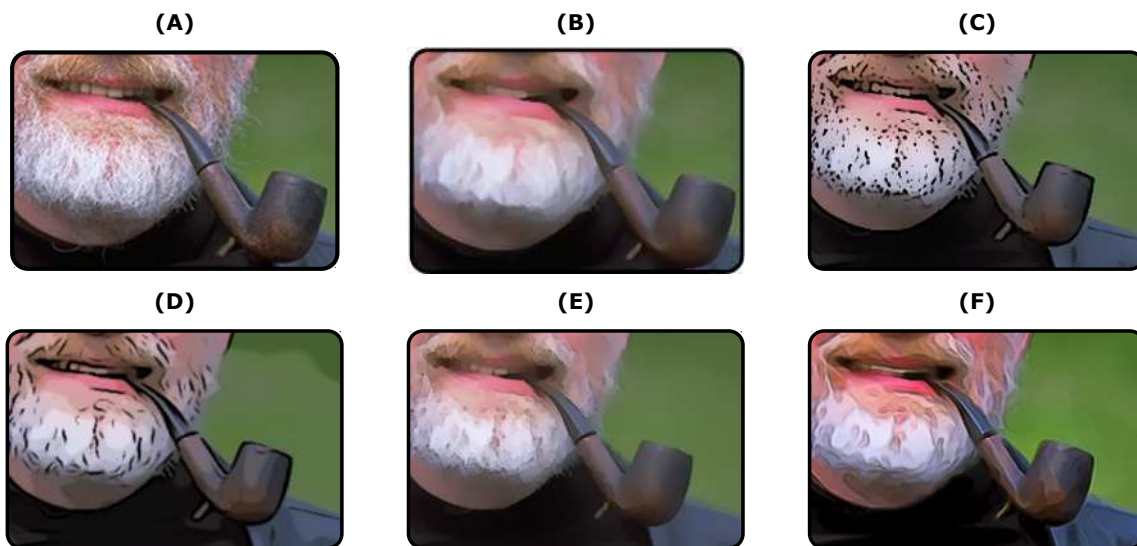
Tabela 5.4 – Sumário dos testes de Tukey-Kramer, considerando avaliações intra-abordagens.

Abordagem	Aspecto	Parâmetros com diferença significativa	Média estimada
Transformada de Domínio	Aspecto 1	G1	3,697
	Aspecto 2	G1	2,000
	Aspecto 3	G1	3,333
Kuwahara Anisotrópico	Aspecto 1	G1	3,571
	Aspecto 2	-	-
	Aspecto 3	G1	3,171
Flow Bilateral+FDoG	Aspecto 1	G1	3,800
	Aspecto 2	-	-
	Aspecto 3	G1	3,333
Bilateral+DoG	Aspecto 1	G2	3,619
	Aspecto 2	G2	2,500
	Aspecto 3	G2	3,515
Abordagem proposta	Aspecto 1	G2	3,704
	Aspecto 2	-	-
	Aspecto 3	G2	3,333

A ausência de diferenças estatisticamente significativas no Aspecto 2, apresentado na abordagens destacadas na Tabela 5.3, suscitou a escolha do grupo de

parâmetros, quando foram considerados os demais aspectos avaliados. Assim sendo, o grupo de parâmetros escolhido foi aquele que apresentou as maiores médias nos Aspectos 1 e 3. Tal escolha ocorreu apenas quando considerada a avaliação intra-abordagem. Destaca-se que o conjunto de parâmetros escolhido foi aquele que apresentou a maior preferência direta dos usuários, o qual também apresentou os maiores valores para a avaliação do estilo de cartum nas renderizações de vídeo consideradas. De posse de tais constatações, apresenta-se, na Figura 5.11, excertos da imagem de Henrik Andersson renderizados com os conjuntos de parâmetros mais bem avaliados.

Figura 5.11 – Renderizações dos grupos de valores de parâmetros mais bem avaliados: (A) Original; (B) Kuwahara Anisotrópico (G1: $\alpha=1$, $N=3$, $\sigma_r=6$); (C) Bilateral+DOG (G2: $\omega=3 \times 3$, $\sigma_{rad}=0,3$); (D) Flow Bilateral+FDoG (G1: $n_a=4$, $N_{bins}=8$); (E) Transformada de Domínio (G1: $\sigma_{rad}=25$, $\sigma_{esp}=0,3$, $N=3$); e (F) Abordagem proposta (G2: $\rho_{ref}=5$, $B=50$, ψ_{bordas} =Welsch).



Após a análise estatística dos dados intra-abordagem, procedeu-se a análise dos dados interabordagem. À semelhança do processo conduzido anteriormente, a fim de elucidar a presença de diferenças estatisticamente significativas quando as cinco abordagens avaliadas são comparadas, foram conduzidos três testes F ANOVA fator único. Cada teste conduzido visou a evidenciar diferenças na avaliação entre abordagens para cada aspecto considerado.

Todos os testes F ANOVA fator único, relativos às comparações interabordagem, são apresentados no Apêndice I. Após análise dos testes F ANOVA fator único, constatou-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 de todos os aspectos, entre todas as abordagens consideradas. Na Tabela 5.5, apresentam-se sumariados os valores F e F_{crit} obtidos nos testes F ANOVA realizados.

Tabela 5.5 – Sumário dos testes F ANOVA, considerando avaliações interabordagens.

Aspecto	Valor F	Valor F _{crit}
Aspecto 1	5,374	2,380
Aspecto 2	3,279	2,380
Aspecto 3	10,627	2,380

Após a realização dos testes F ANOVA, procedeu-se à comparação entre médias de pares de grupos, por meio do teste de Tukey-Kramer, cujos resultados, com um intervalo de confiança de 95%, são sumariados na Tabela 5.6. No Apêndice J, são apresentados todos os dados relativos aos testes de Tukey-Kramer realizados.

Tabela 5.6 – Grupos de similaridade do teste de Tukey-Kramer, considerando avaliações interabordagem.

Aspecto	Categoria	Média estimada	Grupos	
Aspecto 1	Abordagem Proposta	3,441	A	
	Flow Bilateral+FDoG	3,210	A	
	Bilateral+DoG	3,057		B
	Kuwahara Anisotrópico	3,005		B
	Transformada de Domínio	2,925		B
Aspecto 2	Flow Bilateral+FDoG	3,246	B	
	Transformada de Domínio	3,235	B	
	Bilateral+DoG	3,086	B	
	Abordagem Proposta	3,063	B	
	Kuwahara Anisotrópico	2,838		A
Aspecto 3	Abordagem Proposta	3,299	A	
	Flow Bilateral+FDoG	2,906		B
	Kuwahara Anisotrópico	2,902		B
	Bilateral+DoG	2,710		B
	Transformada de Domínio	2,601		B

A partir da análise dos testes de Tukey-Kramer, sumariados na Tabela 5.6, pode-se verificar que não há, estatisticamente, diferença significativa entre a Abordagem Proposta e a abordagem Flow Bilateral+FDoG, quando considerado o Aspecto 1. A inexistência dessa diferença evidencia a capacidade das duas abordagens de gerar renderizações não fotorrealísticas nas quais o tema central do vídeo é preservado com nitidez por áreas uniformemente coloridas delineadas por contornos escuros. Há diferença estatisticamente significativa entre a abordagem proposta e as abordagens Bilateral+DoG, Kuwahara Anisotrópico e Transformada do Domínio. Tal diferença evidencia a menor capacidade destas últimas de gerar regiões bem definidas delineadas por contornos escuros.

Na avaliação do Aspecto 2, relativo à presença de cintilação ou artefatos espúrios, deve-se considerar o menor valor de média obtido, i.e., menor valor implica menor presença de cintilação e menor presença de artefatos espúrios. Ao se analisar os testes de Tukey-Kramer apresentados na Tabela 5.6, concernentes ao Aspecto 2,

pode-se verificar que há diferença estatisticamente significativa entre a abordagem Kuwahara Anisotrópico e todas as demais abordagens consideradas. A presença desta diferença evidencia que a abordagem Kuwahara Anisotrópico apresenta o menor nível de cintilação ou de artefatos espúrios, do ponto de vista subjetivo, quando comparada às demais abordagens consideradas. Observa-se que, quando considerado o Aspecto 2, não há diferença estatisticamente significativa entre a Abordagem Proposta, a transformada de Domínio, a abordagem Bilateral+DoG e a abordagem Flow Bilateral+FDoG.

Ao se analisar os testes de Tukey-Kramer, apresentados na Tabela 5.6, referentes à avaliação do Aspecto 3, i.e., à preferência visual direta entre as abordagens comparadas, pode-se verificar que há diferença estatisticamente significativa entre a abordagem proposta e todas as demais. A preferência visual direta reúne evidências da superioridade, do ponto de vista subjetivo, da abordagem proposta em relação às demais abordagens avaliadas.

Todavia, as análises realizadas anteriormente suscitaram a busca de outras evidências, i.e., buscou-se analisar dados da votação interabordagem quando considerados os grupos intra-abordagem mais bem avaliados. À semelhança do processo conduzido anteriormente e a fim de elucidar a presença de diferenças estatisticamente significativas, foram conduzidos três testes F ANOVA fator único, cujos resultados são apresentados no Apêndice K.

Após a análise dos testes F ANOVA fator único, constatou-se, usando $\alpha=0,05$, que os valores F são *maiores* do que F_{crit} , quando considerados os Aspectos 1 e 2, o que significa que **há diferenças** estatisticamente significativas entre as médias entre estes aspectos. Na Tabela 5.7, resumiam-se os valores F e F_{crit} obtidos nos testes F ANOVA realizados.

Tabela 5.7 – Sumário dos testes F ANOVA, considerando avaliações interabordagens com maiores valores intra-abordagens.

Aspecto	Valor F	Valor F_{crit}
Aspecto 1	2,679	2,400
Aspecto 2	0,842	2,400
Aspecto 3	4,100	2,400

Após a realização dos testes F ANOVA, procedeu-se a comparação entre médias de pares de grupos, por meio do teste de Tukey-Kramer. Um sumário dos resultados do teste, utilizando um intervalo de confiança de 95%, é apresentado na Tabela 5.8. Os resultados de todos os testes de Tukey-Kramer realizados encontram-se no Apêndice L.

Tabela 5.8 – Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.

Aspecto	Categoria	Média estimada	Grupos	
Aspecto 1	Kuwahara Anisotrópico	2,827	A	
	Flow Bilateral+FDoG	3,013		B
	Bilateral+DoG	3,118		B
	Transformada de Domínio	3,138		B
	Abordagem proposta	3,382		B
Aspecto 2	Flow Based	-	-	-
	Transformada de Domínio	-	-	-
	Bilateral+DoG	-	-	-
	Abordagem Proposta	-	-	-
	Kuwahara Anisotrópico	-	-	-
Aspecto 3	Abordagem Proposta	3,345	A	
	Kuwahara Anisotrópico	2,885		B
	Transformada de Domínio	2,761		B
	Flow Bilateral+FDoG	2,711		B
	Bilateral+DoG	2,644		B

A partir da análise dos testes de Tukey-Kramer apresentados na Tabela 5.8, pode-se verificar que não há, estatisticamente, diferença significativa entre a abordagem Kuwahara Anisotrópico e as demais consideradas no tocante ao Aspecto 1. A menor média da abordagem Kuwahara Anisotrópico evidencia a menor capacidade desta abordagem em realizar a renderização na qual há a proeminência de bordas escuras e a maior separação entre áreas de cor uniforme. As demais abordagens comparadas enquadraram-se no mesmo grupo da abordagem proposta. Todavia, a maior média obtida pela abordagem proposta reúne evidências da superioridade desta em relação às demais abordagens comparadas.

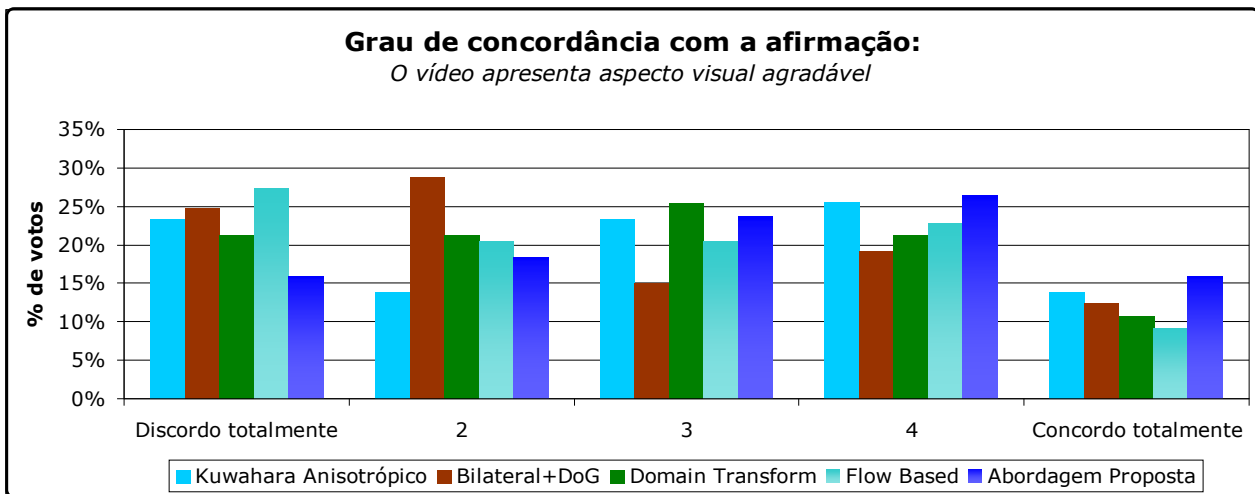
Ao se analisar o teste F ANOVA fator único, apresentado na Tabela 5.7, relativas ao Aspecto 2, pode-se verificar que não há diferença estatisticamente significativa entre nenhum dos pares de abordagens consideradas. A ausência desta diferença evidencia a capacidade da abordagem proposta em gerar renderizações não fotorrealísticas equivalentes àquelas produzidos por abordagens concorrentes, do ponto de vista da apresentação de cintilação ou artefatos espúrios.

Na avaliação do Aspecto 3, relativo à preferência visual direta entre as abordagens comparadas, pode ser verificado que há diferença estatisticamente significativa entre a abordagem proposta e todas as demais consideradas. Reitera-se que a preferência visual direta reúne evidências da superioridade da abordagem proposta em relação às demais abordagens avaliadas.

Os valores percentuais concernentes a votação do Aspecto 3, quando considerada a comparação interabordagens com os grupos intra-abordagem mais bem avaliados, estão ilustrados no gráfico de barras apresentado na Figura 5.12, obtido a partir da

utilização do método da somatória de valores (LIKERT, 1932).

Figura 5.12 – Gráfico de preferência visual direta.



Ao comparar as diversas abordagens, observa-se maior valor percentual relativo à escolha da abordagem proposta (*% de votos 4 e 5*, colunas em degradê branco-azul) em relação às demais consideradas, reiterando a maior preferência visual direta por renderizações obtidas a partir da abordagem proposta.

A seguir, são apresentadas as considerações finais do capítulo.

5.3 Considerações Finais do Capítulo

Neste capítulo, foram apresentados e discutidos resultados da investigação experimental conduzida no âmbito da pesquisa realizada, visando à complementação experimental relativa à: (i) abordagem descrita no Capítulo 3; e (b) avaliação comparativa entre abordagens destinadas à renderização NPR. Após a realização dos experimentos, conduzidos em etapas distintas da pesquisa, constatou-se que:

- A alteração da função de parada-na-aresta exerceu forte influência no grau da suavização produzida pela abordagem proposta e, conseqüentemente, na qualidade visual dos resultados. Foram consideradas três funções de parada-na-aresta, uma das quais foi escolhida em função do maior grau de abstração produzido nos resultados obtidos, quando de sua combinação com as bordas detectadas. A validação de tal escolha ocorreu em função do processo de votação em um experimento envolvendo usuários de teste.

- A calibração do detector de bordas influenciou o estilo resultante do processo de renderização. Foram considerados dois valores, relativos à sobreposição das bordas à imagem suavizada. À semelhança da escolha da função de parada-na-aresta, a validação da escolha do valor para a ponderação da sobreposição de bordas ocorreu

em função do processo de votação realizado.

- A aplicação da CEF conferiu um efeito estilístico aos resultados da abordagem proposta, tais como o realce de elementos curvilíneos, a estilização acentuada dos elementos presentes e a redução do ruído da imagem de entrada. Experimentos relativos à CEF evidenciaram a preferência visual, do ponto de vista subjetivo, por uma quantidade menos acentuada do efeito produzido. Tal preferência foi validada por meio do experimento conduzido.

- A adição da filtragem NLM, ao final da abordagem proposta, conferiu um efeito de abstração alinhado aos propósitos da pesquisa, tais como o aumento do grau de homogeneidade cromática das regiões dos quadros de vídeo avaliados e a estilização acentuada de tais regiões, atuando com preservação das bordas e do aspecto texturizado intrarregional. O tempo de execução elevado, observado no início da pesquisa (acima de 2 minutos para uma imagem de dimensões 512x512 pixels), foi reduzido por meio da utilização de uma implementação computacional alternativa. A validação dos valores para os parâmetros relativos à filtragem NLM foi realizada no experimento conduzido.

- O processo de avaliação envolvendo usuários humanos procurou evidenciar as melhorias associadas à estilização produzida com a utilização da abordagem proposta, quando considerados aspectos subjetivos de avaliação. As diferenças estatisticamente significativas foram obtidas a partir do processamento dos dados coletados do experimento, no qual foram comparadas abordagens destinadas à abstração de quadros de vídeo, ao estilo de cartum.

Resultados referentes à abordagem proposta aplicada a imagens e vídeos, renderizados de forma não fotorrealística, ao estilo de cartuns, podem ser visualizados em <https://dl.dropboxusercontent.com/u/4621189/materialAdicionalTeseFrancisco.zip>

No próximo capítulo, são formuladas as considerações finais relativas à pesquisa realizada.

Capítulo 6

Considerações Finais

Neste capítulo, expõem-se as considerações finais referentes à pesquisa realizada, na qual foi proposta uma abordagem original destinada à renderização não fotorrealística de vídeos digitais, ao estilo de cartuns. Inicialmente, é mostrada a visão contextual da pesquisa realizada. Em seguida, são destacadas contribuições advindas da realização da pesquisa. Por fim, apresentam-se sugestões para investigações futuras.

6.1 Visão Contextual da Pesquisa Realizada

Conforme destacado no Capítulo 1 deste documento, a área de NPR constitui uma área importante e em expansão, comum à Computação Gráfica (CG) e ao Processamento Digital de Imagens (PDI). As pesquisas realizadas nesta área visam à geração de renderizações que se destacam pela expressividade do resultado final, usualmente obtidas a partir da interpretação arbitrária de características presentes na imagem ou vídeo de entrada, ao invés da adesão total às propriedades realistas dos objetos da cena.

A pesquisa realizada teve como objetivo a proposição e avaliação de uma abordagem original destinada à renderização não fotorrealística de vídeos, ao estilo de cartuns. A abordagem proposta foi validada a partir de uma aplicação de software, constituída por seis módulos, capaz de gerar representações NPR de vídeos digitais.

A abordagem proposta tem como diferencial a agregação de módulos que contemplam melhoramentos na técnica típica de suavização por difusão anisotrópica, mais especificamente no que diz respeito: (i) à alteração da equação de parada-na-aresta da técnica de suavização por difusão anisotrópica original; (ii) à utilização de bordas para estilizar e orientar a suavização por difusão anisotrópica modificada; (iii) à combinação das abordagens de detecção de bordas e suavização, visando à redução do número de iterações necessárias para a obtenção de quadros do vídeo abstraído; e (iv) à inclusão de módulos destinados ao aumento da homogeneização cromática intrarregional e da saturação do resultado obtido.

A proposição da referida abordagem levou em consideração o levantamento

bibliográfico realizado (vide Capítulo 2), a partir do qual se pôde verificar a existência de duas grandes categorias de abordagens para a conversão de vídeos reais em vídeos NPR. Enquanto a primeira categoria fundamenta-se na segmentação espaçotemporal de vídeos digitais, a segunda respalda-se na suavização dos quadros do vídeo de entrada, preservando as bordas das regiões da imagem.

Abordagens que empregam a segmentação espaçotemporal dos quadros de entrada ainda representam um problema em aberto na pesquisa da área de PDI e não completamente solucionado para o caso geral, afigurando-se um tópico de pesquisa desafiador. Após a análise experimental de cinco algoritmos destinados à segmentação de imagens, em conjunção com três algoritmos destinados à computação do fluxo óptico, combinados para realizar a segmentação espaçotemporal do vídeo de entrada, *sem o auxílio do usuário*, verificou-se que os resultados obtidos foram insatisfatórios, do ponto de vista da manutenção da coerência temporal e do aspecto visual do vídeo renderizado, os quais apresentaram elevado grau de cintilação (*flickering*).

Após a investigação e a experimentação de técnicas de suavização por difusão anisotrópica, aplicada quadro-a-quadro em vídeos de teste, os resultados mostraram-se mais promissores do que aqueles obtidos com abordagens fundamentadas na segmentação espaçotemporal de vídeos. A experimentação de técnicas fundamentadas em suavização por difusão anisotrópica ensejou um estudo experimental mais aprofundado acerca de funções de parada-na-aresta diferentes daquelas originalmente propostas por Perona e Malik (1990) e por Black et al. (1998). Após tal estudo, experimentou-se a combinação da difusão anisotrópica robusta com um detector de bordas neural (ARRUDA; QUEIROZ; GOMES, 2012), a fim de verificar os benefícios que tal estratégia implicaria à abordagem proposta.

Em seguida, verificou-se a necessidade da inclusão, na abordagem, de dois módulos distintos, destinados: (i) à suavização intrarregional, com preservação de bordas e delimitação de regiões contíguas de cor; e (ii) ao aumento da saturação da imagem processada, aumentando o aspecto de cartum conferido às imagens renderizadas. A melhoria dos resultados obtidos, no tocante à renderização de regiões de textura e à boa coerência temporal entre quadros do vídeo renderizado, pôde ser constatada, a partir dos experimentos apresentados nos Capítulos 3 e 5.

Outro aspecto de destaque desta tese foi a realização de pesquisa bibliográfica relativa à avaliação de abordagens destinadas à renderização não fotorrealística de vídeos. Tal pesquisa ensejou a proposição de uma metodologia de avaliação de

abordagens destinadas à renderização não fotorrealística de vídeos, apresentada no Capítulo 4.

O processo de avaliação comparativa foi conduzido, no qual a abordagem proposta foi comparada a quatro abordagens concorrentes. Após a realização deste experimento, apresentado no Capítulo 5, procedeu-se a análise dos resultados obtidos, visando a evidenciar as melhorias associadas à estilização produzida com a utilização da abordagem proposta, quando considerados aspectos subjetivos de avaliação.

Considera-se que a pesquisa realizada não encerra a investigação dos aspectos elencados, mas propõe um aprofundamento no estudo e aponta para alternativas metodológicas destinadas à NPR. Tais alternativas devem considerar, por exemplo, aspectos relativos à melhor compreensão da opinião humana e aspectos relativos à compreensão do contexto de utilização de renderizações não fotorrealísticas.

Diante do exposto, as hipóteses definidas são rerepresentadas e discutidas. É importante salientar que a aceitação ou a rejeição de tais hipóteses considerou as seguintes restrições: (i) a amostra populacional considerada afigurou-se relevante aos objetivos pré-definidos de pesquisa; (ii) o conteúdo e as características dos vídeos apresentados durante o experimento estiveram alinhados aos objetivos da pesquisa; (iii) o agrupamento de valores de parâmetros para cada abordagem comparada foi realizada de forma a evitar vieses no processo de avaliação; (iv) o experimento subjetivo realizado não visou a abranger toda a gama de enfoques avaliatórios observados durante a pesquisa realizada, i.e., pode haver aspectos subjetivos que tenham escapado ao escopo do experimento realizado.

À luz das restrições supramencionadas e das evidências apresentadas no Capítulo 5, as hipóteses formuladas são confrontadas com as considerações decorrentes da análise dos resultados obtidos:

1. A combinação da filtragem anisotrópica com a ponderação por um detector de bordas neural, em uma abordagem destinada à renderização não fotorrealística de vídeos gera resultados que são mais bem avaliados por um conjunto de usuários, quando comparados a resultados de técnicas concorrentes.

Hipótese **aceita**. Considerando-se a prevalência da preferência visual direta pelos resultados produzidos pela abordagem proposta, com 95% de confiança, à luz dos testes de Tukey-Kramer realizados (vide Tabela 5.8, p. 162).

2. Os resultados gerados a partir da utilização da abordagem proposta para a renderização não fotorrealística contêm artefatos espúrios ou cintilação no vídeo

resultante que são equivalentes àqueles produzidos por abordagens encontradas na literatura da área.

Hipótese **aceita**. Tendo-se em vista a equivalência da abordagem proposta quando considerada a avaliação interabordagem e quando considerados os maiores valores intra-abordagem. De modo geral, os resultados da abordagem Kuwahara Anisotrópico apresentaram (ver Tabela 5.6, p. 160) menor nível de cintilação e artefatos espúrios, quando comparados aos resultados produzidos pelas demais abordagens avaliadas. Tal fato deveu-se aos aspectos inerentes à abordagem, cuja implementação não contém um módulo destinado à sobreposição de bordas à imagem renderizada. A ausência de bordas sobrepostas implicou melhor avaliação dos usuários, conforme pode ser percebido pelos relatos de aspectos negativos do experimento realizado.

3. A abordagem proposta produz resultados que possuem melhor definição de bordas e de estilização, quando comparados a abordagens encontradas na literatura da área.

Hipótese **aceita**. Considerando-se a maior média da votação do aspecto relativo à presença de bordas e delimitação de regiões contíguas de cor (vide Tabela 5.8), com 95% de confiança, à luz dos testes de Tukey-Kramer realizados.

Assim sendo, a partir da proposição da abordagem fundamentada na renderização por difusão anisotrópica modificada, da coleta de dados efetuada durante a realização do experimento relatado e da análise estatística conduzida, conclui-se que a abordagem proposta nesta pesquisa mostrou-se eficaz na geração de renderizações não fotorrealística de vídeos digitais, ao estilo de cartuns, atendendo aos objetivos de pesquisa propostos.

A seguir, são apresentadas as contribuições advindas da realização desta pesquisa. Por fim, apresentam-se sugestões de investigações que podem dar continuidade à pesquisa realizada.

6.2 Contribuições da Pesquisa

Pode-se destacar, enquanto contribuição principal da pesquisa realizada, a concepção de uma abordagem original destinada à renderização não fotorrealística automática de quadros de vídeo. Adicionalmente, podem ser elencadas outras contribuições:

- Proposição de uma combinação original de módulos destinados à renderização não fotorrealística de vídeos digitais, gerando vídeos temporalmente coerentes;
- Proposição de uma estratégia de combinação de bordas geradas por um

detector neural e suavização por difusão anisotrópica com função de parada-na-aresta modificada;

- Proposição da utilização de redes neurais no âmbito da renderização não fotorrealística de vídeos;

- Construção de um compêndio de vídeos diversificado destinados à condução de processos de avaliação subjetiva de técnicas de NPR de vídeos;

- Proposição de um processo de avaliação experimental de resultados obtidos por abordagens NPR de vídeos;

- Proposição da incorporação de um módulo que atua no incremento da saturação da imagem resultante do processo de renderização, acentuando o efeito cartum, a ser integrado a arcabouços destinados à abstração NPR de vídeos;

- Proposição da utilização da filtragem NLM no contexto da NPR, a qual conferiu maior grau de abstração aos resultados obtidos;

- Proposição de modificações ao arcabouço apresentado no estado-da-arte destinados à NPR ao estilo de cartum, as quais conferiram um aspecto diferenciado aos resultados das renderizações obtidos anteriormente.

Destaca-se que uma parte dos resultados da presente pesquisa encontra-se na seguinte publicação: (i) *Non-photorealistic neural sketching: A case study on frontal-face images*, publicada no *Journal of the Brazilian Computer Society* (ARRUDA; QUEIROZ; GOMES, 2012), referente à abordagem utilizada na geração de mapas de borda.

6.3 Investigações Futuras

Apesar dos avanços consequentes da pesquisa conduzida, verifica-se a necessidade de realizar investigações futuras, a fim de superar as dificuldades elencadas ao longo deste documento. Assim sendo, apresentam-se a seguir sugestões de investigações futuras que podem ser derivadas desta pesquisa.

Destaca-se que, apesar da possibilidade de generalização dos resultados obtidos a partir da realização do experimento comparativo, é preciso considerar as limitações das inferências realizadas. Vale ressaltar que o experimento realizado poderia contemplar maior quantidade de participantes. Todavia, além da dificuldade encontrada no processo de recrutamento de indivíduos com conhecimento alinhado aos objetivos de pesquisa, a viabilização das sessões de experimento estava condicionada à participação de usuários que tivessem conhecimento de aspectos relativos à renderização expressiva/não fotorrealística. Assim sendo, considerou-se o

universo amostral caracterizado no Capítulo 4, o qual possibilitou as inferências formuladas neste capítulo.

Diante das condições restritivas supramencionadas, sugere-se, em trabalhos futuros, a utilização de um universo amostral com quantidade superior de participantes ($n > 30$), o qual poderá permitir a formulação de inferências com maior confiança. Há a possibilidade na realização de um experimento avaliatório via Internet, sem a necessidade da realização das sessões de teste *in loco*. Entretanto, a realização do experimento via Internet ensejaria a remodelagem do processo avaliatório, e a falta de controle das condições de teste pode comprometer as inferências e coleta de dados advindos de tal tipo de experimento.

Winnemöller (2013) afirmou que, apesar da conveniência e facilidade na utilização de ferramentas de arte no formato digital, as quais reproduzem automaticamente vários estilos artísticos, tais sistemas podem retirar do usuário o controle do resultado final obtido. A automatização de um estilo artístico, encapsulado em uma ferramenta, permite que um usuário não-artista transforme uma fotografia em um artefato estilístico. Entretanto, uma ferramenta completamente automática pode não seduzir o usuário durante o processo criativo, negando a este o senso de realização.

Winnemöller (2013) reiterou que, no âmbito acadêmico, soluções automatizadas tendem a ser consideradas superiores às soluções que exigem intervenção do usuário, i.e., é possível que uma solução semiautomática apresente-se incompleta quando uma solução automática está disponível. Entretanto, uma solução que integra o usuário, permitindo a este intervir nos resultados produzidos, pode ser vantajoso, uma vez que a percepção semântica/criativa deste usuário poderia melhorar os resultados da renderização.

O envolvimento do usuário no processo de renderização não deve eximir os sistemas e as pesquisas na área de NPR de resolver problemas e desafios encontrados. Tal envolvimento do usuário pode ocorrer de forma gradual, i.e., a abordagem destinada à renderização pode produzir resultados bons o suficiente de forma autônoma, mas pode permitir a intervenção quando o usuário julgar necessário substituir o comportamento padrão do sistema. Tais intervenções devem ser, tanto quanto possível, no âmbito do processo criativo, e não focadas em tarefas repetitivas ou tediosas, nas quais o usuário age apenas enquanto corretor de erros.

Diante do exposto, afigura-se enquanto investigação futura a criação de uma ferramenta de *software* unificada que realize a transformação de vídeos reais em vídeos abstraídos de forma não fotorrealística, a qual implemente a abordagem

proposta nesta pesquisa de doutorado. Destaca-se que a pesquisa comparativa realizada foi possível uma vez que as implementações computacionais das abordagens comparadas foram disponibilizadas pelos autores. A criação e disponibilização de uma ferramenta unificada garantirá que os resultados produzidos também possam ser comparados, em pesquisas futuras.

Após a criação da ferramenta unificada, pretende-se investigar recursos de implementação que tornem possível a mitigação das limitações apresentadas no Capítulo 3. Em especial, a investigação futura procurará implementar recursos disponíveis em placas de vídeo dedicadas (GPU), o qual deverá conferir a capacidade de renderização não fotorrealística em tempo real.

Em seguida, pretende-se investigar os aspectos relativos ao envolvimento do usuário no processo de renderização, i.e., dotar a aplicação de software de capacidades relativas à parametrização dos valores para cada módulo presente na abordagem proposta. Tal parametrização deverá contemplar a modificação da nomenclatura dos parâmetros, e.g., o parâmetro (ρ) relativo à aplicação da CEF poderá ser reescrito, por exemplo, para *quantidade do efeito de estilização*.

Afigura-se enquanto investigação futura a inclusão de módulos adicionais à abordagem proposta, e.g.: (i) módulo opcional destinado ao provimento de estilos de pinceladas artísticas; e (ii) módulo destinado à inclusão de trilhas para ênfase do movimento. Outro aspecto a ser investigado é a separação entre fundo e tema em quadros de vídeo, por meio da utilização de técnica de atenção visual, visando ao reconhecimento de objetos de interesse a serem posteriormente estilizados de forma não fotorrealística.

Outra investigação futura refere-se ao aprofundamento na pesquisa relativa à filtragem NLM. Tal investigação visa à proposição de mecanismos diferenciados de comparação entre blocos de pixels. A proposição de um novo mecanismo deverá visar a renderização NPR, contraposta à redução de ruído (propósito original da filtragem NLM).

Propõe-se a realização de experimentos adicionais, relativos à combinação da abordagem proposta a abordagens destinadas à quantização cromática, à aplicação de algoritmos de difusão de erro (*dithering*).

Por fim, propõe-se o estudo dos aspectos subjetivos, por meio da realização de outros experimentos com participantes humanos, visando a melhor compreensão de preferências visuais relativas à renderização estilística, tal qual a personalização do estilo de renderização.

Referências Bibliográficas

- ACHANTA, R.; HEMAMI, S.; ESTRADA, F.; SUSSTRUNK, S. Frequency-tuned salient region detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 22, 2009, Miami, Florida, USA. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2009. p. 1597–1604.
- AGARWALA, A. SnakeToonz: A semi-automatic approach to creating cel animation from video. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 2, 2002, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2002. v. 1, p. 139–147.
- AGARWALA, A.; HERTZMANN, A.; SALESIN, D. H.; SEITZ, S. M. Keyframe-based tracking for rotoscoping and animation. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 31, 2004, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM, 2004. p. 584–591.
- ALENCAR, J. D.; QUEIROZ, J. D.; GOMES, H. An approach for non-photorealistic rendering that is appealing to human viewers. In: CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 26, 2013, Arequipa, Peru. *Proceedings...* [S.l.], 2013. p. 242–249.
- ALMERAJ, Z.; WYVILL, B.; ISENBERG, T.; GOOCH, A. A.; GUY, R. Computational aesthetics 2008: Automatically mimicking unique hand-drawn pencil lines. *Journal of Computers and Graphics*, Pergamon Press, Inc, Elmsford, NY, USA, v. 33, n. 4, p. 496–508, 2009.
- ARBELAEZ, P.; MAIRE, M.; FOWLKES, C. C.; MALIK, J. From contours to regions: An empirical evaluation. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 22, 2009, Miami, Florida, USA. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2009. p. 2294–2301.
- ARRUDA, F. A. P. V.; QUEIROZ, J. E. R.; GOMES, H. M. Non-photorealistic neural sketching - a case study on frontal-face images. *Journal of the Brazilian Computer Society*, v. 18, n. 3, p. 237-250, 2012.
- ASSUMPÇÃO JÚNIOR, J. M. *Projeto de um sistema de desvio de obstáculos para robôs móveis baseado em computação reconfigurável*. Dissertação (Mestrado em Ciência da Computação) - Instituto de Ciências Matemáticas; de Computação, Universidade de São Paulo (USP), 2009.
- BAKER, S.; SCHARSTEIN, D.; LEWIS, J. P.; ROTH, S.; BLACK, M. J.; SZELISKI, R. A database and evaluation methodology for optical flow. *International Journal in Computer Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 92, n. 1, p. 1–31, mar. 2011.
- BARILE, P.; CIESIELSKI, V.; TRIST, K. Non-photorealistic rendering using genetic programming. In: INTERNATIONAL CONFERENCE ON SIMULATED EVOLUTION AND LEARNING, 7, 2008, Melbourne, Australia. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2008. v. 5361, p. 299–308.
- BARLA, P.; BRESLAV, S.; MARKOSIAN, L.; THOLLOT, J. Interactive hatching and stippling by example. *Rapport de Recherche*, INRIA, p. 17, 2006.
- BARRON, J.L.; FLEET, D.J.; BEAUCHEMIN, S. Performance of optical flow techniques. *International Journal of Computer Vision*, v. 12, n. 1, p. 43-77, 1994.

- BÉNARD, P.; BOUSSEAU, A.; THOLLOT, J. State-of-the-art report on temporal coherence for stylized Animations. *Computer Graphics Forum*, v. 30, p. 2367–2386, 2011.
- BHAT, P.; ZITNICK, C. L.; COHEN, M.; CURLESS, B. GradientShop: A gradient-domain optimization framework for image and video filtering. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 29, n. 2, p. 1-14, 2010.
- BHOWMICK, P.; KLETTE, R. Generation of random digital simple curves with artistic emulation. *Journal of Mathematical Imaging and Vision*, v. 48, n. 1, p. 53–71, 2014.
- BLACK, M. J.; SAPIRO, G.; MARIMONT, D. H.; HEEGER, D. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, IEEE Press, Piscataway, NJ, USA, v. 7, n. 3, p. 421-432, 1998.
- BORGO, R.; CHEN, M.; DAUBNEY, B.; GRUNDY, E.; JÄNICKE, H.; HEIDEMANN, G.; HÖFERLIN, B.; HÖFERLIN, M.; WEISKOPF, D.; XIE, X. A survey on video-based graphics and video visualization. In *EuroGraphics - State of the Art Reports*, p. 1-23, 2011.
- BOUSSEAU, A.; KAPLAN, M.; THOLLOT, J.; SILLION, F. X. Interactive watercolor rendering with temporal coherence and abstraction. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 4, 2006, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2006. v. 1, p. 141–149.
- BOUSSEAU, A.; NEYRET, F.; THOLLOT, J.; SALESIN, D. Video watercolorization using bidirectional texture advection. *ACM Transactions on Graphics*, ACM Press, v. 26, n. 3, p. 104:1-104:7, 2007.
- BOYKOV, Y.; FUNKA-LEA, G. Graph cuts and efficient N-D image segmentation. *International Journal on Computer Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 70, n. 2, p. 109–131, nov. 2006.
- BOYKOV, Y.; KOLMOGOROV, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 9, p. 1124-1137, 2004.
- BRADSKI, G. R. Computer video face tracking for use in a perceptual user interface. *Intel Technology Journal - Q2*, p. 705–740, 1998.
- BRADSKI, G. R.; DAVIS, J. W. Motion segmentation and pose recognition with motion history gradients. *Machine and Vision Applied*, v. 13, n. 3, p. 174–184, 2002.
- BRADSKI, G. R.; KAEHLER, A. *Learning OpenCV*. O'Reilly Media, Inc. 2008.
- BROX T.; WEICKERT J.; BURGETH B.; MRÁZEK P. Non-linear structure tensors. *Image and Vision Computing*. v. 24, n. 1, p. 41–55, 2006.
- BRUHN, A.; WEICKERT, J.; SCHNÖRR, C. Lucas/Kanade meets Horn/Schunck: combining local and global optical flow methods. *International Journal of Computer Vision*, v. 61, n. 3, p. 211-231, 2005.
- BUADES, A.; COLL, B.; MOREL, J.M. A review of image denoising algorithms with a new one. *Multiscale Modeling and Simulation*, v. 4, p. 490–530, 2005.
- BUSSERT, L. Comic books and graphic novels: Digital resources for an evolving form of art and literature, *C&RL News*, 66(2), 2005. Último acesso [28.02.2014]. Disponível em: <<http://library.uwb.edu/staff/bussert/ComicBooks&GraphicNovels.pdf>>.
- CAPCOM. *Street Fighter IV - Capcom Co, Ltd*. 2008. Último acesso [28.02.2014]. Disponível em: <www.streetfighter.com/flash/#/sf4>.
- CARPENDALE, S. Information visualization. In: KERREN, A.; STASKO, J. T.; FEKETE, J.-D.; NORTH, C. (Ed.). Berlin, Heidelberg: Springer-Verlag, 2008. cap. Evaluating Information Visualizations, p. 19-45.
- CHUN, B.-K.; RYU, D.-S.; HWANG, W.-I.; CHO, H.-G. An automated procedure for word balloon placement in cinema comics. In: INTERNATIONAL CONFERENCE ON ADVANCES IN

- VISUAL COMPUTING, 2, 2006, Lake Tahoe, NV, USA. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2006. p. 576–585.
- COHEN, J. The earth is round ($p < .05$). *American Psychologist*, v. 49, n. 12, p. 997-1003, dez. 1994.
- COLLOMOSSE, J. P.; HALL, P. M. Genetic paint: A search for salient paintings. In: EUROPEAN CONFERENCE ON APPLICATIONS OF EVOLUTIONARY COMPUTING, 3, 2005, Lausanne, Switzerland. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2005a. v. 3449, p. 437–447.
- COLLOMOSSE, J. P.; ROWNTREE, D.; HALL, P. M. Stroke surfaces: temporally coherent artistic animations from video. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 11, n. 5, p. 540-549, set. 2005.
- COLLOMOSSE, J.; KYPRIANIDIS, J. E. Artistic stylization of images and video. *Tutorial at Eurographics*, 2011. Último acesso [28.02.2014]. Disponível em: <<http://kahlan.eps.surrey.ac.uk/EG2011/>>
- COLMAN, A. M.; NORRIS, C. E.; PRESTON, C. C. Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports*, n. 80, p. 355–362, 1997.
- COMANICIU D.; MEER, P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 5, p. 603-619, 2002.
- COUPÉ, P.; MANJÓN, J.; ROBLES, M.; COLLINS, D. Adaptive multiresolution non-local means filter for 3D magnetic resonance image denoising. *IET Image Processing*, v. 6, n. 5, p. 558–568, jul. 2012.
- CRESWELL, J. W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4. ed. [S.l.]: SAGE Publications, 2013.
- CRIMINISI, A.; SHARP, T.; ROTHER, C.; P'EREZ, P. Geodesic image and video editing. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 29, n. 5, p. 134:1-134:15, nov. 2010.
- CURTIS, C. J.; ANDERSON, S. E.; SEIMS JOSHUA E. ANDFLEISCHER, K. W.; SALESIN, D. H. Computer-generated watercolor. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 24, 1997, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM Press/Addison-Wesley Publishing Company, 1997. p. 421–430.
- DECARLO, D.; RUSINKIEWICZ, S. Highlight lines for conveying shape. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 5, 2005, San Diego, CA, USA. *Proceedings...* New York, NY, USA: ACM, 2007. p. 63–70.
- DENNIS, J. E.; WELSCH, R. E. Techniques for nonlinear least squares and robust regression. *Communications in statistics - simulation and computation*, v. 7, n. 4, p. 345-359, 1978.
- DOLNICAR, S.; GRUN, B.; LEISCH, F.; ROSSITER, J. Three good reasons not to use five and seven point Likert items. In: CAUTHE NATIONAL CONFERENCE, 21, 2011, Adelaide, Australia. *Proceedings...* [S.l.], 2011. p. 8–11.
- DUKE, D. J.; BARNARD, P. J.; HALPER, N.; MELLIN, M. Rendering and affect. *Computer Graphics Forum*, v. 22, n. 3, p. 359-368, 2003.
- EAKIN, E. Pow! Splat! Take that, you Darwin disparagers! *The New York Times*, p. 11, 2002. Último acesso [28.02.2014]. Disponível em: <<http://www.nytimes.com/2002/11/30/arts/think-tank-pow-splat-take-that-you-darwin-disparagers.html>>.

- FARNEBÄCK, G. Two-frame motion estimation based on polynomial expansion. In: SCANDINAVIAN CONFERENCE ON IMAGE ANALYSIS, 13, 2003, Halmstad, Sweden. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2003. p. 363–370.
- FELIZ, D. A.; LAIDLAW, D. H.; DRURY, F. Using visual design expertise to characterize the effectiveness of 2D scientific visualization methods. *IEEE Visualization*. [S.l.]: IEEE Computer Society, 2005. p. 101.
- FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *International Journal on Computer Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 59, n. 2, p. 167-181, set. 2004.
- FIELD, A. P.; HOLE, G. Book; Book/Illustrated. *How to design and report experiments*. [S.l.]: London ; Thousand Oaks, Calif. : Sage publications Ltd, 2003.
- FOX. *Waking Life - Fox Searchlight Pictures*. 2001. Último acesso [28.02.2014]. Disponível em: <www.imdb.com/title/tt0243017/>.
- GAMBARRA, L. L.; LIMA, J. A. G. de; BATISTA, L. V.; MARQUES, D. S.; SILVA, H. S. da. Otimização do algoritmo Non-local Means utilizando uma implementação em FPGA. In: INTERNATIONAL IBERCHIP WORKSHOP, 18, 2012, Playa del Carmen, Mexico. *Proceedings...* New York, NY, USA: Elsevier, 2012. p. 121–125.
- GASTAL, E. S. L.; OLIVEIRA, M. M. Domain Transform for edge-aware image and video processing. *ACM Transactions on Graphics*, v. 30, n. 4, p. 1-6, 2011.
- GATZIDIS, C.; PAPAKONSTANTINO, S.; BRUJIC-OKRETIC, V.; BAKER, S. Recent advances in the user evaluation methods and studies of non-photorealistic visualization and rendering techniques. In: INTERNATIONAL CONFERENCE ON INFORMATION VISUALISATION, 12, 2008, London, UK. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2008. p. 475–480.
- GEARBOX. *Borderlands - Gearbox Software*. 2009. Último acesso [28.02.2014]. Disponível em: <<http://www.gearboxsoftware.com/games/borderlands/>>.
- GENG, W. *The Algorithms and Principles of Non-photorealistic Graphics: Artistic Rendering and Cartoon Animation*. 1. ed. [S.l.]: Springer, 2010.
- GERSTNER, T.; DECARLO, D.; ALEXA, M.; FINKELSTEIN, A.; GINGOLD, Y.; NEALEN, A. Pixelated image abstraction. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 10, 2012, Annecy, France. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2012. p. 29–36.
- GERVAUTZ, M.; PURGATHOFER, W. *Graphics gems*. In: GLASSNER, A. S. (Ed.). San Diego, CA, USA: Academic Press Professional, Inc, 1990. cap. A Simple Method for Color Quantization: Octree Quantization, p. 287–293.
- GLM, Graphics & Media Lab. *Moscow State University Perceptual Video Quality Tool*, 2013. Último acesso [28.02.2014]. Disponível em: <compression.ru/video/quality_measure/subjective_metrics_info_en.html>.
- GOMES, H. M. *Model Learning in Iconic Vision*. Tese (Doutorado) — University of Edinburgh, 2002.
- GOMES, R. B.; OLIVEIRA, L. M.; NETO, L. S. B.; SANTOS, T. S.; ANDRADE, G. S.; CARVALHO, B. M.; GONCALVES, L. M. G. Producing stylized videos using the AnimVideo rendering tool. *International Journal on Imaging Systems and Technology*, V. 19, N. 2, P. 100–110, 2009.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. [S.l.]: Prentice-Hall, 2007.
- GOOCH, A. A.; LONG, J.; JI, L.; ESTEY, A.; GOOCH, B. S. Viewing progress in non-photorealistic rendering through Heinlein's lens. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 8, 2010, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2010. v. 1, p. 165–171.

- 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. *Human Facial Illustrations: Creation and Evaluation using Behavioral Studies and Functional Magnetic Resonance Imaging*. Tese (Doutorado). University of Utah, 2003.
- GOOCH, B.; GOOCH, A. *Non-Photorealistic Rendering*. [S.l.]: A. K. Peters, Ltd, 2001.
- GOOCH, B.; REINHARD, E.; GOOCH, A. Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 23, n. 1, p. 27-44, jan. 2004.
- GOOSSENS, B.; LUONG, Q.; AELTERMAN, J.; PIZURICA, A.; PHILIPS, W. A GPU-accelerated real-time NLMeans algorithm for denoising color video sequences. In: TALON, J. B.; BONE, D.; PHILIPS, W.; POPESCU, D.; SCHEUNDERS, P. (Ed.). *Lecture Notes in Computer Science*. [S.l.]: Springer, 2010. v. 6475, p. 46-57.
- GREENBERG, S.; BUXTON, B. Usability evaluation considered harmful (some of the time). In: SIGCHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 26, 2008, Florence, Italy. *Proceedings...* New York, NY, USA: ACM, 2008. v. 1, p. 111-120.
- HAEBERLI, P. Paint by numbers: Abstract image representations. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 17, 1990, Dallas, TX, USA. *Proceedings...* New York, NY, USA: ACM, 1990. v. 24, n. 4, p. 207-214.
- HAGGERTY, M. Almost automatic computer painting. *IEEE Computer Graphics and Applications*, IEEE Computer Society, Los Alamitos, CA, USA, v. 11, n. 6, p. 11-12, 1991.
- HALL, P. M.; COLLOMOSSE, J. P.; SONG, Y.-Z.; SHEN, P.; LI, C. RTCams: A new perspective on non-photorealistic rendering from photographs. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 5, p. 966-979, set. 2007.
- HALPER, N.; MELLIN, M.; HERRMANN, C.; LINNEWEBER, V.; STROTHOTTE, T. Psychology and non-photorealistic rendering: The beginning of a beautiful relationship. In: SZWILLUS, G.; ZIEGLER, J. (Ed.). *Mensch & Computer 2003*. [S.l.]: Vieweg+Teubner Verlag, 2003a, (Berichte des German Chapter of the ACM, v. 57). p. 277-286.
- HALPER, N.; MELLIN, M.; HERRMANN, C.; LINNEWEBER, V.; STROTHOTTE, T. Towards an understanding of the psychology of non-photorealistic rendering. In: SCHNEIDER, J.; STROTHOTTE, T.; MAROTZKI, W. (Ed.). *Computational Visualistics, Media Informatics, and Virtual Communities*. [S.l.]: Deutscher Universitätsverlag, 2003b, (Bildwissenschaft, v. 11). p. 67-78.
- HAYS, J.; ESSA, I. Image and video based painterly animation. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 3, 2004, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2004. p. 113-120.
- HECKBERT, P. Color image quantization for frame buffer display. *SIGGRAPH Computer and Graphics*, ACM, New York, NY, USA, v. 16, n. 3, p. 297-307, jul. 1982.
- HERTZMANN, A. Non-photorealistic rendering and the science of art. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 8, 2010, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2010. p. 147-157.
- HERTZMANN, A. Paint by relaxation. In: COMPUTER GRAPHICS INTERNATIONAL, 19, 2001, Hong Kong, China. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2001. p. 47-54.
- HERTZMANN, A. Painterly rendering with curved brush strokes of multiple sizes. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 25, 1998, Orlando, FL, USA. *Proceedings...* New York, NY, USA: ACM, 1998. v. 1, p. 453-460.

- HERTZMANN, A.; PERLIN, K. Painterly rendering for video and interaction. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 1, 2000, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2000. p. 7–12.
- HOGG, R. V.; TANIS, E. A. *Probability and statistical inference*. 4. ed. [S.l.]: MacMillan, 1993.
- HONG, R.; YUAN, X.-T.; XU, M.; WANG, M.; YAN, S.; CHUA, T.-S. Movie2Comics: A feast of multimedia artwork. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA, 18, 2010, Firenze, Italy. *Proceedings...* New York, NY, USA: ACM, 2010. p. 611–614.
- HONG, S.-S.; YOON, J.-C.; LEE, I.-K. e BYUN, S. Interactive system for efficient video cartooning. In *Collaboration techniques: Third international Conference in Computer Vision/Computer Graphics - MIRAGE*, 2007, p. 161-172.
- HORN, B.K.P.; SCHUNCK, B.G. Determining optical flow. *Artificial Intelligence*, vol 17, pp 185-203, 1981.
- HOUAISS, A. *Dicionário Houaiss da Língua Portuguesa*. Ed. Objetiva, 2001.
- HUANG, P. *Video special effects*. Dissertação (Mestrado). University of Surrey, Department of Electronic Engineering, School of Electronics and Physical Science, Guildford, Surrey, UK, 2006.
- HURTUT, T.; LANDES, P.-E.; THOLLOT, J.; GOUSSEAU, Y.; DROUILLHET, R.; COEURJOLLY, J.-F. Appearance-guided synthesis of element arrangements by example. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 7, 2009, New Orleans, Louisiana. *Proceedings...* New York, NY, USA: ACM, 2009. p. 51–60.
- HWANG, W.-I.; LEE, P.-J, CHUN, B.-K.; RYU, D.-S.; CHO, H.-G. Cinema comics: cartoon generation from video stream. In *Computer Graphics Theory and Applications*, 2006.
- IANEVA, T. I.; VRIES, A. P. de; ROHRIG, H. Detecting cartoons: A case study in automatic video-genre classification. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO, 2003, 2003, Baltimore, Maryland, USA. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2003. v. 2, p. 449–452.
- IGARASHI, T.; MATSUOKA, S.; TANAKA, H. Teddy: A sketching interface for 3D freeform design. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 26, 1999, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM Press/Addison-Wesley Publishing Company, 1999. v. 1, p. 409–416.
- INGLIS, T. C.; KAPLAN, C. S. Generating op-art lines. In: INTERNATIONAL SYMPOSIUM ON COMPUTATIONAL AESTHETICS IN GRAPHICS, VISUALIZATION, AND IMAGING, 7, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 25–32.
- INTERNATIONAL TELECOMMUNICATION UNION. *ITU-R BT.500-12: Methodology for the subjective assessment of the quality of television pictures*. Geneva, Switzerland, 2012. Disponível em: <<http://www.itu.int/rec/R-REC-BT.500/en>>.
- ISENBERG, T. Evaluating and validating non-photorealistic and illustrative rendering. In: ROSIN, P.; COLLOMOSSE, J. (Ed.). *Image and Video-Based Artistic Stylisation*. [S.l.]: Springer London, 2013, (Computational Imaging and Vision, v. 42). p. 311-331.
- ISENBERG, T.; NEUMANN, P.; CARPENDALE, S.; SOUSA, M. C.; JORGE, J. A. Non-photorealistic rendering in context: An observational study. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 4, 2006, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2006. p. 115–126.
- JACKSON, C. D.; ACEVEDO, D.; LAIDLAW, D. H.; DRURY, F.; VOTE, E.; KEEFE, D. Designer-critiqued comparison of 2D vector visualization methods: A pilot study. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES: SKETCHES & APPLICATIONS, 30, 2003, San Diego, CA. *Proceedings...* New York, NY, USA: ACM, 2003. p. 1–10.

- KAGAYA, M.; BRENDDEL, W.; DENG, Q.; KESTERSON, T.; TODOROVIC, S.; NEILL, P. J.; ZHANG, E. Video painting with space-time-varying style parameters. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society Press, v. 17, n. 1, p. 74-87, 2010.
- KAKALIOS, J. Adding pow! to your physics class with comic-book lessons, *Curriculum Review*, p. 14-15, out, 2002.
- KALMAN, R. E. A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, n. 82, p. 35-45, 1960.
- KANG, D.; OHN, Y.-J.; HAN, M.-H.; YOON, K.-H. Animation for ancient tile mosaics. In: EUROGRAPHICS SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 9, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 157-166.
- KANG, H.; LEE, S. Shape-simplifying image abstraction. *Computer Graphics Forum*, v. 27, n. 7, p. 1773-1780, 2008.
- KANG, H.; LEE, S.; CHUI, C. K. Coherent line drawing. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 5, 2007, San Diego, CA, USA. *Proceedings...* New York, NY, USA: ACM, 2007. p. 43-50.
- KANG, H.; LEE, S.; CHUI, C. K. Flow-based image abstraction. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 15, n. 1, p. 62-76, jan. 2009.
- KAPTEIN, M.; ROBERTSON, J. Rethinking statistical analysis methods for CHI. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 30, 2012, Austin, Texas, USA. *Proceedings...* New York, NY, USA: ACM, 2012. p. 1105-1114.
- KEEFE, D. F.; KARELITZ, D. B.; VOTE, E. L.; LAIDLAW, D. H. Artistic collaboration in designing VR visualizations. *IEEE Computer Graphics and Applications*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 25, n. 2, p. 18-23, mar. 2005.
- KENDRICKS, N. Comic-book artist uses work to take serious look at how hard life can be. *The San Diego Union Tribune*, p. E-8, 2000.
- KERLINGER, F. N.; LEE, H. B. *Foundations of behavioral research*. [S.l.]: Wadsworth, 2000.
- KIM, B.; ESSA, I. Video-based non-photorealistic and expressive illustration of motion. In *Computer Graphics International*, p. 32-35, 2005.
- KIM, D.; SON, M.; LEE, Y.; KANG, H.; LEE, S. Feature-guided image stippling. In: EUROGRAPHICS CONFERENCE ON RENDERING, 19, 2008, Sarajevo, Bosnia and Herzegovina. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2008. v. 27, p. 1209-1216.
- KIM, H. Y. *Projeto de Operadores pela Aprendizagem, Difusao Anisotropica e Marca d'água de Autenticação*. Tese (livre-docência). Escola Politécnica, USP, 2004.
- KLEIN, A. W.; SLOAN, P.-P. J.; FINKELSTEIN, A.; COHEN, M. F. Stylized video cubes. In: SYMPOSIUM ON COMPUTER ANIMATION, 29, 2002, San Antonio, Texas, USA. *Proceedings...* New York, NY, USA: ACM, 2002. p. 15-22.
- KRISSIAN, K.; WESTIN, C. F.; KIKINIS, R.; VOSBURGH, K. G. Oriented speckle reducing anisotropic diffusion. *Transactions on Image Processing*, IEEE Press, Piscataway, NJ, USA, v. 16, n. 5, p. 1412-1424, 2007.
- KROON, D.-J.; SLUMP, C. H.; MAAL, T. J. J. Optimized anisotropic rotational invariant diffusion scheme on cone-beam CT. In: INTERNATIONAL CONFERENCE ON MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION: PART III, 13, 2010, Beijing, China. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2010. p. 221-228.

- KUWAHARA M.; HACHIMURA K.; EIHO S.; KINOSHITA M. *Digital processing of biomedical images*. Plenum Press, p. 187–203, 1976.
- KYPRIANIDIS, J. E. Image and video abstraction by multi-scale anisotropic Kuwahara filtering. In: EUROGRAPHICS SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 9, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 55–64.
- KYPRIANIDIS, J. E. *Structure adaptive stylization of images and video*. Tese (Doutorado). University of Potsdam, 2013.
- KYPRIANIDIS, J. E.; COLLOMOSSE, J.; WANG, T.; ISENBERG, T. State of the 'art': A taxonomy of artistic stylization techniques for images and video. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 19, n. 5, p. 866-885, 2013.
- KYPRIANIDIS, J. E.; DÖLLNER, J. Image abstraction by structure adaptive filtering. In: EUROGRAPHICS UK THEORY AND PRACTICE OF COMPUTER GRAPHICS, 3, 2008, Manchester, UK. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2008. p. 51–58.
- KYPRIANIDIS, J. E.; KANG, H. Image and video abstraction by coherence-enhancing filtering. *Computer Graphics Forum*, v. 30, n. 2, p. 593-602, 2011.
- KYPRIANIDIS, J. E.; KANG, H.; DÖLLNER, J. Image and video abstraction by anisotropic Kuwahara filtering. *Computer Graphics Forum*, Wiley-Blackwell, v. 28, n. 7, p. 1955-1963, 2009.
- LAIDLAW, D. H.; KIRBY, R. M.; JACKSON, C. D.; DAVIDSON, J. S.; MILLER, T. S.; SILVA, M. da; WARREN, W. H.; TARR, M. J. Comparing 2D vector and visualization methods: a user study. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 11, n. 1, p. 59-70, jan. 2005.
- LAKE, A.; MARSHALL, C.; HARRIS, M.; BLACKSTEIN, M. Stylized rendering techniques for scalable real-time 3D animation. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 1, 2000, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2000. p. 13–20.
- LEE, H.; KWON, S.; LEE, S. Real-time pencil rendering. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 4, 2006, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2006. p. 37–45.
- LEE, H.-C.; LEE, C.-H.; YOON, K. H. Motion based painterly rendering. *Computer Graphics Forum*, v. 28, n. 4, p. 1207-1215, 2009.
- LEMPITSKY, V. S.; ROTH, S.; ROTHER, C. FusionFlow: Discrete-continuous optimization for optical flow estimation. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 21, 2008, Anchorage, AL, USA. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2008.
- LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. *Estatística: teoria e aplicações*. 1. ed. [S.l.]: LTC, 2000. ISBN 85-216-1211-7.
- LI, H.; MOULD, D. Structure-preserving stippling by priority-based error diffusion. In: GRAPHICS INTERFACE CONFERENCE, 16, 2011, St. John's, Newfoundland, Canada. *Proceedings...* School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada: Canadian Human-Computer Communications Society, 2011. p. 127–134.
- LI, H.; MOULD, D.; DAVIES, J. Structure and aesthetics in non-photorealistic images. In: GRAPHICS INTERFACE CONFERENCE, 18, 2013, Regina, Saskatchewan, Canada. *Proceedings...* Toronto, Ontario, Canada: Canadian Information Processing Society, 2013. p. 181–188.

- LI, W.; AGRAWALA, M.; CURLESS, B.; SALESIN, D. Automated generation of interactive 3D exploded view diagrams. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 27, n. 3, p. 101:1–101:7, 2008.
- LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology*, v. 22, n. 140, p. 1–55, 1932.
- LING, U.; YAN, C.; LIU, C.; WANG, X.; LI, H. Adaptive tone-preserved image detail enhancement. *The Visual Computer*, v. 28, n. 6, p. 733–742, 2012.
- LITWINOWICZ, P. Processing images and video for an impressionist effect. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 24, 1997, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM Press/Addison-Wesley Publishing Company, 1997. p. 407–414.
- LIU, C.; TORRALBA, A.; FREEMAN, W. T.; DURAND, F.; ADELSON, E. H. Motion magnification. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 24, n. 3, p. 519–526, jul. 2005.
- LUCAS, B. D.; KANADE, T. An iterative image registration technique with an application to stereo vision. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 7, 1981, Vancouver, BC, Canada. *Proceedings...* San Francisco, CA, USA: Morgan Kaufmann Publishers, 1981. v. 2, p. 674–679.
- LUFT, T.; COLDITZ, C.; DEUSSEN, O. Image enhancement by unsharp masking the depth buffer. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 25, n. 3, p. 1206–1213, 2006.
- LUFT, T.; KOBIS, F.; ZINSER, W.; DEUSSEN, O. Watercolor illustrations of cad data. In: EUROGRAPHICS CONFERENCE ON COMPUTATIONAL AESTHETICS IN GRAPHICS, VISUALIZATION AND IMAGING, 4, 2008, Lisbon, Portugal. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2008. p. 57–63.
- LUM, E. B.; MA, K.-L. Non-photorealistic rendering using watercolor inspired textures and illumination. In *Ninth Pacific Conference on Computer Graphics and Applications*, 2001. p. 322–330.
- MACIEJEWSKI, R.; ISENBERG, T.; ANDREWS, W. M.; EBERT, D. S.; SOUSA, M. C.; CHEN, W. Measuring stipple aesthetics in hand-drawn and computer-generated images. *IEEE Computer Graphics and Applications*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 28, n. 2, p. 62–74, mar. 2008.
- MANDRYK, R. L.; MOULD, D.; LI, H. Evaluation of emotional response to non-photorealistic images. In: ACM SIGGRAPH/EUROGRAPHICS SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 9, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 7–16.
- MANTIUK, R. K.; TOMASZEWSKA, A.; MANTIUK, R. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, John Wiley & Sons, Inc, New York, NY, USA, v. 31, n. 8, p. 2478–2491, 2012.
- MARKOSIAN, L. *Art-based modeling and rendering for computer graphics*. Tese (Doutorado), Brown University, 2000.
- MARKOSIAN, L.; KOWALSKI, M. A.; GOLDSTEIN, D.; TRYCHIN, S. J.; HUGHES, J. F.; BOURDEV, L. D. Real-time nonphotorealistic rendering. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 24, 1997, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM Press/Addison-Wesley Publishing Company, 1997. p. 415–420.
- MARONNA, R.A.; MARTIN, D.R.; YOHAI, V. J. *Robust statistics: theory and methods*. Wiley Series in Probability and Statistics. 2006.
- MARTIN, D. R.; FOWLKES, C.; TAL, D.; MALIK, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological

statistics. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 8, 2001, Vancouver, Canada. *Proceedings...* Los Alamitos, CA, USA: IEEE Computer Society Press, 2001. v. 2, p. 416–423.

MARTÍN, D.; ARROYO, G.; LUZÓN, M. V.; ISENBERG, T. Scale-dependent and example-based grayscale stippling. *Computer and Graphics*, v. 35, n. 1, p. 160-174, 2011.

MCCANE, B.; NOVINS, K.; CRANNITCH, D.; GALVIN, B. On benchmarking optical flow. *Computer Vision and Image Understanding*, v. 84, n. 1, p. 126-143, 2001.

MCCLLOUD, S.; *Understanding comics: the invisible art*. Northampton, MA: Kitchen Sink Press. 1993.

MEDHI, I.; PATABANDHIGE, C. R.; TOYAMA, K. Cartoon generation for text-free user interfaces. In: INTERNATIONAL WORKSHOP ON COMPUTER VISION APPLICATIONS FOR DEVELOPING REGIONS, 1, 2007, Rio de Janeiro, Brazil. *Proceedings...* [S.l.], 2007.

MEIER, B. J. Painterly rendering for animation. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 23, 1996, New Orleans, LA, USA. *Proceedings...* New York, NY, USA: ACM, 1996. p. 477–484.

MELLO, V. B.; JUNG, C. R.; WALTER, M. Virtual woodcuts from images. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES IN AUSTRALIA AND SOUTHEAST ASIA, 5, 2007, Perth, Australia. *Proceedings...* New York, NY, USA: ACM, 2007. p. 103–109.

MENDRIK, A.; VONKEN, E.-J.; RUTTEN, A.; VIERGEVER, M. A.; GINNEKEN, B. Noise reduction in computed tomography scans using 3D anisotropic hybrid diffusion with continuous switch. *IEEE Transactions on Medical Imaging*, v. 28, n. 10, p. 1585–1594, 2009.

MI, X.; DECARLO, D.; STONE, M. Abstraction of 2D shapes in terms of parts. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 7, 2009, New Orleans, Louisiana, USA. *Proceedings...* New York, NY, USA: ACM, 2009. p. 15–24.

MOULD, D. Stipple placement using distance in a weighted graph. In: EUROGRAPHICS CONFERENCE ON COMPUTATIONAL AESTHETICS IN GRAPHICS, VISUALIZATION AND IMAGING, 3, 2007, Alberta, Canada. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2007. p. 45–52.

MOULD, D. Texture-preserving abstraction. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 10, 2012, Annecy, France. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2012. p. 75–82.

MOULD, D.; MANDRYK, R. L.; LI, H. Emotional response and visual attention to non-photorealistic images. *Computer and Graphics*, v. 36, n. 6, p. 658-672, 2012.

NIENHAUS, M.; DÖLLNER, J. Sketchy drawings. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS, VIRTUAL REALITY, VISUALISATION AND INTERACTION IN AFRICA, 3, 2004, Cape Town, South Africa. *Proceedings...* New York, NY, USA: ACM, 2004. p. 73–81.

O'DONOVAN, P.; MOULD, D. Felt-based rendering. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 4, 2006, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2006. p. 55–62.

OPENCV. Intel® Open Computer Vision Library. Disponível em: <<http://sourceforge.net/projects/opencvlibrary/>>, 2006. Acesso em 01/05/2012.

ORZAN, A.; BOUSSEAU, A.; BARLA, P.; THOLLOT, J. Structure-preserving manipulation of photographs. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 5, 2007, San Diego, CA. *Proceedings...* New York, NY, USA: ACM, 2007. p. 103–110.

- OSHER, S.; RUDIN, L. I. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, Society for Industrial and Applied Mathematics, v. 27, n. 4, p. 919–940, 1990.
- OSTROMOUKHOV, V.; HERSCH, R. D. Multi-color and artistic dithering. In: ANNUAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 26, 1999, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM Press/Addison-Wesley Publishing Company, 1999. p. 425–432.
- PALETTECAD. *PaletteCAD GmbH - 3D design software for interior design*. 2013. Último acesso [28.02.2014]. Disponível em: <www.palettecad.com>.
- PALMA, G.; PICCIALI, F.; MICHELE, P. D.; CUOMO, S.; COMERCI, M.; BORRELLI, P.; ALFANO, B. 3D non-local means denoising via multi-GPU. In: FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS,, 5, 2013, Krakow, Poland. *Proceedings...* Los Alamitos, CA, USA: IEEE Computer Society Press, 2013. p. 495–498.
- PAPARI, G.; PETKOV, N.; CAMPISI, P. Artistic edge and corner enhancing smoothing. *IEEE Transactions on Image Processing*, IEEE Press, Piscataway, NJ, USA, v. 16, n. 10, p. 2449–2462, 2007.
- PARIS, S.; DURAND, F. A topological approach to hierarchical segmentation using mean-shift. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. p. 18–23.
- PARK, S.-J.; SHIN, J.-K.; LEE, M. Biologically inspired saliency map model for bottom-up visual attention. In: INTERNATIONAL WORKSHOP ON BIOLOGICALLY MOTIVATED COMPUTER VISION, 2, 2002, Tübingen, Germany. *Proceedings...* London, UK: Springer-Verlag, 2002. p. 418–426.
- PEREIRA, L. F. *Efeito de desenho com GIMP*. 2008. Último acesso [28.02.2014]. Disponível em: <luygimp.deviantart.com/art/efeito-desenho-92972578>.
- PERONA, P.; MALIK, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Washington, DC, USA, v. 12, n. 7, p. 629–639, 1990.
- POLYGRAM. *What Dreams May Come - Polygram Filmed Entertainment*. 1998. Último acesso [28.02.2014]. Disponível em: <www.imdb.com/title/tt0120889/>.
- POULI, T.; REINHARD, E. Extended papers from NPAR 2010: Progressive color transfer for images of arbitrary dynamic range. *Computer and Graphics*, Pergamon Press, Inc, Elmsford, NY, USA, v. 35, n. 1, p. 67–80, fev. 2011.
- PROESMANS, M.; VAN GOOL, L.; PAUWELS E.; OOSTERLINCK, A. Determination of optical flow and its discontinuities using non-linear diffusion. In *3rd European Conference on Computer Vision*, 1994. p. 295–304.
- PROTTER, M.; ELAD, M.; TAKEDA, H.; MILANFAR, P. Generalizing the non-local means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, v. 18, n. 1, p. 36–51, 2009.
- RAMPONI, G. A rational edge-preserving smoother. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. *Proceedings...* [S.l.: s.n.], 1995. v. 1, p. 151–154.
- RAUTEK, P.; BRUCKNER, S.; GRÖLLER, E.; VIOLA, I. Illustrative visualization: new technology or useless tautology? *SIGGRAPH Computer and Graphics*, ACM, New York, NY, USA, v. 42, n. 3, p. 4:1–4:8, 2008.
- REYNOLDS, D. A. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, p. 91–108, 1995.
- RICHEMS, P. Feature-guided image stippling. In: INTERNATIONAL CONFERENCE ON COMPUTER AIDED ARCHITECTURAL DESIGN FUTURES, 8, 1999, Atlanta, USA. *Proceedings...* Dordrecht, Netherlands: Kluwer Academic Publisher, 1999. p. 381–398.

- ROBSON, L. S.; SHANNON, H. S.; GOLDENHAR, L. M.; HALE, A. R. *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to show whether a safety intervention really works*. [S.l.]: Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2001.
- ROCKSTAR. *Grand Theft Auto V - Rockstar Games*. 2013. Último acesso [28.02.2014]. Disponível em: <www.rockstargames.com/V>.
- ROSIN, P.; COLLOMOSSE, J. *Image and Video-Based Artistic Stylisation*. [S.l.]: Springer, 2013.
- RYU, D.-S.; PARK, S.-H.; LEE, J.-w.; LEE, D.-H.; CHO, H.-G. CineToon: A semi-automated system for rendering black/white comic books from video streams. In: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY WORKSHOPS, 8, 2008, Khulna, Bangladesh. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2008. p. 336–341.
- SAINT-MARC, P.; CHEN, J.-S.; MEDIONI, G. Adaptive smoothing: a general tool for early vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 13, n. 6, p. 514–529, 1991.
- SALESIN, D. H. *Non-photorealistic animation and rendering: 7 grand challenges*. 2002. Keynote talk at the Second International Symposium on Non-Photorealistic Animation and Rendering - Annecy, France, jun 3-5.
- SAND, P.; TELLER, S. Particle video: long-range motion estimation using point trajectories. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006. p. 2195-2202.
- SANTELLA, A. *The art of seeing: bisual perception in design and evaluation of non-photorealistic Rendering*. Tese (Doutorado), Rutgers University, New Brunswick, New Jersey, USA, 2005.
- SANTELLA, A.; DECARLO, D. Abstracted painterly renderings using eye-tracking data. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 2, 2002, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2002. p. 75–ff.
- SANTELLA, A.; DECARLO, D. Visual interest and NPR: An evaluation and manifesto. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 3, 2004, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2004. p. 71–150.
- SAUVAGET, C.; BOYER, V. Abstraction of photographs: A comics style approach. In: INTERNATIONAL CONFERENCE ON EUROPEAN COMPUTING CONFERENCE, 3, 2009, Tbilisi, Georgia. *Proceedings...* Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society, 2009. p. 215–220.
- SAWADA, T.; TOYOURA, M.; MAO, X. Film comic generation with eye tracking. In: LI, S.; SADDIK, A.; WANG, M.; MEI, T.; SEBE, N.; YAN, S.; HONG, R.; GURRIN, C. (Ed.). *Advances in Multimedia Modeling*. [S.l.]: Springer Berlin Heidelberg, 2013, (Lecture Notes in Computer Science, v. 7732). p. 467–478.
- SAYEED, R.; HOWARD, T. State-of-the-art of non-photorealistic rendering for visualisation. In: EUROGRAPHICS UK THEORY AND PRACTICE OF COMPUTER GRAPHICS, 24, 2006, Middlesbrough, UK. *Proceedings...* [S.l.]: Pergamon Press, 2006. p. 1–10.
- SCHLECHTWEG, S.; GERMER, T.; STROTHOTTE, T. RenderBots: multi-agent systems for direct image generation. *Computer Graphics Forum*, Blackwell Publishing Ltd, v. 24, n. 2, p. 137–148, 2005.
- SCHMIDT, R.; ISENBERG, T.; JEPP, P.; SINGH, K.; WYVILL, B. Sketching, scaffolding, and inking: A visual history for interactive 3D modeling. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 5, 2007, San Diego, CA, USA. *Proceedings...* New York, NY, USA: ACM, 2007. p. 23–32.

- SCHUMANN, J.; STROTHOTTE, T.; LASER, S.; RAAB, A. Assessing the effect of non-photorealistic rendered images in CAD. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 13, 1996, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 1996. p. 35–41.
- SECORD, A. Weighted Voronoi stippling. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 2, 2002, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2002. p. 37–43.
- SESHADRINATHAN, K.; SOUNDARARAJAN, R.; BOVIK, A. C.; CORMACK, L. K. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, IEEE Press, Piscataway, NJ, USA, v. 19, n. 6, p. 1427–1441, 2010.
- SETLUR, V.; WILKINSON, S. Automatic stained glass rendering. In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTER GRAPHICS, 24, 2006, Hangzhou, China. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2006. p. 682–691.
- SHET, R. N.; LAI, K. H.; EDIRISINGHE, E. A.; CHUNG, P. W. H. Use of neural networks in automatic caricature generation: An approach based on drawing style capture. In: IBERIAN CONFERENCE ON PATTERN RECOGNITION AND IMAGE ANALYSIS, 2, 2005, Estoril, Portugal. *Proceedings...* Berlin, Heidelberg: Springer-Verlag, 2005. p. 343–351.
- SHI, J.; MALIK, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, p. 888–905, 2000.
- SKETCHUP. *SketchUp - 3D for Everyone*. 2000. Último acesso [28.02.2014]. Disponível em: <sketchup.google.com>.
- SNAVELY, N.; ZITNICK, C. L.; KANG, S. B.; COHEN, M. Stylizing 2.5-D video. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 4, 2006, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2006. p. 63–69.
- SOHN, A. *Oxygen television series - Avenue Amy*. 1999. Último acesso [28.02.2014]. Disponível em: <www.amysohn.com/etc/multi-media/>.
- SON, M.; LEE, Y.; KANG, H.; LEE, S. Structure grid for directional stippling. *Graphical Models*, v. 73, n. 3, p. 74–87, 2011.
- SONG, Y.-Z.; ROSIN, P. L.; HALL, P. M.; COLLOMOSSE, J. Arty shapes. In: EUROGRAPHICS CONFERENCE ON COMPUTATIONAL AESTHETICS IN GRAPHICS, VISUALIZATION AND IMAGING, 4, 2008, Lisbon, Portugal. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 2008. p. 65–72.
- STRASSMANN, S. Hairy brushes. *SIGGRAPH Computer Graphics*, ACM, New York, NY, USA, v. 20, n. 4, p. 225–232, ago. 1986.
- STREIT, L. M. *Importance-driven halftoning*. Dissertação (Mestrado) Faculty of Graduate Studies and Research - University of Alberta, Edmonton, Alta, Canada, 2009. UML Order No. GAXMQ-34422.
- STROTHOTTE, T.; PREIM, B.; RAAB, A.; SCHUMANN, J.; FORSEY, D. R. How to render frames and influence people. *Computer Graphics Forum*, v. 13, n. 3, p. 455–466, 1994.
- STROTHOTTE, T.; SCHLECHTWEG, S. *Non-photorealistic Computer Graphics: Modeling, Rendering and Animation*. [S.l.]: Morgan Kaufmann, 2002.
- SUN, D.; ROTH, S.; BLACK, M. Secrets of optical flow estimation and their principles. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 23, 2010, San Francisco, CA, USA. *Proceedings...* [S.l.], 2010. p. 2432–2439.
- TABIK, S.; GARZON, E. M.; GARCIA, I.; FERNANDEZ, J. J. Multiprocessing of anisotropic nonlinear diffusion for filtering 3D images. In: EUROMICRO INTERNATIONAL CONFERENCE ON PARALLEL, DISTRIBUTED, AND NETWORK-BASED PROCESSING, 14, 2006, Dubrovnik, Croatia. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 2006. p. 21–27.

- TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 6, 1998, Bombay, India. *Proceedings...* Washington, DC, USA: IEEE Computer Society, 1998. p. 839-.
- TOONPAINT. *ToonPAINT by Toon-FX*. 2011. Último acesso [28.02.2014]. Disponível em: <insatiablegenius.com/products/toonpaint>.
- TREAVETT, S. M. F.; CHEN, M. Statistical techniques for the automated synthesis of non-photorealistic images. In: EUROGRAPHICS UK CHAPTER, 15, 1997, Anglia, UK. *Proceedings...* Aire-la-Ville, Switzerland: Eurographics Association, 1997. p. 201-210.
- TRESSET, P.; LEYMARIE, F. F. Generative portrait sketching. In: CONFERENCE ON VIRTUAL SYSTEMS AND MULTIMEDIA, 11, 2005, Ghent, Belgium. *Proceedings...* Xi an, China: Hal Twaites, 2005. p. 739-748.
- UBISOFT. *Prince of Persia: sands of time - Ubisoft Entertainment*. 2008. Último acesso [28.02.2014]. Disponível em: <<http://www.ubi.com/US/Games/Info.aspx?pid=657>>.
- UBISOFT. *XIII - Ubisoft Entertainment*. 2003. Último acesso [28.02.2014]. Disponível em: <<http://www.ubi.com/us/games/info.aspx?pid=39>>.
- VALVE. *Team Fortress 2 - Valve Corporation*. 2007. Último acesso [28.02.2014]. Disponível em: <www.teamfortress.com>.
- VAN LAERHOVEN, T.; LIESENBORGS, J.; VAN REETH, F. Real-time watercolor painting on a distributed paper model. In *Computer Graphics International*, p. 640-643, 2004.
- VELHO, L.; TEIRA, R.; GONES, J. *Introdução aos Espaços de Escala*, 12^a Escola de Computação, 2000.
- WALLRAVEN, C.; BÜLTHOFF, H. H.; CUNNINGHAM, D. W.; FISCHER, J.; BARTZ, D. Evaluation of real-world and computer-generated stylized facial expressions. *ACM Transactions on Applied Perception*, ACM, New York, NY, USA, v. 4, n. 3, 2007.
- WANG, J. *Foreground segmentation in images and video: methods, systems and applications*, Tese (Doutorado), University of Washington, 2007.
- WANG, J.; AGRAWALA, M.; COHEN, M. F. Soft-scissors: An interactive tool for real-time high quality matting. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 26, n. 3, jul. 2007. ISSN 0730-0301.
- WANG, J.; DRUCKER, S. M.; AGRAWALA, M.; COHEN, M. F. The cartoon animation filter. *ACM SIGGRAPH 2006 Papers*. New York, NY, USA: ACM, 2006, p. 1169-1173.
- WANG, J.; XU, Y.; SHUM, H.-Y.; COHEN, M. F. Video tooning. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 31, 2004, Los Angeles, CA, USA. *Proceedings...* New York, NY, USA: ACM, 2004. p. 574-583.
- WANG, T.; COLLOMOSSE, J. Probabilistic motion diffusion of labeling priors for coherent video segmentation. *IEEE Transactions on Multimedia*, 2012.
- WARNER. *A Scanner Darkly - Warner Independent Pictures*. 2006. Último acesso [28.02.2014]. Disponível em: <www.imdb.com/title/tt0405296/>.
- WEICKERT, J. *Anisotropic Diffusion in Image Processing*. [S.l.]: ECMI Series, Teubner-Verlag, Stuttgart, Germany, 1998.
- WEICKERT, J. Coherence-enhancing diffusion filtering. *International Journal on Computer Vision*, Kluwer Academic Publishers, v. 31, n. 2-3, p. 111-127, 1999.
- WINNEMÖLLER, H. NPR in the wild. In: ROSIN, P. L.; COLLOMOSSE, J. P. (Ed.). *Image and Video-Based Artistic Stylisation*. [S.l.]: Springer, 2013. p. 353-374.
- WINNEMÖLLER, H. *Perceptually-motivated non-photorealistic graphics*. Tese (Doutorado), Northwestern University, 2006.

- WINNEMÖLLER, H. XDoG: Advanced image stylization with eXtended Difference-of-Gaussians. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 9, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 147–156.
- WINNEMÖLLER, H.; OLSEN, S. C.; GOOCH, B. Real-time video abstraction. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 25, n. 3, p. 1221-1226, 2006.
- WINTER, J. C. F. de; DODOU, D. Five-point likert items: t-test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation*, v. 15, n. 11, p. 1+, 2010. ISSN 1531-7714.
- WITKIN, A. P. Scale-space filtering. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 8, 1983, Karlsruhe, West Germany. *Proceedings...* San Francisco, CA, USA: Morgan Kaufmann Publishers, 1983. v. 2, p. 1019–1022.
- XU, J.; KAPLAN, C. S. Artistic thresholding. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 6, 2008, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2008. p. 39–47.
- XU, M.; NI, B.; TANG, J.; YAN, S. Image re-emotionalizing. In: *The Era of Interactive Media*. [S.l.]: Springer New York, 2013. p. 3–14.
- YAUCH, C, STEUDEL, H. Complementary Use of Qualitative and Quantitative Cultural Assessment Methods. *Organizational Research Methods*, 2003, v. 6, p. 465-481.
- YILMAZ, A.; JAVED, O.; e SHAH, M. Object tracking: a survey. *ACM Computing Surveys*, v. 38, Article 13, 2006.
- YU, L. *Touching 3D data: Interactive Visualization of Cosmological Simulations*. Tese (Doutorado). University of Groningen, jun. 2013.
- ZENG, K.; ZHAO, M.; XIONG, C.; ZHU, S.-C. From image parsing to painterly rendering. *ACM Transactions on Graphics*, ACM, New York, NY, USA, v. 29, n. 1, p. 2:1-2:11, dez. 2009.
- ZHANG, S.-H.; LI, X.-Y.; HU, S.-M.; MARTIN, R. R. Online video stream abstraction and stylization. *IEEE Transactions on Multimedia*, IEEE Press, Piscataway, NJ, USA, v. 13, n. 6, p. 1286–1294, 2011.
- ZHAO, H.; JIN, X.; MAO, X. Real-time directional stylization of images and videos. *Multimedia Tools and Applications*, p. 1-15, 2011.
- ZHAO, M.; ZHU, S.-C. Portrait painting using active templates. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 9, 2011, Vancouver, BC, Canada. *Proceedings...* New York, NY, USA: ACM, 2011. p. 117–124.
- ZHAO, M.; ZHU, S.-C. SISLEY - The abstract painter. In: INTERNATIONAL SYMPOSIUM ON NON-PHOTOREALISTIC ANIMATION AND RENDERING, 8, 2010, Annecy, France. *Proceedings...* New York, NY, USA: ACM, 2010. p. 99–107.

Apêndice A

Experimentos envolvendo Segmentação, Rastreamento de Objetos e Efeitos de Movimento para a Renderização Não Fotorrealística de Vídeos

Este apêndice contém fundamentos das técnicas de processamento digital de imagens (PDI) que foram alvos de experimentos em três das etapas da abordagem proposta para a geração de representações não fotorrealísticas de vídeos, fundamentada na segmentação espaçotemporal de vídeos.

Conforme ressaltado no Capítulo 1, a pesquisa ora documentada focaliza a rerepresentação do conteúdo original do vídeo de forma não fotorrealística. Na revisão bibliográfica, contida no Capítulo 2, duas abordagens destacaram-se, as quais se fundamentam no uso de: (i) algoritmos de segmentação espaçotemporal do vídeo de entrada; e (ii) algoritmos de suavização dos quadros do vídeo de entrada. A pesquisa inicialmente desenvolvida, relativa aos algoritmos que visam à segmentação espaçotemporal de vídeos, é apresentada a seguir.

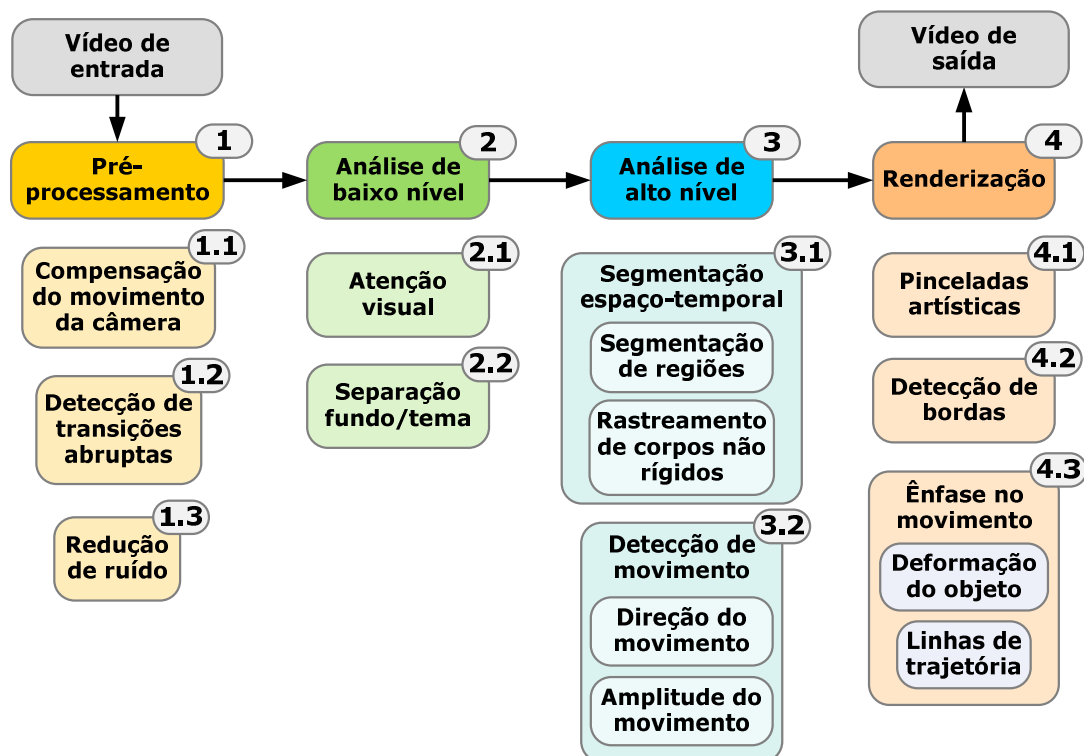
Inicialmente, uma abordagem pertencente à categoria fundamentada em algoritmos de segmentação espaçotemporal foi investigada. Embora os resultados preliminares sugerissem que tal abordagem seria promissora, limitações encontradas na resolução de problemas associados às diferentes etapas da abordagem implicaram o redirecionamento da pesquisa. Entretanto, destaca-se que a pesquisa realizada antes desse redirecionamento merece um detalhamento. Assim sendo, a abordagem fundamentada na segmentação espaço-temporal de vídeos que foi inicialmente investigada, composta por quatro etapas, é apresentada a seguir.

A.1 Abordagem Fundamentada na Segmentação Espaçotemporal de Vídeos

Uma síntese gráfica da abordagem fundamentada na segmentação espaçotemporal de vídeos é apresentada na Figura A.1, tendo sua composição se inspirado nos sistemas propostos por Wang et al. (2004), Collomosse, Rowntree e Hall (2005), Kim e Essa

(2005) e Huang (2006). Em seguida, serão brevemente descritos detalhes relativos a cada módulo.

Figura A.1 – Abordagem proposta, fundamentada na segmentação espacotemporal de quadros de vídeo.



Módulo 1 – Pré-processamento

O módulo de pré-processamento tinha como propósito melhorar o vídeo de entrada para que artefatos relatados na literatura tivessem menor influência nos módulos posteriores. O Módulo 1 foi dividido em:

1.1 Compensação de Movimento de Câmera: módulo destinado a realizar a compensação de movimento (*motion compensation*) do filme de entrada, facilitando a renderização dos módulos subsequentes e reduzindo ou eliminando efeitos de tremulações advindos do processo de aquisição do vídeo de entrada.

1.2 Detecção de Transições Abruptas: módulo destinado a delimitar o espaço de busca da análise e da renderização, possibilitando a renderização em tomadas de vídeo, ao invés da renderização do vídeo completo, em uma única etapa. O uso desta técnica não é mencionado na literatura pesquisada, constituindo, portanto, um aspecto original da primeira abordagem proposta.

1.3 Redução de Ruído: módulo destinado à melhoria do resultado final e da melhoria da abstração por meio da suavização de regiões dos quadros do vídeo de entrada.

Módulo 2 – Análise de Baixo Nível

2.1 Atenção Visual: módulo que deveria, a partir de mapas de saliência (PARK et al., 2002; ACHANTA et al., 2009) e técnicas de atenção visual, reunir informações para auxiliar módulos subsequentes, ressaltando o objeto de interesse da cena, o qual deveria ser renderizado com maior acuidade.

2.2 Separação Fundo/Tema: módulo que separaria o tema do vídeo (*foreground*) do fundo da cena (*background*), a fim de que efeitos de renderização, e.g., ênfase no movimento ou linhas de trajetória, pudessem ser aplicados apenas ao assunto do vídeo.

Módulo 3 – Análise de Alto Nível

3.1 Segmentação Espaço-temporal: este módulo seria destinado à segmentação do vídeo de forma espaço-temporal, mantendo a coerência temporal entre objetos do vídeo de entrada. A segmentação seria realizada por meio da acumulação de evidências de cor, textura e/ou forma dos objetos, após a escolha de um algoritmo de segmentação mais adequado à tarefa proposta. Adicionalmente, a informação semântica poderia ser usada para guiar a segmentação.

3.2 Detecção de Movimento: módulo que proveria ao Módulo de Renderização a orientação e a escala, assim como auxiliaria na manutenção da coerência temporal intraobjeto. O uso de algoritmos para a computação de fluxo óptico denso se mostraria mais adequado a tal tarefa, visto que, conforme evidenciado pelos autores na área de NPR para a renderização de vídeos, erros na computação do fluxo óptico esparsos geram inconsistências na renderização final.

Módulo 4 – Renderização

4.1 Pinceladas Artísticas: módulo que, a partir de uma gama de estilos disponíveis, proveria a transposição de pinceladas artísticas (*strokes*) para o vídeo, tomando como base para a coerência temporal as informações obtidas no módulo anterior.

4.2 Detecção de Bordas: módulo destinado à detecção de bordas da cena retratada, na qual o resultado da renderização e o estilo desenhado seriam enfatizados por meio da sobreposição das bordas no quadro de saída. A coerência temporal do desenho das bordas também levaria em conta as informações do movimento obtidas no módulo anterior.

4.3 Ênfase do Movimento: módulo destinado à incorporação de linhas ou

curvas de trajetória que enfatizariam o movimento. A ênfase do movimento seria obtida por meio da deformação do objeto de interesse, seguida do mapeamento da textura do tema para o objeto deformado.

Ao se analisar a abordagem inicialmente proposta, a qual objetiva a renderização não fotorrealística de vídeos digitais, optou-se pela implementação computacional inicial dos seguintes módulos: (i) 3.1 *Segmentação Espaço-temporal*; (ii) 3.2 *Detecção de Movimento*; e (iii) 3.3 *Ênfase do Movimento*. O início da implementação de tais módulos justificou-se pelos relatos das dificuldades encontradas por autores da área, em especial no tocante aos algoritmos de segmentação espaço-temporal, os quais conjugam algoritmos de segmentação de imagens com algoritmos para computação do fluxo óptico, provendo a segmentação espaço-temporal do vídeo de entrada.

Após a pesquisa e a realização de experimentos, o algoritmo de segmentação escolhido foi o algoritmo *mean-shift* topológico (PARIS; DURAND, 2007), o qual apresentou, visualmente, os resultados mais estáveis. A avaliação foi feita de forma subjetiva, já que os resultados apresentados pelos demais algoritmos testados em uma sequência de quadros apresentaram o efeito de cintilação (*flickering*) muito acentuado. O detalhamento dos experimentos relativos a este algoritmo é apresentado na Seção A.2.

Durante a realização de experimentos, observou-se que quadros de entrada consecutivos que apresentavam mudanças insignificantes, tanto visual quanto computacionalmente³³, geravam resultados muito díspares na segmentação, para os mesmos valores de parâmetros utilizados.

O algoritmo para a computação do fluxo óptico escolhido foi aquele proposto por Farnebäck (2003), que se destacou, após a realização de experimentos, pela estabilidade na detecção da magnitude e direção do movimento, pela gama de parâmetros a serem calibrados, que proporcionariam uma detecção mais lenta e mais precisa ou uma detecção mais rápida e mais imprecisa, assim como por ter apresentado o ajuste da densidade na computação do fluxo óptico. O detalhamento dos experimentos relativos a este algoritmo é apresentado na Seção A.3.

Adicionalmente, o detalhamento dos experimentos relativos aos algoritmos utilizados no processo de ênfase no movimento é apresentado na Seção A.4.

Embora os resultados obtidos a partir dos algoritmos de segmentação e de

³³ Ao subtrair quadros de entrada consecutivos, verificava-se a apresentação de mudanças inferiores a 5% de um quadro para outro.

computação do fluxo óptico supramencionados terem sido satisfatórios para uma gama variada de situações e imagens e apesar de serem os referidos algoritmos populares na literatura da área, além de possuírem implementações computacionais disponíveis, a tarefa de integrar as regiões segmentadas e rastrear-las a partir das informações de movimento afigurou-se, ao mesmo tempo, desafiadora e frustrante. Os desafios encontrados, refletidos nas frustrações associadas aos resultados obtidos, relacionaram-se à incapacidade de integração, de forma satisfatória, dos resultados dos dois algoritmos, a fim de gerar um vídeo abstraído temporalmente coerente, i.e., que apresentasse o mínimo de cintilação ao longo dos quadros.

Outro problema surge do fato de, tipicamente, haver supersegmentação nas abordagens fundamentadas em agrupamento ou exagero nas regiões detectadas, fazendo com que regiões similares apareçam separadas pela mudança na textura ou brilho da região a ser segmentada. Arbelaez et al. (2009) introduziram uma nova perspectiva de avaliação para os algoritmos de segmentação, ao empregarem bases canônicas de imagens de referência no processo de avaliação de segmentadores, assim como a comparação objetiva usando gráfico precisão versus revocação (*precision versus recall*), *ground-truth* e medida-F. No entanto, parece não haver consenso na literatura da área a respeito do melhor algoritmo destinado à segmentação de imagens.

No tocante às técnicas de segmentação, Arbelaez et al. (2009) afirmaram que várias tarefas no campo da visão computacional podem se beneficiar de uma redução na complexidade no processamento computacional, a partir da utilização de algoritmos de segmentação: ao invés de realizar o processamento de milhões de pixels, a tarefa restringe-se ao processamento de algumas centenas de *super-pixels*. Adicionalmente, os autores afirmaram que, dentre as mais variadas técnicas para a segmentação, propostas ao longo dos anos, três técnicas se destacam, a saber: (i) a técnica de agrupamento de regiões a partir de grafos, proposta por Felzenszwalb e Huttenlocher (2004); (ii) a técnica *mean-shift*, proposta por Comaniciu e Meer (2002); e (iii) a técnica dos cortes normalizados de grafos, proposta por Shi e Malik (2000).

Wang e Collomosse (2012) ressaltaram a dificuldade encontrada na geração de abstrações de vídeos a partir de algoritmos de segmentação, de forma temporalmente coerente. Os autores ressaltaram que, apesar de avanços em técnicas de segmentação, tal processo ainda permanece desafiador para a segmentação de vídeos de moderada complexidade, devido a fatores tais como mudanças na iluminação e a

oclusão de objetos. Os autores propuseram uma técnica probabilística que melhora a estimativa de movimento gerada a partir de um algoritmo que realizou a computação do fluxo óptico.

Limitações foram encontradas no emprego das técnicas de segmentação e rastreamento de movimento, os quais não produziram resultados satisfatórios, nem do ponto de vista subjetivo, nem do ponto de vista da coerência temporal. Considerou-se, adicionalmente, a revisão bibliográfica realizada e o tempo delimitado para a pesquisa, no contexto de um programa de doutoramento e julgou-se pertinente redirecionar a abordagem inicialmente considerada para a conversão de vídeos em vídeos renderizados não fotorrealisticamente. A pesquisa realizada após este redirecionamento está apresentada nos 6 capítulos desta tese.

A seguir, são apresentados os experimentos envolvendo técnicas de segmentação (Seção A.1). A Seção A.2 contém a descrição dos experimentos envolvendo algoritmos para computação do fluxo óptico. Finalmente, a Seção A.3 contém a descrição dos experimentos relativos à inclusão de efeitos de movimento em vídeo.

A.2 Experimentos relativos às Técnicas de Segmentação

Conforme ressaltado no Capítulo 2, Wang et al. (2004) verificaram o aparecimento de artefatos indesejáveis quando técnicas de NPR adequadas a imagens estáticas são aplicadas a quadros de vídeo. Após breve experimentação envolvendo a técnica proposta por Arruda et al. (2008), alguns artefatos ficaram evidentes quando a técnica foi aplicada a quadros adjacentes.

Visando atender aos objetivos da pesquisa proposta e visando a redução das dificuldades encontradas ao longo da realização dos experimentos conduzidos, foram testadas várias técnicas de segmentação de imagens. Tais técnicas foram aplicadas a quadros consecutivos de um vídeo, apresentados na Figura A.2, sendo os resultados obtidos apresentados a seguir. As técnicas exploradas foram: *mean-shift*, *mean-shift* topológico, k-médias, cortes normalizados de grafos e transformada *watershed* orientada.

Figura A.2 – Quadros de entrada.



No âmbito desta pesquisa, associa-se o *agrupamento de pixels em regiões de uma imagem* à busca da similaridade dos valores de seus atributos, com o propósito de verificar a pertinência de um dado pixel a um determinado grupo possível. São utilizadas técnicas de busca não supervisionada, visto que a similaridade entre os atributos é intrínseca aos dados, não necessitando, portanto, de treinamento prévio com grupos pré-determinados. O processo é normalmente iterativo, necessitando a intervenção do usuário no tocante à definição prévia da quantidade de classes, assim como da configuração satisfatória dos parâmetros a serem considerados.

O algoritmo *mean-shift* (COMANICIU; MEER, 2002) fundamenta-se em uma técnica estatística, segundo a qual se busca o pico entre distribuições de probabilidade em gradientes de densidade. Esta técnica é denominada *média por deslocamento* e pode ser empregada para a segmentação ou a suavização de imagens. Durante a aplicação do *mean-shift*, informações sobre regiões densas da imagem são armazenadas e separadas em agrupamentos, de acordo com dois parâmetros, um espacial e outro radial. Cada pixel da imagem é associado a um agrupamento pertencente à moda, após sucessivas iterações do algoritmo. A Figura A.3 contém um exemplo da aplicação da técnica *mean-shift*, tendo como base a implementação "*Edge Detection and Image SegmentatiON - EDISON*"³⁴, com as seguintes configurações: $\sigma_s=10$, $\sigma_r=6.5$, $\text{minRegion}=10$, medSpeedup . Comentários acerca dos resultados podem ser encontrados ao final desta seção.

Figura A.3 – Mean-shift aplicado a quadros consecutivos de um vídeo.



Uma variante do algoritmo *mean-shift* que foi proposta por Paris e Durand (2007), denominada *mean-shift* topológico. Os autores modificaram o algoritmo *mean-shift* a partir da aplicação da teoria das variedades topológicas, de modo a reinterpretar uma imagem na forma de um relevo e pré-computar uma grade para a representação das coordenadas espaciais dos pixels. Para cada ponto da grade, é computada a função de densidade e armazenadas as modas de cada ponto. Os pixels são então agrupados de acordo com a similaridade, resultando nas regiões do algoritmo *mean-shift*. A alteração proposta acelera o processamento em imagens com alta resolução. No entanto, o método apresenta algumas falhas, tais como: (i) a

³⁴ Software de segmentação disponível em: <http://coewww.rutgers.edu/riul/research/code.html>

natureza puramente digital do método, sem uma formalização e simplificação real do método *mean-shift*, pois é feita a pré-computação de alguns parâmetros da imagem; e (ii) o descarte de regiões da imagem fundamentadas em critérios puramente heurísticos. A Figura A.4 contém um exemplo da aplicação da técnica *mean-shift* topológico. Comentários acerca dos resultados podem ser encontrados ao final desta seção.

Figura A.4 – Mean-shift topológico aplicado a quadros consecutivos de um vídeo.



O algoritmo K-médias (*k-means*) destina-se à descoberta de agrupamentos e deriva seu nome do fato de ser iniciado com um conjunto de k “sementes”, escolhidas como suposição inicial para os centróides ou médias. O algoritmo é implementado em quatro passos: (i) realiza-se a partição dos objetos em k grupos não vazios; (ii) definem-se as sementes como centróides dos grupos da partição atual; (iii) associa-se cada objeto ao grupo cuja semente lhe for mais próxima; e (iv) retorna-se ao passo ii, atualizando os centróides até que não mais se modifiquem. A Figura A.5 contém um exemplo da aplicação da técnica, para $k = 20$. Comentários acerca dos resultados podem ser encontrados ao final desta seção.

Figura A.5 – K-médias aplicado a quadros consecutivos de um vídeo.

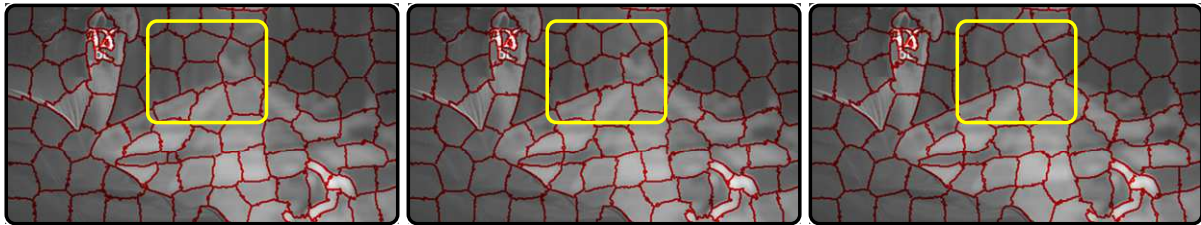


Técnicas fundamentadas em cortes de grafos também se destacaram recentemente (BOYKOV; JOLLY, 2001; BOYKOV; KOLMOGOROV, 2004). São técnicas semiautomáticas, nas quais o usuário deve definir previamente o número de partições que o resultado deverá conter.

O algoritmo denominado *cortes normalizados (normalized cuts)*, proposto por Shi e Malik (2000), trata a segmentação como um problema de partição de um grafo, sendo a imagem descrita como um grafo completamente conectado e os pixels (formando vértices) conectados aos vizinhos por arestas. As arestas são ponderadas de acordo com a similaridade da vizinhança dos pixels. Tais arestas são removidas de

acordo com um critério de similaridade entre pixels vizinhos, devendo as arestas removidas totalizarem o menor peso possível. A Figura A.6 contém um exemplo da aplicação da técnica, com o corte pré-definido para resultar em 100 regiões. Comentários acerca dos resultados podem ser encontrados ao final desta seção.

Figura A.6 – Cortes normalizados de grafos aplicado a quadros consecutivos de um vídeo.



O algoritmo *globalPb-owt* (ARBELAEZ et al., 2009) integra informações de brilho, cor e textura locais, obtidas após a aplicação de operadores de gradientes locais em excertos da imagem, à informações espectrais globais, obtidas após a partição da imagem a partir de grafos. A junção das informações locais e globais fornece os contornos da imagem. Essas informações de contorno são, então, empregadas como entrada para um algoritmo de agrupamento, uma especialização do algoritmo *watershed* denominada transformada *watershed orientada*, a qual agrupa as regiões de contornos anteriormente obtidas e organiza tais grupos de forma hierárquica e probabilística, resultando em um mapa de contornos não binário, ou seja, o usuário deve definir qual o nível de interesse para a formação da imagem final. Dada a natureza dual da abordagem, ela é denominada *globalPb-owt* (*globalized probability of boundary + oriented watershed transform*). A Figura A.7 contém um exemplo da aplicação da técnica a quadros de um vídeo. Comentários acerca dos resultados podem ser encontrados ao final desta seção.

Figura A.7 – GlobalPb+owt aplicado a quadros consecutivos de um vídeo.



Conforme ressaltado anteriormente, o algoritmo *mean-shift* topológico foi o que apresentou, visualmente, os resultados mais estáveis. Uma avaliação visual foi realizada nas figuras presentes nesta seção, nas quais foram destacadas regiões que obtiveram segmentações díspares para quadros do vídeo de entrada consecutivos. Os quadros testados não apresentam variação significativa. Adicionalmente, os resultados apresentados ao aplicar os algoritmos de segmentação diferentes do *mean-shift*

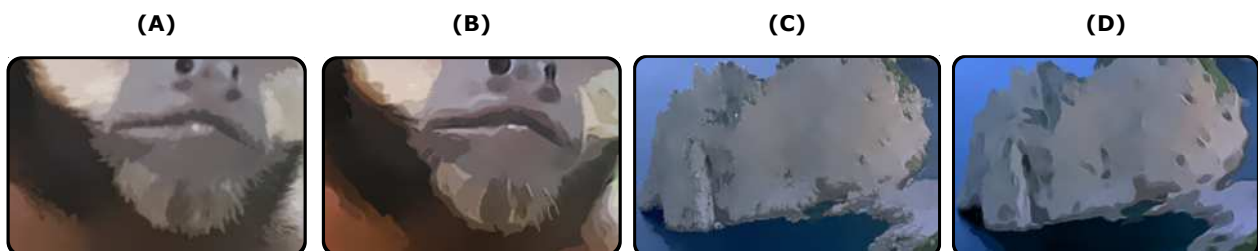
topológico à todo o conteúdo do vídeo de entrada apresentaram efeito de cintilação muito acentuado.

Ressalta-se que o algoritmo *globalPb-owt* (ARBELAEZ et al., 2009) apresenta bons resultados, mas inferiores ao *mean-shift* topológico, tanto na apresentação dos resultados quanto no desempenho computacional, requerendo quantidades expressivas de memória (4GB de memória principal para uma imagem de 620x350 pixels) e tempo computacional elevado ($t > 1$ hora para a renderização de 60 quadros).

Adicionalmente, são apresentados nas Figura A.8A e Figura A.8C excertos de imagens obtidas a partir da aplicação do processo de segmentação *mean-shift* topológico, aplicado à imagem de entrada sem nenhum pré-processamento. Nas Figura A.8B e Figura A.8D, são apresentados excertos de imagens obtidas após a aplicação da abordagem proposta, destinada à renderização não fotorrealística (apresentada no Capítulo 3), em conjunção ao processamento por segmentação com a utilização do *mean-shift* topológico.

Percebe-se que há maior abstração quando o processo de segmentação é realizado após a aplicação do processo de renderização não fotorrealística, quando comparado ao processo de segmentação realizado na imagem obtida sem tal processamento. Entretanto, após a realização de experimentos em 2 vídeos (*Crew* e *Cowgirl*, apresentados no Capítulo 4), a segmentação não apresentou resultados satisfatórios, do ponto de vista subjetivo, no processo de segmentação aplicado a vídeos. Um exemplo do vídeo segmentado pode ser visualizado no *hyperlink* que contém o material adicional desta pesquisa (apresentado ao final do Capítulo 4).

Figura A.8 – Mean-shift topológico após a renderização utilizando a abordagem proposta.



A.3 Experimentos envolvendo a Computação do Fluxo Óptico

A pesquisa relativa à detecção e rastreamento automático de movimento em vídeos digitais serviria como sustentáculo para a conversão de vídeos em vídeos renderizados não fotorrealisticamente, no estilo de cartuns. Comumente a detecção e rastreamento de movimento em vídeo são realizados por algoritmos que realizam a computação de fluxo óptico. As informações de movimento não foram utilizadas de

forma conjunta nas técnicas estudadas anteriormente, no âmbito da renderização não fotorrealística de vídeo. Observou-se, na literatura pesquisada, que as técnicas para a extração de informações do movimento destinadas à renderização de efeitos do tipo *cartum* podem ser usadas para: (i) desenhar listras de movimento e linhas-fantasma; (ii) computar descritores de forma, cor e movimento que possibilitem a introdução de deformações de movimento no vídeo de saída; e (iii) enfatizar o movimento, com base na quantidade de movimento do assunto do vídeo de entrada.

Adicionalmente, o uso de um detector e estimador de movimento aprimorado, possibilitaria a segmentação espaçotemporal de vídeos, atuando de forma conjunta a algoritmos de segmentação estudados e relatados na Seção A.2. A computação do fluxo óptico apresentou-se enquanto técnica mais comumente empregada para estimar o movimento em vídeos, com vistas à renderização estilizada.

Assumpção Júnior (2009), em um contexto diferente daquele no qual se insere o presente trabalho, avaliou as dificuldades da implementação em *hardware*, de um sistema de desvio de obstáculos para robôs móveis, utilizando uma única câmera, em conjunto com algoritmos para computação do fluxo óptico. Em sua proposição, o autor utilizou os referidos algoritmos para a geração do mapa de navegação do robô, em contraste com mapas pré-definidos (e.g., usando GPS e mapas estáticos fornecidos por terceiros).

Assumpção Júnior (2009) definiu que a computação do fluxo óptico ocorre quando movimentos de objetos no espaço tridimensional em relação à câmera provocam movimentos correspondentes em duas dimensões nas imagens capturadas. O autor ressaltou que o módulo desenvolvido pode ser aplicado a tarefas de segmentação de imagens que empreguem as descontinuidades do fluxo obtido, indicando o movimento de um objeto com relação ao fundo, a fim de guiar um robô. Adicionalmente, a visão estéreo pode ser combinada à computação do fluxo óptico a fim de que sejam aproveitadas as vantagens de cada método.

De acordo com Assumpção Júnior (2009), Horn e Schunck (1981) estabeleceram os princípios que fundamentam os algoritmos modernos para a computação do fluxo óptico, nos quais se define que uma textura que não varia no tempo e que se move paralelamente ao plano da imagem representa o fluxo de movimento. Supondo restrições aos vetores de fluxo, ou seja, definindo-se que os vetores devam variar suavemente ao longo da imagem, uma solução algorítmica é definida, de modo que, de forma iterativa, vetores sejam calculados nos pontos nos

quais o gradiente seja passível de cálculo. Sucessivas iterações do algoritmo calculam os vetores dos pixels vizinhos. O processamento é finalizado quando todos os valores dos vetores desconhecidos sejam calculados.

As iterações mencionadas podem ser restringidas de formas diferentes. Cada restrição apresentada suscita o aparecimento de um algoritmo para computação de fluxo óptico diferente. Visando a organização e comparação dos algoritmos para computação de fluxo óptico, Barron et al. (1994) agruparam os algoritmos em quatro grupos distintos, dos quais se destacam as técnicas diferenciais e as técnicas de convolução. Os autores utilizaram imagens em escala de cinza, o que reduz a quantidade de informação a ser tratada pelo algoritmo utilizado.

O algoritmo original de Horn e Schunck (1981), classificado como uma técnica diferencial, avalia as derivadas das imagens de entrada para impor restrições. São utilizados filtros espaciais e temporais nas imagens de entrada para facilitar o cálculo das derivadas de pixels vizinhos. No decorrer das iterações dos algoritmos do tipo diferencial, há parâmetros que são ajustados para selecionar a velocidade de convergência. Uma velocidade alta reduz a quantidade de cálculos necessários, mas gera resultados com menor precisão.

Nas técnicas de convolução para a computação do fluxo óptico, uma parte da imagem atual é comparada com partes da imagem anterior. Diferentes vetores de deslocamentos são testados para determinar qual a solução que minimiza as diferenças entre as duas imagens. Este vetor será a resposta apenas para o ponto central dos trechos comparados. Outros vetores são encontrados após a repetição do processo descrito em outras regiões da imagem.

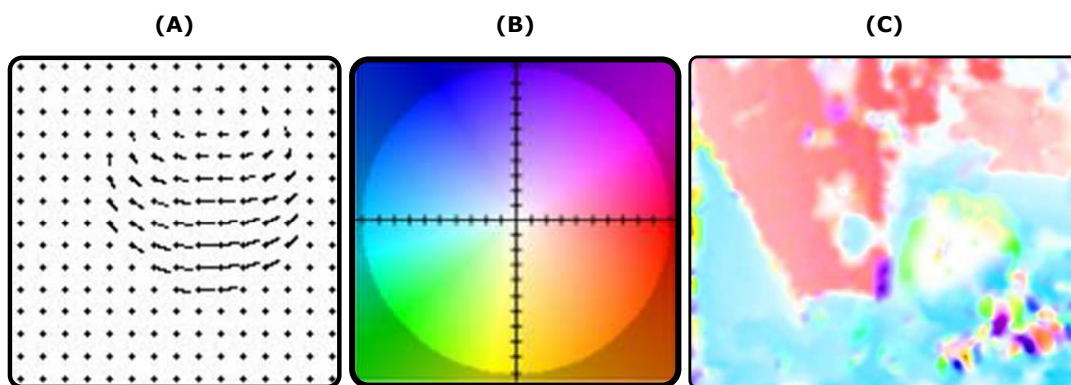
Barron et al. (1994) propuseram a primeira comparação numérica, detalhada e objetiva de algoritmos para a computação de fluxo óptico, propondo o primeiro banco de imagens de referência para testes de tais algoritmos. Os autores adotaram uma avaliação objetiva, ao averiguar que uma avaliação visual, para observadores humanos, se afigurava demasiadamente subjetiva. Os autores subtraíram o vetor de cada ponto dos campos de fluxo óptico gerados, obtendo uma única medida de disparidade entre uma imagem de referência e o campo de fluxo obtido por um algoritmo de computação de fluxo óptico considerado. Baker et al. (2007) ressaltaram a importância da proposição de uma medida de disparidade que contemple a computação da magnitude e ângulo de tais vetores.

A Figura A.9 contém exemplos de representação do fluxo óptico. Cada vetor

ocupa diversos pontos na imagem, para que seja ressaltada a dimensão e direção, o que faz com que apenas um subconjunto possa ser mostrado em uma imagem. Tal representação é denominada esparsa. Alguns algoritmos não calculam uma resposta para cada pixel. Entretanto, para a representação em blocos, tal como ocorre nos algoritmos de convolução, tal representação é satisfatória. A alternativa é a representação de fluxo óptico de forma densa, no qual cada vetor é indicado por uma cor, representando, assim, vetores de movimento relativos a pixels individuais.

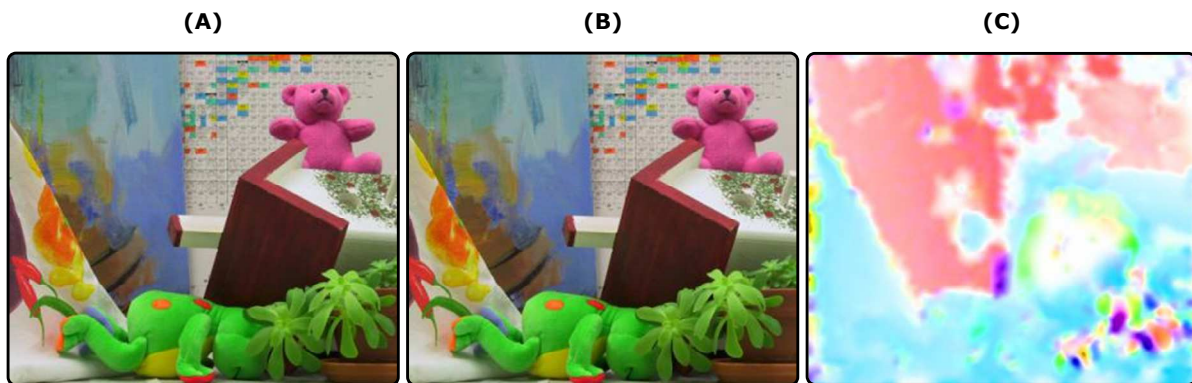
A obtenção de dados de movimento com pouco erro de medição e com qualidade satisfatória é um fator importante a ser considerado na avaliação de algoritmos para a computação de fluxo óptico. A alternativa é o uso de imagens sintéticas, já que a existência do modelo de referência provê os vetores reais de movimento, enquanto uma projeção no plano da câmera gera o fluxo óptico verdadeiro. McCane et al. (2001) selecionaram cenas naturais com objetos poligonais e côncavos, nas quais apenas a câmera se movimentava. Um algoritmo proposto pelos autores auxilia um operador no cálculo do fluxo óptico.

Figura A.9 - Representação de vetores de deslocamento em algoritmos de fluxo óptico: (A) Esparso; (B) Legenda para denso; e (C) Denso.



Baker et al. (2007) propuseram uma forma de avaliação que considera um conjunto de imagens e um repositório de referência de algoritmos para a computação de fluxo óptico. Os autores propuseram a avaliação de cenas complexas, com objetos flexíveis (e.g., tecidos). A solução encontrada pelos autores para a geração das imagens de referência (*ground-truth*) e dos vetores reais de fluxo para a avaliação foi encobrir os objetos a serem avaliados com marcadores invisíveis à câmera, em conjunto com outra câmera de alta resolução e sensível apenas aos marcadores. Na Figura A.10, é mostrada a resposta esperada para o fluxo óptico entre os quadros 10 e 11, disponibilizado por Baker et al. (2007). A legenda de cores para o movimento é a mesma adotada na Figura A.9B.

Figura A.10 – Exemplo de imagem usada na avaliação de fluxo óptico (BAKER et al., 2007): (A) Imagem 1; (B) Imagem 2; e (C) Fluxo detectado.



Diante do exposto, no tocante aos algoritmos destinados à computação do fluxo óptico e visando a atender aos objetivos da pesquisa de técnicas para a renderização não fotorrealística de vídeos, fundamentadas na segmentação espaçotemporal, foram testados dois algoritmos para computação de fluxo óptico, a saber: Lucas-Kanade/Horn Schunck (BRUHN; WEICKERT; SCHNÖRR, 2005) e o algoritmo de Farnebäck (FARNEBÄCK, 2003).

Baker et al. (2007) empregaram quatro medidas de qualidade para classificar os algoritmos para computação de fluxo óptico, duas para a exatidão de fluxo (erro angular e erro de ponto final) e duas para a qualidade da interpolação intraquadros. Os autores não definiram qual a melhor métrica a ser adotada a fim de comparar algoritmos, pois a definição do melhor algoritmo ou da melhor medida para a avaliação de tais algoritmos é função do domínio de aplicação destes algoritmos. Os autores deixaram a opção e avaliação a critério do pesquisador.

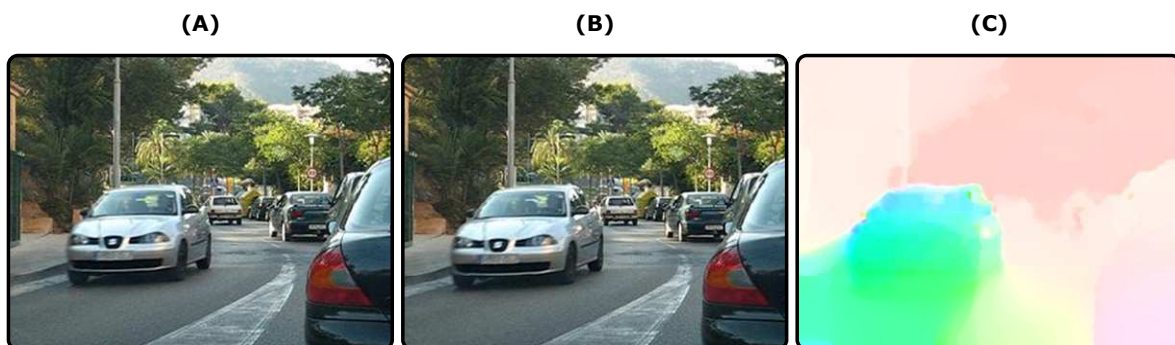
No mesmo contexto, Gomes et al. (2009) escolheram o método *FusionFlow*, proposto por Lempitsky et al. (2008) que, na página Web que retrata a pesquisa em algoritmos para a computação do fluxo óptico da Middlebury University (BAKER et al., 2007), é atualmente classificado como um dos quatro melhores métodos para a computação do fluxo óptico, no que concerne as quatro medidas supracitadas. O método *FusionFlow* (LEMPITSKY et al., 2008) formula a computação do fluxo óptico enquanto um problema de corte de grafo, combinado à um conjunto de soluções usando cortes mínimos de grafos. A função de energia utilizada é dividida em dois termos: (i) um de dados, que mede quão bem o campo de fluxo descreve a correspondência entre pixels nas duas imagens; e (ii) outro espacial, que penaliza as mudanças no fluxo horizontal e vertical entre pixels adjacentes.

A distância entre dois pixels, no termo de dados do *FusionFlow*, é a distância Euclidiana no espaço RGB, a qual é calculada após a realização de uma filtragem

passa-altas dos dados de entrada, a fim de melhorar a robustez dos dados de entrada. A fusão se dá entre (i) o *método Lucas-Kanade* (LUCAS; KANADE, 1981), que geralmente produz estimativas precisas para as regiões de textura, mas não para regiões sem textura e (ii) o *método Horn-Schunck* (HORN; SCHUNCK, 1981), que produz estimativas de fluxo corretas para regiões de movimento suave, mas suaviza em demasia o movimento em regiões com movimento descontínuo.

Resultados similares foram obtidos por Bruhn, Weickert e Schnörr (2005), que propuseram a técnica *Lucas-Kanade/Horn Schunck*. O processo é iterativo e começa pela escolha aleatória de um dos dois métodos (Horn-Schunck ou Lucas-Kanade), sendo tal aleatoriedade repetida duas outras vezes e os métodos combinados de forma aleatória, a partir de sua fusão com o vetor de fluxo inicial. Devido ao acesso facilitado por uma implementação já disponível de tal técnica, alguns testes foram realizados. A Figura A.11 contém um exemplo da aplicação da técnica. Comentários acerca dos resultados podem ser encontrados ao final deste apêndice.

Figura A.11 – Lucas-Kanade/Horn-Schunck, por Bruhn, Weickert e Schnörr (2005): (A) Imagem 1; (B) Imagem 2; e (C) Fluxo detectado.



Conforme mencionado, os métodos de Horn e Schunck (1981) e Lucas-Kanade (1981) possuem limitações. De acordo com Bradski e Kaehler (2008), tais métodos mostraram-se robustos ao longo do tempo para uma gama variada de tarefas de análise de movimento, mesmo tendo sido propostos na década de 80. Os métodos de fluxo óptico denso foram então relegados apenas ao interesse puramente acadêmico. No entanto, ao longo da pesquisa, um algoritmo de fluxo óptico denso destacou-se: o *algoritmo de Farnebäck* (2003). A avaliação de tal algoritmo não se encontra disponível na página Web da Middlebury University (BAKER et al., 2007).

Os resultados obtidos por Bruhn, Weickert e Schnörr (2005) mostraram-se inferiores àqueles obtidos a partir do algoritmo de Farnebäck (2003). O algoritmo de Farnebäck possui limitações, em especial a acurácia na computação do fluxo óptico ao medir movimentos rápidos. Ressalta-se ainda a vasta gama de parâmetros a serem calibrados no algoritmo de Farnebäck, visando uma detecção mais lenta e mais

precisa, uma detecção mais rápida e mais imprecisa, ou ainda um ajuste na densidade do cálculo do fluxo óptico. O excesso de parâmetros a serem calibrados pode constituir um desafio a ser considerado ao utilizar o algoritmo de Farnebäck.

Diante dos resultados expostos e após a visualização de outros vídeos nos quais algoritmos para a computação do fluxo óptico foi testado, o *algoritmo de Farnebäck* foi o que apresentou, visualmente, os resultados mais estáveis. Outras técnicas para a computação de fluxo óptico foram testadas (PROESMANS et al., 1994; HORN; SCHUNCK, 1981), mas não foram aqui relatadas em virtude da imprecisão do fluxo óptico, mesmo em um vídeo com câmera fixa e sem nenhum movimento ocorrendo na cena, o que inviabilizou sua utilização.

Sun, Roth e Black (2010) relataram que as pesquisas acerca de algoritmos de fluxo óptico, após cerca de 30 anos, obtêm hoje um bom nível de confiabilidade e precisão. Os autores propuseram a adição de uma etapa de pré-processamento à computação do fluxo óptico, i.e., a aplicação de um filtro da mediana nos passos intermediários durante a computação do fluxo óptico, durante as várias iterações do algoritmo. Os autores relataram tendências de pesquisa que visam o aumento da robustez dos algoritmos para computação do fluxo óptico: (i) a inclusão de informações a respeito da estrutura da cena; (ii) o uso de informações de bordas contornos dos objetos presentes na cena; ou (iii) a inclusão de informações a respeito das superfícies e fronteiras entre objetos presentes na cena.

As imagens apresentadas na pesquisa realizada por Baker et al. (2007) foram inspiradas em situações do mundo real. As imagens apresentam desafios à algoritmos de computação de fluxo óptico, tais quais: mudanças complexas na iluminação da cena, ruído intrínseco do sensor no momento da captura, efeitos atmosféricos, efeitos de transparência nos objetos da cena. Esses desafios foram transferidos para as métricas computadas pelos autores, sugerindo a adição, em trabalhos futuros na área de avaliação de algoritmos para computação de fluxo óptico, de cenários pautados em situações reais.

A.4 Experimentos de Efeitos de Movimento em Vídeo

Foram testadas duas técnicas que adicionam efeitos de movimento à vídeos digitais: (i) *Continuously Adaptive Mean Shift (camshift)*, proposta por Bradski (1998) e *motion history image*, proposta por Bradski e Davis (2000).

A implementação disponível do algoritmo *camshift* foi utilizada, no contexto desta pesquisa, como método para a separação entre fundo/tema de uma cena,

realizada de forma semiautomática. Esse algoritmo combina o algoritmo *mean-shift* clássico proposto por Comaniciu e Meer (2002) com um passo adicional de redimensionamento de regiões, baseada na cor indicada pelo canal H (*hue* /tonalidade) do modelo cromático HSV. A seleção do objeto de interesse é feita de forma manual pelo usuário. Após a seleção do objeto de interesse, o centróide que localiza o objeto de interesse é deslocado de acordo com a probabilidade de encontrar o objeto selecionado. A motivação para o uso do algoritmo *camshift* foi a seleção do objeto de interesse a ser renderizado de forma não fotorrealística.

A técnica denominada *mhi* pode ser utilizada de forma complementar ao algoritmo *camshift* visando à adição de efeitos especiais em vídeo. Em tal técnica, a sequência de silhuetas é condensada em imagens consecutivas, enquanto a informação do movimento dominante é armazenada. Diante dos resultados obtidos, houve uma tentativa de agregar os dois algoritmos testados. No entanto, tal ação concomitante dos dois algoritmos mostrou-se inviável para a adição de efeitos especiais em vídeo, mesmo tendo sido feita uma avaliação de forma subjetiva dos resultados apresentados. A seguir, são relatadas considerações a respeito dos algoritmos para rastreamento de movimento, que podem nortear pesquisas relacionadas à adição de efeitos especiais em vídeo, as quais se afiguram enquanto investigações futuras de pesquisa.

Yilmaz, Javed e Shah (2006) ressaltaram progressos nas pesquisas em técnicas de rastreamento de objetos. Em especial, progressos relativos ao desempenho em tempo real das técnicas de rastreamento em cenários simples. Entretanto, o rastreamento de objetos em cenários simples restringe a aplicabilidade das técnicas, enquanto assumem alguns fatores constantes durante o rastreamento do objeto, tornando o problema tratável, mas falhando em cenários de uso reais (e.g., vigilância automatizada, aplicações de interação homem-máquina, monitoramento de tráfego urbano ou navegação veicular autônoma). Os fatores incluem a oclusão mínima do objeto rastreado, a iluminação constante da cena, alto contraste em relação ao fundo da cena, entre outros. Os autores ressaltaram ainda que rastreadores futuros devam incluir informações a respeito do ambiente nos quais estes se inserem, e.g., a posição e direção de linhas no asfalto podem ajudar na orientação dos rastreadores que lidam com tráfego urbano. Yilmaz, Javed e Shah (2006) afirmam ainda que capacidades de aprendizagem possam ser agregadas aos rastreadores a serem desenvolvidos em pesquisas futuras, visando à identificação automática dos objetos da cena sem a necessidade de conhecimento prévio do ambiente.

Apêndice B

Resultados Obtidos a Partir da Utilização da Abordagem Proposta

Neste apêndice são apresentados resultados adicionais obtidos a partir da utilização da abordagem proposta para a renderização não fotorrealística de imagens e vídeos. A exposição de tais resultados foi motivada por sua utilização em várias publicações da literatura da área de NPR. Adicionalmente, a presença dessas imagens evidencia aspectos relativos à abordagem proposta que podem não ter sido evidenciados nos Capítulos 3 e 4 deste documento, e.g., a renderização não fotorrealística de texturas da natureza, a renderização de nuvens.

Imagem original

Imagem renderizada

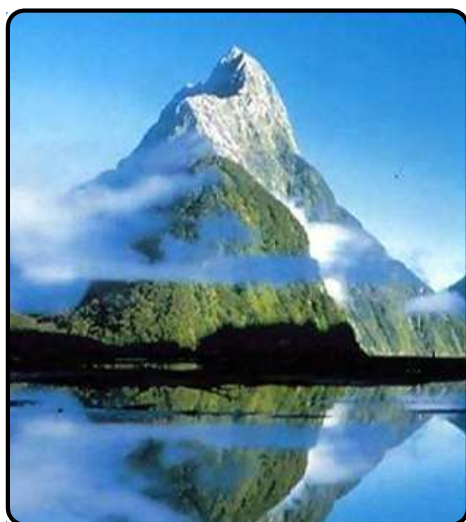


Imagem original



Imagem renderizada



Imagem original



Imagem renderizada



Apêndice C

Resultados Obtidos a Partir de Técnicas Destinadas à Quantização Cromática

Neste apêndice, são apresentados dois resultados obtidos a partir da utilização da abordagem proposta para a abstração de imagens em conjunção com um processo de quantização cromática. O processo de quantização cromática foi aplicado após a renderização da imagem mediante a utilização de todos os módulos da abordagem proposta.

A motivação para o relato destes experimentos adveio da existência de um processo de quantização cromática nas abordagens propostas por Winnemöller, Olsen e Gooch (2006), Kyprianidis e Döllner (2008) e Kyprianidis (2013). A inserção desse processo à abordagem proposta afigura-se um tópico de investigação futura, conquanto este produziu resultados insatisfatórios, do ponto de vista de coerência temporal, ao ser aplicado à vídeos digitais.

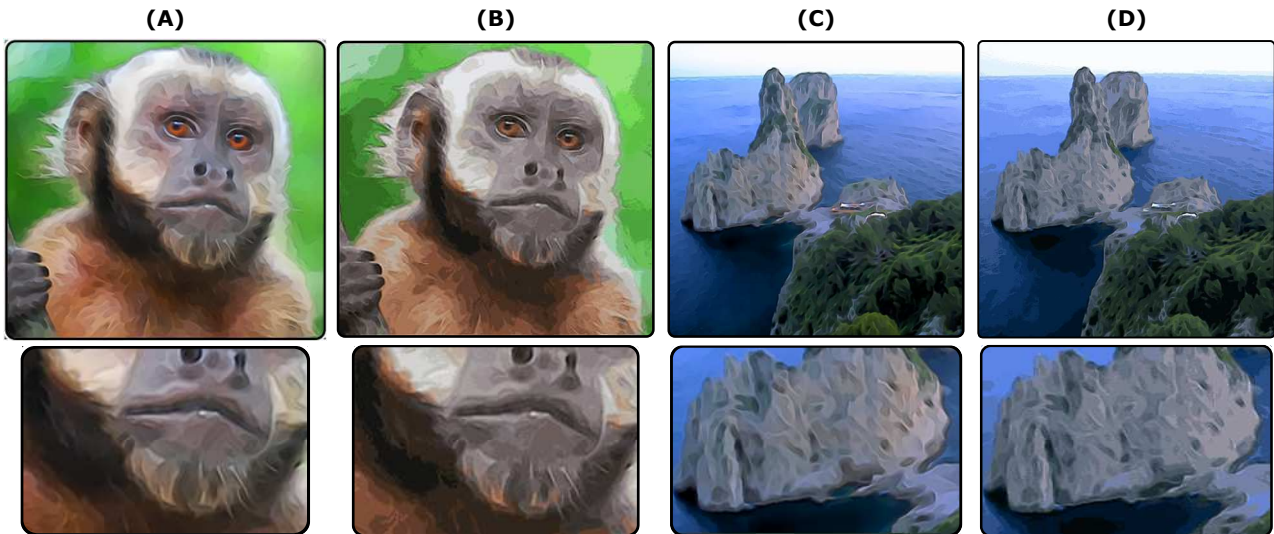
As renderizações apresentadas neste apêndice foram obtidas por meio da utilização do *plug-in* Ximagic Quantizer³⁵, especificamente desenvolvido para o programa Adobe Photoshop. A utilização de tal *plug-in* possibilitou o processamento de imagens para a quantização cromática e *dithering* (inserção de ruído de modo a mitigar erros de quantização visíveis ao reduzir a profundidade de bits das amostras).

No Ximagic Quantizer são disponibilizadas dez técnicas para quantização cromática, as quais podem ser utilizadas com o intuito de reduzir o número de cores presente na imagem processada, e.g.: quantização por popularidade (GERVAUTZ; PURGATHOFER, 1990); *median cut* (HECKBERT, 1982). Na Figura C.1A e Figura C.1C são apresentados resultados obtidos a partir da utilização da abordagem proposta para renderização não fotorrealística. As imagens apresentadas na Figura C.1B e Figura C.1D foram geradas a partir da utilização do *plug-in* Ximagic Quantizer, consideradas as opções: Denis Lee v3.0, *calculated palette*, 64 cores, 6 bits por cor. Observa-se maior abstração do fundo da cena do macaco prego, e maior segmentação

³⁵ Disponível em: http://www.ximagic.com/q_index.html

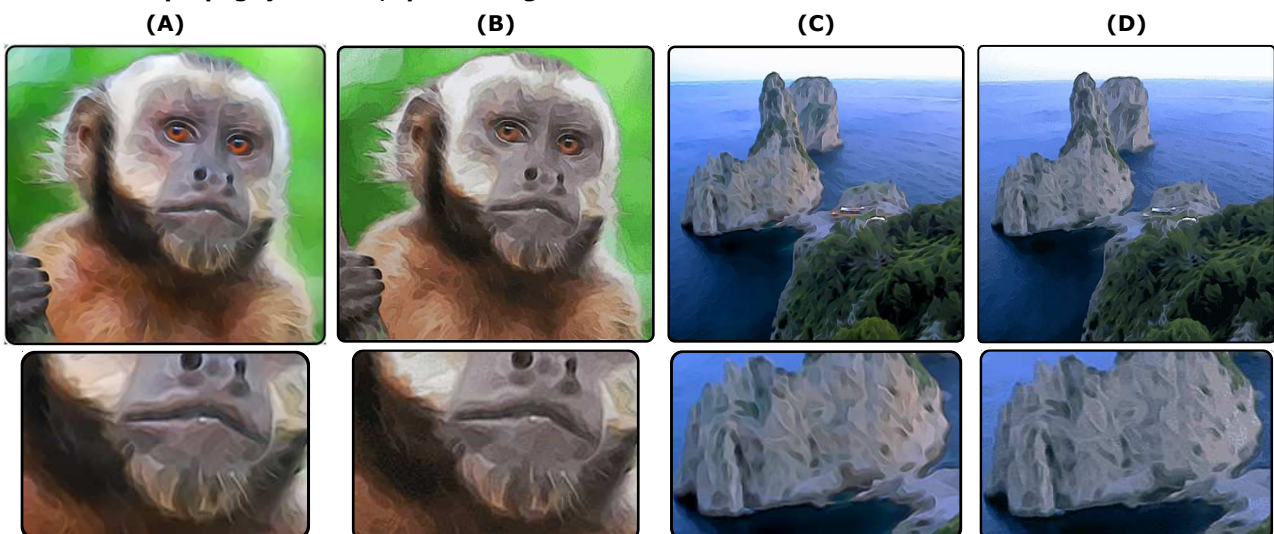
cromática da região de mar da ilha de Capri.

Figura C.1 – Resultados da aplicação de quantização cromática, com algoritmo Denis Lee v3.0, *calculated palette*, 64 cores, 6 bits por cor.



No *plug-in* Ximagic Quantizer são disponibilizados nove técnicas destinadas à aplicação de difusão de erro, e.g.: *Floyd-Steinberg*, *Ulichney*, *Fan*, *Riemersma*. O processo de *dithering* utilizando a técnica *Riemersma* diferencia-se da técnica *Floyd-Steinberg* por reduzir a difusão de erro a uma vizinhança próxima ao pixel a ser processado. Nas Figura C.2A Figura C.2C são apresentados resultados obtidos a partir da utilização da abordagem proposta destinada à renderização não fotorrealística. Nas Figura C.2B e Figura C.2D são apresentados resultados obtidos a partir da utilização do *plug-in* Ximagic Quantizer: *dithering*, propagação=50%, *space filling curve*, algoritmo *Riemersma*. Observa-se que o efeito de quantização (e.g., transição abrupta entre regiões de cor distintas) é reduzido após a utilização de uma técnica de *dithering*.

Figura C.2 – Resultados da aplicação de quantização cromática e *dithering*, com algoritmo *Riemersma*, propagação=50%, *space filling curve*.



Apêndice D

Caracterização dos Parâmetros das Abordagens Comparadas

Este apêndice contém detalhes relativos aos parâmetros e valores associados a cada uma das abordagens comparadas no processo de avaliação (ver Seção 5.2, p. 146). O detalhamento consistiu na apresentação de todos os parâmetros, na apresentação de todos os valores utilizados no processo de pré-seleção e no destaque daqueles valores cuja variação resultou em renderizações significativamente diferentes, do ponto de vista subjetivo, da imagem de entrada.

D.1 Abordagem proposta por Winnemöller, Olsen e Gooch (2006)

A abordagem proposta por Winnemöller, Olsen e Gooch (2006) contém três módulos principais (apresentados no Capítulo 2), destinados: (i) à filtragem por meio da utilização do filtro bilateral (TOMASI; MANDUCHI, 1998); (ii) à detecção de bordas por meio da diferença-de-Gaussianas (DoG); e (iii) à redução de cores por meio da pseudoquantização do canal de luminância.

Três parâmetros são inerentes ao filtro bilateral: (i) as dimensões da vizinhança do pixel a ser filtrado (ω), (ii) a contribuição da distância radial ao processo de suavização (σ_{radial}); e (iii) a contribuição da distância espacial ao processo de suavização (σ_{espacial}).

No módulo para a detecção de bordas, no qual ocorre a filtragem da imagem por diferença de Gaussianas (DoG), os seguintes parâmetros característicos são: (i) a cor da borda; (ii) a espessura da borda; (iii) a sensibilidade do detector de bordas (τ); valor máximo para a computação do gradiente das bordas (∇_{max}); e (iv) valores para os desvios-padrão das máscaras Gaussianas (σ_a e σ_b).

No módulo destinado à pseudoquantização do canal de luminância, há dois parâmetros: (i) a quantidade de cores presente na renderização (n_{bins}); e (ii) o grau de inclinação da equação da tangente hiperbólica utilizada para suavizar as transições entre as cores (ϕ_q).

Os valores dos parâmetros supramencionados estão sumariados na Tabela D.1, na qual se destacam os valores cuja variação resultou em renderizações significativamente diferentes, do ponto de vista subjetivo, da imagem de entrada.

Tabela D.1 – Parâmetros da abordagem proposta por Winnemöller, Olsen e Gooch (2006).

Parâmetro	Valor
ω	3x3 ; 5x5
σ_{radial}	0,1 ; 0,3
σ_{espacial}	5
τ	0,99
∇_{max}	0,3
σ_a	1
σ_b	1,6* σ_a
\mathbf{n}_{bins}	20
φ_q	3,4

D.2 Abordagem proposta por Kyprianidis (2009)

A abordagem proposta por Kyprianidis, Kang e Döllner (2009) fundamenta-se na utilização de uma máscara anisotrópica de formato elipsoidal, subdividida em setores triangulares. Os parâmetros associados a esta abordagem são: (i) o valor do desvio padrão utilizado na suavização Gaussiana do tensor de estrutura (σ_t); (ii) o valor para o fator de ponderação da contribuição dos valores de setores da máscara em formato de elipse (α); (iii) a quantidade de setores na qual a máscara em formato de elipse é subdividida (\mathbf{N}); (iv) o raio da máscara em formato de elipse, em pixels (σ_r); (v) o fator de suavização entre as regiões que compõe a máscara em formato de elipse (s); e (vi) um valor para o aguçamento da imagem de saída (q).

Os valores para todos os parâmetros inerentes à abordagem proposta por Kyprianidis, Kang e Döllner (2009) são apresentados na Tabela D.2, sendo destacados aqueles cuja variação resultou em renderizações significativamente diferentes, do ponto de vista subjetivo, da imagem de entrada.

Tabela D.2 – Parâmetros da abordagem proposta por Kyprianidis, Kang e Döllner (2009).

Parâmetro	Valor
σ_t	2
α	1 ; 2
\mathbf{N}	3 ; 8
σ_r	6 ; 8
s	33,33%
q	8

D.3 Abordagem proposta por Gastal e Oliveira (2011)

A abordagem proposta por Gastal e Oliveira (2011) realiza a renderização não fotorrealística por meio de uma transformação de domínio, na qual a imagem de entrada (bidimensional) é transcrita de forma unidimensional.

A nomenclatura utilizada para a descrição dos parâmetros da abordagem proposta por Gastal e Oliveira (2011) assemelha-se àquela adotada na descrição do filtro bilateral: (i) a contribuição da distância radial ao processo de suavização (σ_{radial}); (ii) a contribuição da distância espacial ao processo de suavização (σ_{espacial}); (iii) a quantidade de iterações a serem aplicadas à imagem de entrada (N); (iii) o tipo de interpolação utilizada no processo de filtragem (t). O tipo de interpolação utilizada pode ser a convolução normalizada (*normalized convolution* - NC), convolução interpolada (*interpolated convolution* - IC) ou a convolução interpolada e a filtragem recursiva (*recursive filtering* - RC).

Os valores de todos os parâmetros inerentes à abordagem proposta por Gastal e Oliveira (2011) são apresentados na Tabela D.3, na qual se destacam os valores cuja variação resultou em renderizações significativamente diferentes, do ponto de vista subjetivo, da imagem de entrada.

Tabela D.3 – Parâmetros da abordagem proposta por Gastal e Oliveira (2011).

Parâmetro	Valor
σ_{radial}	25 ; 40
σ_{espacial}	0,3 ; 0,5
N	3 ; 6
t	NC

D.4 Abordagem proposta por Kyprianidis e Döllner (2008)

A abordagem proposta por Kyprianidis e Döllner (2008) realiza a estilização de imagens por meio do emprego do filtro bilateral, em conjunção à sobreposição de bordas detectadas com a utilização do filtro FDoG. Esta abordagem assemelha-se àquela proposta por Winnemöller, Olsen e Gooch (2006), na qual estão presentes nove parâmetros. Essa quantidade de parâmetros é incrementada na abordagem proposta por Kyprianidis e Döllner (2008), a qual totaliza treze. A seguir, apenas os parâmetros adicionais são caracterizados.

O módulo destinado a extração do ETF é sucedido por um módulo destinado à suavização do tensor de fluxo. A quantidade de suavização aplicada ao tensor de estrutura é indicada por um parâmetro, o σ_{etf} .

A filtragem com o filtro bilateral fundamentado no fluxo ETF é encapsulado em

um módulo que possui quatro parâmetros. Na filtragem por meio do filtro bilateral, utilizam-se as distâncias espacial e radial entre os pixels, de forma iterativa. Os parâmetros característicos são: (i) a contribuição da distância radial ao processo de suavização (σ_{radial}); (ii) a contribuição da distância espacial ao processo de suavização (σ_{espacial}); (ii) a quantidade de iterações do filtro bilateral após a pseudoquantização das cores (n_a); a quantidade de iterações do filtro bilateral após a aplicação do filtro de detecção de bordas (n_e).

No módulo destinado à detecção de bordas, no qual ocorre a filtragem da imagem por diferença de Gaussianas baseada em fluxo (FDoG), os parâmetros característicos são: (i) a cor da borda; (ii) a espessura da borda; (iii) a sensibilidade do detector de bordas (τ); (iv) valores para os desvios-padrão das máscaras Gaussianas (σ_a e σ_b); (v) a quantidade de iterações do filtro FDoG (n_{fdog}); e (vi) a dimensão da suavização aplicada ao longo da linha de borda (σ_m).

O módulo destinado à pseudoquantização do canal de luminância possui três parâmetros, a saber: (i) a quantidade de cores a serem reduzidas no processo de pseudoquantização (n_{bins}); e (ii) o grau de inclinação da equação da tangente hiperbólica utilizada para suavizar as transições entre as cores (ϕ_q); e (iii) as dimensões da máscara da suavização aplicada à imagem com cores reduzidas (ω_s).

A quantidade de suavização aplicada à imagem, ao final do processamento, é indicada por um parâmetro (ω_{final}) associado às dimensões da máscara de suavização.

Os valores de todos os parâmetros inerentes à abordagem proposta por Kyprianidis e Döllner (2008) são apresentados na Tabela D.4, na qual destacam-se aqueles cuja variação resultou em renderizações significativamente diferentes, do ponto de vista subjetivo, da imagem de entrada.

Tabela D.4 – Parâmetros da abordagem proposta por Kyprianidis e Döllner (2008).

Parâmetro	Valor	Parâmetro	Valor
σ_{etf}	2	σ_b	$1,6 * \sigma_a$
σ_{radial}	0,04	n_{fdog}	1
σ_{espacial}	3	σ_m	3
n_a	4 ; 6	n_{bins}	8 ; 12
n_e	1 ; 3	ϕ_q	3,4
τ	0,99	ω_s	3x3
σ_a	1	ω_{final}	3x3

Apêndice E

Documentação Pré-teste

Termo de Confidencialidade

Comprometo-me a manter completo e absoluto sigilo, em relação a quaisquer dados, materiais, informações transmitidas, documentos, especificações técnicas ou comerciais, de que venha a ter conhecimento ou acesso de forma verbal e/ou escrita; ou que a mim venha a ser confiado em razão da sessão de votação intitulada "Avaliação de Abstrações de Vídeos Digitais", realizada em ____ de _____ de 2013, não podendo, sob qualquer pretexto, reproduzir, divulgar, ceder, vender, doar, explorar, comercializar, revelar, utilizar ou de qualquer modo dar conhecimento a terceiros estranhos.

Declaro estar ciente de que, na forma da lei, sou responsável civilmente pela divulgação indevida, descuidada ou incorreta utilização das informações de natureza confidencial que me tenham sido reveladas.

Campina Grande, ____ de _____ de 2013.

Nome: _____

CPF: _____

RG: _____

Termo de Cessão de Dados

Autorizo a utilização dos dados de votos e de registros computacionais capturados durante a sessão de votação intitulada "Avaliação de Abstrações de Vídeos Digitais", realizada em ____ de _____ de 2013, para fins de análise de dados e geração de relatórios que serão utilizados no trabalho de doutorado do aluno Francisco de Assis Pereira Vasconcelos de Arruda (COPIN/UFCG).

Campina Grande, ____ de _____ de _____.

Nome: _____

CPF: _____

RG: _____

Termo de Consentimento Livre e Esclarecido

Avaliação de Abstrações de Vídeos Digitais

Você está sendo convidado (a) a participar da pesquisa acima citada. O documento abaixo contém informações necessárias a respeito da pesquisa que está sendo conduzida. Sua colaboração neste estudo será de extrema importância, e a desistência em participar, a qualquer momento, não causará nenhum prejuízo a você.

Eu, _____, abaixo assinado(a), concordo de livre e espontânea vontade em participar como voluntário(a) da sessão de votação referente ao estudo intitulado "Avaliação de Abstrações de Vídeos Digitais". Declaro que obtive todas as informações necessárias, bem como todos os eventuais esclarecimentos quanto às dúvidas por mim apresentadas.

Estou ciente de que:

- I) **As seções de votação se fazem necessárias para que se possa validar uma nova abordagem para a criação de abstrações de vídeos digitais, proposta no trabalho de doutorado do aluno Francisco de Assis Pereira Vasconcelos de Arruda (COPIN/UFCG). A validação se dará a partir de votação envolvendo a apresentação de vídeos, par a par, gerados a partir de cinco abordagens de abstração de vídeos digitais, os quais são avaliados de acordo com três critérios: (i) aspectos estéticos; (ii) coerência temporal e (iii) preferência visual direta.**
- II) **Após a realização da votação, responderei a um questionário de levantamento do perfil do usuário. Além disso, responderei a uma pergunta a respeito do contexto de uso das abordagens melhor avaliadas, e responderei a uma pergunta a respeito de melhorias no processo de votação.**
- III) **Tenho a liberdade de desistir ou de interromper a colaboração neste estudo no momento em que desejar, sem necessidade de qualquer explicação;**
- IV) **Concordo que meu nome e assinatura sejam divulgados na tese submetida à COPIN (Coordenação de Pós-Graduação em Ciência da Computação), como prova da minha participação na votação.**
- V) **Observações Complementares:**

Caso me sinta prejudicado (a) por participar desta pesquisa, poderei recorrer ao diretor do Centro de Engenharia Elétrica e Informática, ao CEP/HUAC, do Comitê de Ética em Pesquisas em Seres Humanos do Hospital Universitário Alcides Carneiro ou a Delegacia Regional de Campina Grande.

Campina Grande, ____ de _____ de 2013.

Nome: _____

CPF: _____

RG: _____

Telefone: _____

Testemunha: _____

Responsáveis pelo projeto: _____

José Eustáquio Rangel de Queiroz
Herman Martins Gomes
(Orientadores)

Apêndice F

Questionário Pós-teste

Delineamento do Perfil do Usuário

Avaliação de Abstrações de Vídeos Digitais

Obs.: Marque apenas 1 alternativa em cada item.

1 - Seu grau de instrução:

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

2 - Você é do sexo:

- Masculino
- Feminino

3 - Você é:

- Destro (Direito)
- Canhoto (Esquerdo)
- Ambidestro (Direito e Esquerdo)

4 - Você utiliza óculos ou lentes de contato?

- Sim
- Não

5 - Você pertence à faixa etária de:

- Menos de 18 anos
- 18-23 anos
- 24-29 anos
- 30 -35 anos
- Acima de 35 anos

**6 - Você tem experiência prévia no uso de sistemas computacionais (computador)?
Caso sua resposta seja Não, assinale a opção "Não se aplica" nas questões 7, 8, 9 e 10.**

- Sim
- Não

7 - Há quanto tempo você utiliza sistemas computacionais (computador)?

- Menos de 3 meses

- Entre 3 meses e 1 ano
- Mais de 1 ano
- Não se aplica

8 - Com que frequência você utiliza sistemas computacionais (computador)?

- Diariamente
- Ocasionalmente
- Algumas vezes por semana
- Algumas vezes por mês
- Não se aplica

9 - Qual a plataforma computacional que você utiliza com mais frequência?

- Windows
- Linux
- Outra
- Não se aplica

10 - Qual o seu nível de conhecimento em Informática?

- Básico
- Intermediário
- Avançado
- Não se aplica

11 - Em qual contexto você usaria as rederizações dos vídeos melhor avaliados durante a votação?

12 - Houve algum aspecto desagradável durante a votação? Em caso afirmativo, qual?

Apêndice G

Testes F ANOVA Fator Único Intra-abordagens

Este apêndice contém os resultados dos testes F ANOVA fator único para cada uma das abordagens consideradas no experimento de avaliação conduzido (Tabelas G.1 a G.5).

Tabela G.1 – Teste ANOVA para valores da abordagem Transformada de Domínio.

Aspecto	ANOVA: fator único						$\alpha = 0,05$
	Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Aspecto 1	Entre grupos	15,32184	3	5,107279	3,334524	0,021923	2,682809
	Dentro dos grupos	177,6698	108	1,531636			
	Total	192,9917	111				
Aspecto 2	Entre grupos	48,44804	3	16,14935	10,50402	3,64E-06	2,682809
	Dentro dos grupos	178,3436	108	1,537445			
	Total	226,7917	111				
Aspecto 3	Entre grupos	21,2533	3	7,084433	5,038178	0,002562	2,682809
	Dentro dos grupos	163,1134	108	1,40615			
	Total	184,3667	111				

Tabela G.2 – Teste ANOVA para valores da abordagem Kuwahara Anisotrópico.

Aspecto	ANOVA: fator único						$\alpha = 0,05$
	Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Aspecto 1	Entre grupos	24,68763	3	8,229209	5,680521	0,001095	2,682809
	Dentro dos grupos	188,3273	108	1,448672			
	Total	213,0149	111				
Aspecto 2	Entre grupos	6,078016	3	2,026005	1,004196	0,39323	2,682809
	Dentro dos grupos	262,2802	108	2,01754			
	Total	268,3582	111				
Aspecto 3	Entre grupos	14,83397	3	4,944656	3,307164	0,022316	2,682809
	Dentro dos grupos	194,3675	108	1,495135			
	Total	209,2015	111				

Tabela G.3 – Teste ANOVA para valores da abordagem Bilateral+DoG.

Aspecto	ANOVA: fator único						$\alpha = 0,05$
	Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Aspecto 1	Entre grupos	15,32184	3	5,107279	3,334524	0,021923	2,682809
	Dentro dos grupos	177,6698	108	1,531636			
	Total	192,9917	111				
Aspecto 2	Entre grupos	37,06743	3	12,35581	8,248250	5,54E-05	2,682809
	Dentro dos grupos	158,7871	108	1,497992			
	Total	195,8545	111				
Aspecto 3	Entre grupos	35,41762	3	11,80587	7,094210	0,000217	2,682809
	Dentro dos grupos	176,4006	108	1,664156			
	Total	211,8182	111				

Tabela G.4 – Teste ANOVA para valores da Flow Bilateral+FDoG.

Aspecto	ANOVA: fator único						$\alpha = 0,05$
	Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Aspecto 1	Entre grupos	19,3174	3	6,439133	4,095809	0,008522	2,682809
	Dentro dos grupos	169,7897	108	1,572127			
	Total	189,1071	111				
Aspecto 2	Entre grupos	8,612683	3	2,870894	1,954170	0,125232	2,682809
	Dentro dos grupos	158,6641	108	1,469112			
	Total	167,2768	111				
Aspecto 3	Entre grupos	19,86227	3	6,620757	4,231145	0,007194	2,682809
	Dentro dos grupos	168,9949	108	1,564767			
	Total	188,8571	111				

Tabela G.5 – Teste ANOVA para valores da Abordagem Proposta.

Aspecto	ANOVA: fator único						$\alpha = 0,05$
	Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Aspecto 1	Entre grupos	16,91568	3	5,638561	3,406036	0,020721	2,682809
	Dentro dos grupos	158,9243	108	1,655462			
	Total	175,84	111				
Aspecto 2	Entre grupos	6,807971	3	2,269324	1,100600	0,352815	2,682809
	Dentro dos grupos	197,942	108	2,061896			
	Total	204,75	111				
Aspecto 3	Entre grupos	16,36907	3	5,456355	3,213146	0,02635	2,682809
	Dentro dos grupos	163,0209	108	1,698135			
	Total	179,39	111				

Apêndice H

Testes de Tukey-Kramer Intra-abordagens

Neste apêndice, são apresentados os resultados dos testes de Tukey-Kramer para cada uma das abordagens consideradas na avaliação realizada (Tabela H.2), os quais visaram a ressaltar as diferenças entre médias de cada um dos grupos intra-abordagem considerados na avaliação conduzida nesta pesquisa. Visando a facilitar a compreensão destas tabelas apresentam-se, inicialmente, os grupos de valores de parâmetros considerados (Tabela H.1).

Tabela H.1 – Grupos comparados nos testes ANOVA e Tukey-Kramer.

Transf. de Domínio				Kuwahara Anisotrópico			
$\sigma_{radial}=25$		$\sigma_{radial}=40$		$\alpha=1$		$\alpha=2$	
$\sigma_{espacial}=0,3$		$\sigma_{espacial}=0,5$		$N=3$		$N=8$	
$N=3$	$N=6$	$N=3$	$N=6$	$\sigma_r=6$	$\sigma_r=8$	$\sigma_r=6$	$\sigma_r=8$
G1	G2	G3	G4	G1	G2	G3	G4

Flow Bilateral+FDoG				Bilateral+DoG			
$n_a=4$		$n_a=6$		$\omega=3 \times 3$		$\omega=5 \times 5$	
$N_{bins}=8$	$N_{bins}=12$	$N_{bins}=8$	$N_{bins}=12$	$\sigma_{rad}=0,1$	$\sigma_{rad}=0,3$	$\sigma_{rad}=0,1$	$\sigma_{rad}=0,3$
G1	G2	G3	G4	G1	G2	G3	G4

Abordagem proposta			
$\rho_{cef}=5$		$\rho_{cef}=7$	
$\psi_{bordas}=W$		$\psi_{bordas}=T$	
$B=10$	$B=50$	$B=10$	$B=50$
G1	G2	G3	G4

Tabela H.2 – Grupos de similaridade obtidos no teste de Tukey-Kramer, para aspectos intra-abordagem.

Abordagem	Aspecto	Grupo de parâmetros	Média estimada	Grupos	
Transformada de Domínio	Aspecto 1	G3	2,636	A	
		G4	2,675	A	
		G2	3,167	A	
		G1	3,697		B
	Aspecto 2	G1	2,000	A	
		G2	2,333		B
		G4	3,325		B
		G3	3,606		B
	Aspecto 3	G4	2,275	A	
		G3	2,455	A	
		G2	2,875	A	
		G1	3,333		B
Kuwahara Anisotrópico	Aspecto 1	G4	2,383	A	
		G2	2,517	A	
		G3	3,091	A	
		G1	3,571		B
	Aspecto 2	G1	-	-	-
		G2	-	-	-
		G3	-	-	-
		G4	-	-	-
	Aspecto 3	G3	2,348	A	
		G4	2,435	A	
		G2	2,690	A	
		G1	3,171		B
Flow Bilateral+FDoG	Aspecto 1	G2	2,680	A	
		G3	2,808	A	
		G4	3,267	A	
		G1	3,800		B
	Aspecto 2	G1	-	-	-
		G2	-	-	-
		G3	-	-	-
		G4	-	-	-
	Aspecto 3	G3	2,231	A	
		G2	2,480	A	
		G4	2,767	A	
		G1	3,333		B
Bilateral+DoG	Aspecto 1	G4	2,382	A	
		G3	3,000	A	
		G1	3,455	A	
		G2	3,619		B
	Aspecto 2	G2	2,500	A	
		G1	2,545		B
		G3	3,381		B
		G4	3,952		B
	Aspecto 3	G4	2,029	A	
		G3	2,524	A	
		G1	2,905	A	
		G2	3,515		B
Abordagem proposta	Aspecto 1	G4	2,593	A	
		G3	2,773	A	
		G1	3,130	A	
		G2	3,704		B
	Aspecto 2	G1	-	-	-
		G2	-	-	-
		G3	-	-	-
		G4	-	-	-
	Aspecto 3	G4	2,296	A	
		G3	2,409		B
		G1	2,565		B
		G2	3,333		B

Apêndice I

Testes F ANOVA Fator Único Interabordagens

Este apêndice contém os resultados dos testes F ANOVA fator único referentes às abordagens avaliadas nesta pesquisa (Tabelas I.1 a I.3). Apresentam-se todos os valores para os testes F ANOVA envolvendo as cinco abordagens comparadas, os quais visaram a ressaltar a existência de diferenças entre médias entre cada uma das abordagens consideradas.

Tabela I.1 – Teste F ANOVA interabordagem para o Aspecto 1.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	34,60546	4	8,651366	5,374824	0,000275	2,380133
Dentro dos grupos	1746,426	1085	1,609609			
Total	1781,031	1089				

Tabela I.2 – Teste F ANOVA interabordagem para o Aspecto 2.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	22,92964	4	5,732410	3,279932	0,011014	2,380133
Dentro dos grupos	1927,737	1085	1,747722			
Total	1950,667	1089				

Tabela I.3 – Teste F ANOVA interabordagem para o Aspecto 3.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	72,37548	4	18,09387	10,62794	1,89E-08	2,380133
Dentro dos grupos	1847,192	1085	1,702481			
Total	1919,568	1089				

Apêndice J

Testes de Tukey-Kramer Interabordagens

Este apêndice contém os resultados dos testes de Tukey-Kramer para cada uma das abordagens consideradas no experimento de avaliação conduzido, visando a ressaltar as diferenças entre médias de cada um dos grupos intra-abordagem considerados. Com o propósito de facilitar a compreensão destas tabelas, apresentam-se, nas Tabelas J.1 a J.3, os grupos de valores de parâmetros considerados. Visando a facilitar a compreensão destas tabelas, apresenta-se na Tabela J.4 um sumário dos grupos de valores de parâmetros considerados.

Tabela J.1 – Teste de Tukey-Kramer para o Aspecto 1.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Abord. Proposta vs Flow Bilateral+FDoG	0,231	1,884	2,728	0,326	Não
Abord. Proposta vs Bilateral+DoG	0,384	3,194	2,728	0,012	Sim
Abord. Proposta vs Kuwahara Anisotr.	0,436	3,473	2,728	0,005	Sim
Abord. Proposta vs Transf. Domínio	0,516	4,154	2,728	0,000	Sim
Flow Bilateral+FDoG vs Bilateral+DoG	0,153	1,302	2,728	0,690	Não
Flow Bilateral+FDoG vs Kuwahara Anisotr.	0,205	1,669	2,728	0,453	Não
Flow Bilateral+FDoG vs Transf. Domínio	0,285	2,347	2,728	0,130	Não
Bilateral+DoG vs Kuwahara Anisotr.	0,052	0,434	2,728	0,993	Não
Bilateral+DoG vs Transf. Domínio	0,132	1,113	2,728	0,800	Não
Kuwahara Anisotr. vs Transf. Domínio	0,080	0,644	2,728	0,968	Não
Valor crítico d de Tukey:	-	-	3,858	-	-

Tabela J.2 – Teste de Tukey-Kramer para o Aspecto 2.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Kuwahara Anisotr. vs Flow Bilateral+FDoG	0,407	3,183	2,728	0,013	Sim
Kuwahara Anisotr. vs Transf. Domínio	0,397	3,062	2,728	0,019	Sim
Kuwahara Anisotr. vs Bilateral+DoG	0,247	1,975	2,728	0,278	Sim
Kuwahara Anisotr. vs Abord. Proposta	0,225	1,753	2,728	0,401	Sim
Flow Bilateral+FDoG vs Transf. Domínio	0,011	0,085	2,728	1,000	Não
Flow Bilateral+FDoG vs Bilateral+DoG	0,160	1,308	2,728	0,686	Não
Flow Bilateral+FDoG vs Abord. Proposta	0,182	1,457	2,728	0,590	Não
Transf. Domínio vs Bilateral+DoG	0,149	1,203	2,728	0,750	Não
Transf. Domínio vs Abord. Proposta	0,172	1,354	2,728	0,657	Não
Bilateral+DoG vs Abord. Proposta	0,023	0,185	2,728	1,000	Não
Valor crítico d de Tukey:	-	-	3,858	-	-

Tabela J.3 – Teste de Tukey-Kramer para o Aspecto 3.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Abord. Proposta vs Flow Bilateral+FDoG	0,393	3,061	2,728	0,019	Sim
Abord. Proposta vs Kuwahara Anisotr.	0,397	3,025	2,728	0,021	Sim
Abord. Proposta vs Bilateral+DoG	0,589	4,685	2,728	< 0,0001	Sim
Abord. Proposta vs Transf. Domínio	0,698	5,375	2,728	< 0,0001	Sim
Flow Bilateral+FDoG vs Kuwahara Anisotr.	0,004	0,033	2,728	1,000	Não
Flow Bilateral+FDoG vs Bilateral+DoG	0,196	1,599	2,728	0,498	Não
Flow Bilateral+FDoG vs Transf. Domínio	0,305	2,406	2,728	0,114	Não
Kuwahara Anisotr. vs Bilateral+DoG	0,192	1,526	2,728	0,546	Não
Kuwahara Anisotr. vs Transf. Domínio	0,301	2,318	2,728	0,139	Não
Bilateral+DoG vs Transf. Domínio	0,109	0,880	2,728	0,904	Não
Valor crítico d de Tukey:	-	-	3,858	-	-

Tabela J.4 – Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.

Aspecto	Categoria	Média estimada	Grupos	
Aspecto 1	Abordagem Proposta	3,441	A	
	Flow Bilateral+FDoG	3,210	A	
	Bilateral+DoG	3,057		B
	Kuwahara Anisotrópico	3,005		B
	Domain Transform	2,925		B
Aspecto 2	Flow Bilateral+FDoG	3,246	B	
	Domain Transform	3,235	B	
	Bilateral+DoG	3,086	B	
	Abordagem Proposta	3,063	B	
	Kuwahara Anisotrópico	2,838		A
Aspecto 3	Abordagem Proposta	3,299	A	
	Flow Bilateral+FDoG	2,906		B
	Kuwahara Anisotrópico	2,902		B
	Bilateral+DoG	2,710		B
	Domain Transform	2,601		B

Apêndice K

Testes F ANOVA Fator Único Interabordagens com Maiores Valores Intra-Abordagens

Este apêndice contém os resultados dos testes F ANOVA fator único único (Tabelas K.1 a K.3) associados às abordagens consideradas no experimento de avaliação conduzido. Apresentam-se todos os valores para os testes F ANOVA realizados para as cinco abordagens comparadas. Tais testes visaram a ressaltar a existência de diferenças entre médias interabordagem, quando considerados os melhores valores intra-abordagem.

Tabela K.1 – Teste F ANOVA interabordagem, Aspecto 1, considerando maiores valores intra-abordagem.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	17,03985	4	4,259963	2,679864	0,031769	2,400041
Dentro dos grupos	505,4988	318	1,589619			
Total	522,5387	322				

Tabela K.2 – Teste F ANOVA interabordagem, Aspecto 2, considerando maiores valores intra-abordagem.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	6,047258	4	1,511815	0,842104	0,499293	2,400041
Dentro dos grupos	570,9001	318	1,795283			
Total	576,9474	322				

Tabela K.3 – Teste F ANOVA interabordagem, Aspecto 3, considerando maiores valores intra-abordagem.

ANOVA: fator único						$\alpha = 0,05$
Fonte da variação	SQ	gl	MQ	F	valor-P	F crítico
Entre grupos	29,46267	4	7,365667	4,100028	0,003048	2,400041
Dentro dos grupos	483,2563	318	1,796492			
Total	512,719	322				

Apêndice L

Testes de Tukey-Kramer Interabordagens com Maiores Valores Intra-abordagens

Este apêndice contém os resultados dos testes de Tukey-Kramer para cada uma das abordagens consideradas na avaliação experimental conduzida, os quais visaram a ressaltar as diferenças entre médias de cada um dos grupos interabordagem considerados, quando utilizados os valores dos parâmetros intra-abordagem mais bem avaliados. Visando a facilitar a compreensão destas tabelas apresentam-se, na Tabela L.3, os grupos de valores de parâmetros considerados. Conforme ressaltado no Capítulo 5 (ver Seção 5.2.3, p. 161), há diferenças estatisticamente significativas entre as médias dos Aspectos 1 e 3. Assim sendo, não apresenta-se, neste apêndice, os resultados tabulados do teste de Tukey-Kramer para o aspecto 2.

Tabela L.1 – Teste de Tukey-Kramer para o Aspecto 1.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Kuwahara Anisotr. vs Abord. proposta	-0,706	-2,708	1,932	0,053	Sim
Kuwahara Anisotr. vs Transf. Domínio	-0,562	-2,259	1,932	0,158	Sim
Kuwahara Anisotr. vs Bilateral+DoG	-0,530	-2,198	1,932	0,181	Sim
Kuwahara Anisotr. vs Flow Bilateral+FDoG	-0,476	-1,976	1,932	0,278	Sim
Flow Bilateral+FDoG vs Abord. Proposta	-0,230	-0,958	1,932	0,874	Não
Flow Bilateral+FDoG vs Transf. Domínio	-0,085	-0,376	1,932	0,996	Não
Flow Bilateral+FDoG vs Bilateral+DoG	-0,053	-0,244	1,932	0,999	Não
Bilateral+DoG vs Abord. Proposta	-0,176	-0,736	1,932	0,948	Não
Bilateral+DoG vs Transf. Domínio	-0,032	-0,140	1,932	1,000	Não
Transf. Domínio vs Abord. proposta	-0,145	-0,585	1,932	0,977	Não
Valor crítico d de Tukey:	-	-	3,858	-	-

Tabela L.2 – Teste de Tukey-Kramer para o Aspecto 3.

Contraste	Diferença	Diferença padronizada	Valor crítico	Pr > Dif	Significante para $\alpha = 0,05$
Abord. Proposta vs Flow Bilateral+FDoG	0,393	3,061	2,728	0,019	Sim
Abord. Proposta vs Kuwahara Anisotr.	0,397	3,025	2,728	0,021	Sim
Abord. Proposta vs Bilateral+DoG	0,589	4,685	2,728	< 0,0001	Sim
Abord. Proposta vs Transf. Domínio	0,698	5,375	2,728	< 0,0001	Sim
Flow Bilateral+FDoG vs Kuwahara Anisotr.	0,004	0,033	2,728	1,000	Não
Flow Bilateral+FDoG vs Bilateral+DoG	0,196	1,599	2,728	0,498	Não
Flow Bilateral+FDoG vs Transf. Domínio	0,305	2,406	2,728	0,114	Não
Kuwahara Anisotr. vs Bilateral+DoG	0,192	1,526	2,728	0,546	Não
Kuwahara Anisotr. vs Transf. Domínio	0,301	2,318	2,728	0,139	Não
Bilateral+DoG vs Transf. Domínio	0,109	0,880	2,728	0,904	Não
Valor crítico d de Tukey:	-	-	3,858	-	-

Tabela L.3 – Grupos de similaridade obtidos por meio do teste de Tukey-Kramer.

Aspecto	Categoria	Média estimada	Desvio padrão	Grupos	
Aspecto 1	Kuwahara Anisotrópico	2,827	1,175	A	
	Flow Bilateral+FDoG	3,013	0,906		B
	Bilateral+DoG	3,118	0,952		B
	Transformada de Domínio	3,138	1,269		B
	Abordagem proposta	3,382	1,004		B
Aspecto 2	Flow Based	-	-	-	-
	Transformada de Domínio	-	-	-	-
	Bilateral+DoG	-	-	-	-
	Abordagem Proposta	-	-	-	-
	Kuwahara Anisotrópico	-	-	-	-
Aspecto 3	Abordagem Proposta	3,345	1,194	A	
	Kuwahara Anisotrópico	2,885	1,407		B
	Transformada de Domínio	2,761	1,157		B
	Flow Bilateral+FDoG	2,711	1,322		B
	Bilateral+DoG	2,644	1,149		B

Apêndice M

Tempo de Execução da Abordagem Proposta

Este apêndice contém os resultados dos testes de tempo de execução da abordagem proposta. No contexto da implementação computacional, foram realizados testes que visaram à aferição do tempo computacional médio para a renderização de 20 imagens. As imagens renderizadas tinham dimensões médias de 512x415 pixels, com a utilização de três canais do espaço cromático RGB. O computador utilizado para a aferição possuía um processador Intel Core i5 (x86-64, modelo M 460), com *clock* de 2,53 GHz, quatro núcleos de processamento e 4 GB de memória principal, utilizando o sistema operacional Microsoft Windows 7 (versão 64 bits). O tempo médio de execução para cada módulo e o tempo médio total (22,1 segundos por imagem, $\pm 1,8$ segundos) podem ser visualizados no Quadro M.1.

Quadro M.1 – Tempo de execução médio para cada módulo, considerando a renderização de 20 imagens.

Módulo	Tempo de execução médio
Módulo A – Ênfase na orientação e estrutura local	4,5 segundos
Módulo B – Extração de bordas utilizando redes neurais	3,6 segundos
Módulo C – Suavização por difusão anisotrópica modificada (3 canais)	6,5 segundos
Módulo D – Realce de sombras e superposição das bordas	3,5 segundos
Módulo E – Filtragem com uso de médias não locais	3,5 segundos
Módulo F – Ajuste de saturação	0,5 segundos
Tempo médio total:	22,1 segundos

Ressalta-se que a abordagem proposta, em seu estágio atual, afigura-se inadequada à renderização de imagens e vídeos em tempo real. Entretanto, Kyprianidis et al. (2013) afirmaram que a implementação computacional de forma paralela de algoritmos destinados à filtragem espacial, em especial com o emprego de GPU, pode ser realizada de forma direta, com poucas modificações relativas às implementações computacionais existentes em linguagem de programação de alto nível. Com efeito, duas técnicas utilizadas na abordagem proposta já possuem

implementação computacional que faz uso de GPU: (i) CEF utilizada no módulo A; e (ii) filtragem NLM, utilizada no módulo E.

Com a intenção de verificar a viabilidade de tais implementações em GPU, assim como de investigar a possibilidade da redução do tempo médio de execução da renderização de imagens a partir da abordagem proposta, foram testadas as duas implementações supramencionadas. Os testes foram realizados em 20 quadros de vídeo com dimensões médias de 512x415 pixels, em um computador com processador Intel Core2Duo (x86-64 modelo E7200), com *clock* de 2,53 GHz, 2 GB de memória principal e sistema operacional Microsoft Windows 7 (versão 64 bits), equipado com GPU Nvidia GeForce GT240, com 1 GB.

No teste da técnica adotada no módulo A, a implementação da CEF, que faz uso da GPU³⁶, realizou o processamento a uma taxa média de 7 quadros por segundo³⁷, totalizando 0,11 segundos por imagem. No teste da técnica utilizada no módulo E, a implementação da NLM que faz uso da GPU³⁸ realizou o processamento a uma taxa média de 13 FPS, totalizando 0,2 segundos para filtragem de um quadro de vídeo. No tocante à utilização de GPU para a filtragem com a técnica NLM, outras pesquisas podem ser fontes de consulta e investigação de aprimoramentos futuros (GOOSSENS et al., 2010; PALMA et al., 2013).

A partir das modificações nas implementações destes dois módulos, o tempo médio de renderização foi reduzido para 14,4 segundos (\pm 0,9 segundos). Cabe ressaltar que a implementação de forma paralela, seja mediante o emprego de múltiplos núcleos de um processador ou a partir da utilização de placas gráficas dedicadas, pode ter complexidade elevada pelo uso concomitante da detecção de bordas e da difusão anisotrópica inerente à abordagem proposta. Diante do exposto, a implementação da abordagem proposta para a renderização de quadros de vídeos em tempo real afigura-se um objetivo de investigação futura, devendo contemplar uma versão computacional que faça uso de GPU ou programação multinúcleo.

³⁶ Disponível em <https://code.google.com/p/cefabs/>

³⁷ *Frames per second - FPS*

³⁸ Disponível em <http://docs.nvidia.com/cuda/cuda-samples/index.html> - Image denoising - NLM

Apêndice N

Detalhamento do Módulo de Detecção de Bordas

Este apêndice contém a publicação intitulada *Non-photorealistic neural sketching: A case study on frontal-face images*, apresentada no *Journal of the Brazilian Computer Society* (ARRUDA; QUEIROZ; GOMES, 2012). Tal publicação refere-se à abordagem utilizada na geração do mapa de bordas neural, presente no Módulo B da abordagem proposta (apresentada no Capítulo 3, p. 99).

