

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA

Coordenação de Pós-Graduação em Ciência da Computação

DISSERTAÇÃO

ABORDAGEM PARA A ESTILIZAÇÃO DE VÍDEOS DIGITAIS

ALBERTO FAGNER FERREIRA DE BARROS

ORIENTADORES:

JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ

HERMAN MARTINS GOMES

CAMPINA GRANDE, PARAÍBA

FEVEREIRO DE 2019

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
Coordenação de Pós-Graduação em Ciência da Computação

ABORDAGEM PARA A ESTILIZAÇÃO DE VÍDEOS DIGITAIS

ALBERTO FAGNER FERREIRA DE BARROS

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

Área de concentração: Ciência da Computação

Linha de pesquisa: Metodologia e Técnicas da Computação

José Eustáquio Rangel de Queiroz

Herman Martins Gomes

Orientadores

Campina Grande – Paraíba – Brasil

Fevereiro de 2019

B277a Barros, Alberto Fagner Ferreira de.
Abordagem para a estilização de vídeos digitais / Alberto Fagner Ferreira de Barros. – Campina Grande, 2019.
125 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2019.

"Orientação: Prof. Dr. José Eustáquio Rangel de Queiroz, Prof. Dr. Herman Martins Gomes".

Referências.

1. Cartuns. 2. Segmentação. 3. Vídeo digital. I. Queiroz, José Eustáquio Rangel de. II. Gomes, Herman Martins. III. Título.

CDU 621.397:74(043)

"ABORDAGEM PARA A ESTILIZAÇÃO DE VÍDEOS DIGITAIS"

ALBERTO FAGNER FERREIRA DE BARROS

DISSERTAÇÃO APROVADA EM 04/02/2019

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

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

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

EDMAR CANDEIA GURJÃO, D.Sc, UFCG
Examinador(a)

CAMPINA GRANDE - PB

Dedico esta dissertação a Deus e a todos aqueles que me incentivaram nesta caminhada. A meus pais, José e Verônica.

AGRADECIMENTOS

A Deus, que me fortaleceu durante esta jornada acadêmica em busca de conhecimento e crescimento pessoal e que me permitiu, apesar das dificuldades e obstáculos enfrentados, concluir mais uma etapa nesta longa estrada da vida.

A meus orientadores, Rangel e Herman, que foram de grande auxílio durante esses anos do mestrado, tirando-me muitas das dúvidas relacionadas a conceitos e muitas das ideias expostas nesta dissertação, assim como dispoendo sempre de paciência em suas orientações.

A todos os companheiros do Laboratório de Percepção Computacional que, durante minha estadia ali, foram sempre prestativos e bem-humorados. Em especial, a Carlos e a Ronnyldo, pelo auxílio prestado em dúvidas sobre algumas implementações de código.

A João Bosco e Francisco de Assis pela ajuda indispensável no desenvolvimento desta pesquisa.

Às funcionárias da Copin, que me auxiliaram nos diversos processos abertos durante esta jornada.

Ao CNPq pois, a presente pesquisa foi realizada com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

A meus pais...

... amigos...

... e todos que, de alguma maneira, colaboraram para que esta dissertação de mestrado se tornasse possível.

RESUMO

A Renderização Não Fotorrealística é uma área de pesquisa que reside na intersecção entre as áreas de Computação Gráfica, Processamento Digital de Imagens e Arte. A partir de imagens e vídeos capturados de cenas do mundo real, busca-se gerar representações estilizadas e abstratas da realidade, com o propósito de comunicar visualmente a um observador humano ideias ou conceitos. Neste contexto, a presente dissertação tem como objetivo a proposição e a validação de uma abordagem automática para a conversão de vídeos digitais fotorrealísticos em vídeos não fotorrealísticos ao estilo de cartuns. A abordagem proposta consiste das seguintes etapas: (1) suavização dos quadros de vídeo com preservação de bordas, a partir do emprego de um filtro bilateral adaptativo, visando a homogeneizar suas regiões; (2) quantização do espaço cromático, visando à segmentação de regiões e à redução do número de cores dos quadros de vídeo; (3) determinação dos componentes conectados nos quadros, a fim de reduzir efeitos de cintilação e aumentar a coerência temporal do vídeo renderizado; (4) realce dos contornos dos objetos presentes nos quadros, via um detector de bordas; e (5) ajuste do contraste e da saturação dos quadros do vídeo, a fim de melhorar sua visualização. Como resultado da pesquisa, foi implementado um protótipo da abordagem proposta. Além disto, diferentes combinações de parâmetros foram experimentalmente validadas, a partir da sondagem da opinião de voluntários acerca de vídeos selecionados em uma base pública de vídeos digitais. Análises estatísticas sobre os dados coletados do estudo experimental evidenciaram as combinações de parâmetros que produziram os melhores resultados para os participantes do experimento, permitindo, desta forma, identificar as melhores configurações para a utilização satisfatória da abordagem proposta.

PALAVRAS-CHAVE: Cartuns. Segmentação. Vídeo Digital. Filtragem Bilateral.

ABSTRACT

Non-photorealistic rendering is a research area that lies in the intersection of Computer Graphics, Digital Image Processing and Art. Using images and videos captured from real world scenes stylized and abstracted representations of reality are generated to visually communicate ideas or concepts to a human observer. Within this context, the present dissertation propose and validate an automatic approach to photorealistic video cartoonization. The approach is composed by the following steps: (1) Edge preserving smooth filtering (Adaptative Bilateral Filter) to smooth video frame regions, (2) Color quantization for region segmentation and to reduce the number of colors in the video, (3) Determination of connected components to reduce flickering effects and temporal coherence enhancement of the rendered video, (4) Edge enhancement using an edge detector, and (5) Saturation and contrast adjustments in the video frames in order to improve visualization. As a result a prototype implementing the proposed approach has been produced. The prototype parameters were experimentally tested and validated using quantitative and qualitative tests consisting of an assessment of the opinion of volunteers on videos extracted from a public digital database. Statistical analyses were conducted upon the collected data in order to indentify parameter combinations and techniques that yielded the best results for the participants of the experiment, thus indicating the best operating conditions for the successful use of the proposed approach.

KEYWORDS: Cartoonization. Segmentation. Digital Video. Bilateral Filter.

LISTA DE QUADROS

Quadro 1	-	Publicações que utilizam informação espacial	25
Quadro 2	-	Publicações que utilizam informação temporal	26
Quadro 3	-	Parâmetros utilizados pelas abordagens	119

LISTA DE FIGURAS

Figura 1	- Sequência de quadros ilustrando problemas de cintilação	17
Figura 2	- Sequência de quadros ilustrando problemas de coerência espaço-temporal	18
Figura 3	- Exemplo de resultado da abordagem de Ben e Zelnik.....	27
Figura 4	- Exemplo do método de extração de linhas de Wang <i>et al.</i>	28
Figura 5	- Exemplo de resultado da abordagem de Su, Luo e Artusi.	29
Figura 6	- Exemplo da abordagem de Rosin e Lai	30
Figura 7	- Exemplo de resultado da abordagem de Lu <i>et al.</i>	31
Figura 8	- Exemplo da extração de segmentos de Xue <i>et al.</i> (2014) .	32
Figura 9	- Exemplo de resultado da abordagem de Chen, Turk e MacIntyre de 2012	33
Figura 10	- Exemplo de resultado da abordagem de Song e Xu	34
Figura 11	- Exemplo de resultado da abordagem de Zhang <i>et al.</i>	35
Figura 12	- Exemplo de resultado da abordagem de Lin, Lin e Wang ..	36
Figura 13	- Exemplo de resultado da abordagem de Cao <i>et al.</i>	37
Figura 14	- Exemplo de resultado da abordagem de Lu <i>et al.</i> 2015	38
Figura 15	- Exemplo de resultado da abordagem de Kang <i>et al.</i>	39
Figura 16	- Exemplo de resultado da abordagem de Yoon, Lee e Kang	40
Figura 17	- Exemplos de resultados da abordagem de Bénard <i>et al.</i> ...	41
Figura 18	- Exemplo de resultado da abordagem de Klette	42
Figura 19	- Exemplo de resultado da abordagem de Jiang <i>et al.</i>	43
Figura 20	- Exemplo de resultado da abordagem de Gangopadhyay, Pachori e Raman	44
Figura 21	- Imagem digital bidimensional	47
Figura 22	- Vizinhanças de pixels mais comuns utilizadas	48
Figura 23	- Exemplo de aplicação de convolução em uma imagem	49
Figura 24	- Exemplo de aplicação de filtros espaciais	50
Figura 25	- Exemplo de aplicação do filtro bilateral	51
Figura 26	- Espectro visível da luz	52
Figura 27	- Sistema de cor RGB	53
Figura 28	- Sistema de cor HSV	54
Figura 29	- Exemplo de redução de cores	55
Figura 30	- Exemplo de segmentação de cores	55

Figura 31	- Pixel central conectado a 4 pixels adjacentes	56
Figura 32	- Exemplo de 3 regiões de pixels conectadas distintas	57
Figura 33	- Diagrama dos módulos da abordagem proposta	58
Figura 34	- Exemplo da aplicação do FABL no módulo	59
Figura 35	- Ilustração das possíveis cores no sistema RGB em forma de cubo ao utilizar 256 níveis de quantização por canal	61
Figura 36	- Ilustração das possíveis cores no sistema RGB fatiado em cubos de 52x52x52 níveis de cinza	62
Figura 37	- Ilustração das possíveis cores no sistema RGB fatiado em cubos de 32x32x32 níveis de cinza	63
Figura 38	- Exemplo de saída do módulo de segmentação de cores	64
Figura 39	- Pixels pertencentes a um nível de quantização de um quadro de vídeo conectados espacialmente	64
Figura 40	- Pixels de mesmo identificador recoloridos	65
Figura 41	- Aumento da saturação original de um quadro de vídeo em 50%	66
Figura 42	- Detecção de bordas de um quadro de vídeo	67
Figura 43	- Quadro exemplo do processamento realizado pela abordagem proposta	68
Figura 44	- Cintilação provocada pela mudança de nível de quantização de alguns pixels	69
Figura 45	- Quadro inicial de cada um dos 10 vídeos selecionados	74
Figura 46	- Interface gráfica online utilizada nos testes	76
Figura 47	- Foto do ambiente de testes	77
Figura 48	- Gráfico do grau de instrução dos participantes	80
Figura 49	- Gráfico referente ao sexo dos participantes	80
Figura 50	- Gráfico referente a escrita dos participantes	81
Figura 51	- Gráfico referente ao uso de óculos e lentes dos participantes	81
Figura 52	- Gráfico referente a faixa etária dos participantes	82
Figura 53	- Gráfico referente a experiência prévia com sistemas computacionais dos participantes	82
Figura 54	- Gráfico referente ao tempo de uso de sistemas computacionais dos participantes	83
Figura 55	- Gráfico referente a frequência do uso de sistemas computacionais dos participantes	83
Figura 56	- Gráfico referente a plataforma mais utilizada pelos participantes	84
Figura 57	- Gráfico referente a nível de conhecimento em informática dos participantes	84
Figura 58	- Gráfico referente aos contextos de utilização dos vídeos renderizados sugeridos pelos participantes	85
Figura 59	- <i>Boxplot</i> dos votos dos participantes para o critério 1	86
Figura 60	- <i>Boxplot</i> dos votos dos participantes para o critério 2	87

Figura 61	-	<i>Boxplot</i> dos votos dos participantes para o critério 3	87
Figura 62	-	<i>Boxplot</i> dos votos dos participantes para o critério 4	88
Figura 63	-	<i>Boxplot</i> dos votos dos participantes para o critério 5	88
Figura 64	-	<i>Boxplot</i> dos votos dos participantes para o critério 6	89
Figura 65	-	Exemplo de cartuns modernos	103
Figura 66	-	Exemplo de Saída gerada pelo módulo de Subtração de Plano de Fundo	112
Figura 67	-	Exemplo do cálculo de fluxo óptico Farneback de um quadro	113
Figura 68	-	Exemplo de combinação do Fluxo óptico Farneback calculado em 3 quadros consecutivos, utilizando-se uma vizinhança 3x3	114
Figura 69	-	Exemplo de aplicação da operação de morfologia matemática de fechamento em um quadro	114
Figura 70	-	Máscara resultante da combinação das máscaras de MOG2 e cálculo do fluxo óptico Farneback de um quadro	115
Figura 71	-	Exemplo de um quadro de entrada	116
Figura 72	-	Exemplo de aplicação de estilização a um quadro de vídeo, utilizando-se FABL	116
Figura 73	-	Exemplo de lacunas nos objetos detectados em um quadro de vídeo processado	117
Figura 74	-	Exemplo de máscara resultante do processamento de um quadro nos módulos removidos	118
Figura 75	-	Exemplo de <i>ground truth</i> correspondente a um quadro de vídeo marcado manualmente	118
Figura 76	-	Quadro inicial de cada um dos 10 vídeos selecionados processados pelo filtro Anisotrópico de Kuwahara	122
Figura 77	-	Quadro inicial de cada um dos 10 vídeos selecionados processados pelo filtro Coherence-Enhancing Filtering	123
Figura 78	-	Quadro inicial de cada um dos 10 vídeos selecionados processados pelo filtro XDoG	124
Figura 79	-	Quadro inicial de cada um dos 10 vídeos selecionados processados pela abordagem proposta	125

LISTA DE TABELAS

Tabela 1	-	Parâmetros utilizados em cada uma das configurações da abordagem proposta	79
Tabela 2	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 1	90
Tabela 3	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 2	90
Tabela 4	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 3	90
Tabela 5	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 4	91
Tabela 6	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 5	91
Tabela 7	-	Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 6	91
Tabela 8	-	Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 1	92
Tabela 9	-	Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 3	92
Tabela 10	-	Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 4	92
Tabela 11	-	Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 5	92
Tabela 12	-	Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 6	93
Tabela 13	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 1	93
Tabela 14	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 2	94
Tabela 15	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 3	94
Tabela 16	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 4	94
Tabela 17	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 5	94
Tabela 18	-	Teste para verificação de diferenças estatísticas entre as abordagens para o critério 6	95
Tabela 19	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 1	95
Tabela 20	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 2	95
Tabela 21	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 3	96
Tabela 22	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 4	96
Tabela 23	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 5	97
Tabela 24	-	Teste para verificação da mais bem avaliada entre as abordagens para o critério 6	97

Tabela 25 - Tempo de processamento de 10 quadros de um vídeo 98

LISTA DE SIGLAS E ABREVIATURAS

Codec	Codificador/ D ecodificador
FABL	<i>Flow-Based Adaptive Bilateral</i> (Bilateral Adaptativo Baseado em Fluxo)
FBL	<i>Flow-Based Bilateral</i> (Bilateral Baseado em Fluxo)
HSV	<i>Hue Saturation Value</i> (Matiz Saturação Valor)
ITU-R	<i>International Telecommunication Union Radiocommunication</i> (União Internacional de Telecomunicação e Radio Comunicação)
LSD	<i>Line Segment Detector</i> (Detector de Segmento de Linha)
LTV	<i>Local Total Variation</i> (Variação Total Local)
NPR	<i>Non-Photorealistic Rendering</i> (Renderização não fotorrealística)
OpenCV	<i>Open Source Computer Vision Library</i> (Biblioteca de Código Aberto de Visão Computacional)
RGB	<i>Red Green Blue</i> (Vermelho Verde Azul)
XDoG	<i>eXtended Difference-of-Gaussians</i> (Diferença de Gaussianas Estendida)
TCLE	Termo de C onsentimento L ivre e E esclarecido

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	15
1.1 Motivação/Contextualização da Pesquisa.....	15
1.2 Objetivo Geral	20
1.3 Objetivos Específicos.....	20
1.4 Hipóteses	20
1.5 Estrutura do Documento	21
CAPÍTULO 2 - REVISÃO BIBLIOGRÁFICA	22
2.1 Descrição do Processo de Revisão Bibliográfica	22
2.2 Taxonomia Adotada na Classificação das Publicações Revisadas	23
2.3 Discussão das Publicações Seleccionadas	26
2.4 Considerações Finais	46
CAPÍTULO 3 - FUNDAMENTAÇÃO TEÓRICA	47
3.1 Filtragem Espacial	47
3.2 Detecção de Bordas.....	51
3.3 Processamento de Imagens Coloridas.....	52
3.4 Componentes Conectados	56
3.5 Considerações Finais	57
CAPÍTULO 4 - ABORDAGEM PROPOSTA	58
4.1 Descrição da Abordagem.....	58
4.2 Suavização com Preservação de Bordas	59
4.3 Segmentação de Cores	60
4.4 Conexão de Componentes	64
4.5 Posterização.....	65
4.6 Ajuste de Saturação	65
4.7 Realce de Contornos.....	66
4.8 Limitações da Abordagem Proposta	68
4.9 Considerações Finais do Capítulo	70
CAPÍTULO 5 - VALIDAÇÃO DA ABORDAGEM PROPOSTA	71
5.1 Planejamento Experimental	71
5.2 Aspectos Relativos à Análise dos Resultados Experimentais	77
5.3 Experimentos Envolvendo a Abordagem Proposta.....	78
5.4 Delineamento do Perfil dos Participantes do Experimento	80
5.5 Resultados da Avaliação do Processo de Renderização de Vídeos	86
5.6 Resultados da Avaliação Comparativa entre Abordagens	89
5.7 Detalhes de Implementação.....	98
5.8 Considerações Finais do Capítulo	98
CAPÍTULO 6 - CONSIDERAÇÕES FINAIS	100
6.1 Conclusões e Contribuições	100
6.2 Proposições para Trabalhos Futuros.....	102
REFERÊNCIAS BIBLIOGRÁFICAS	104
APÊNDICE A	109
APÊNDICE B	112
APÊNDICE C	119

Capítulo 1

Introdução

No presente capítulo, contextualiza-se a área de pesquisa na qual esta dissertação de mestrado se insere (Seção 1.1). Também é comentada a motivação para a realização da pesquisa, assim como os objetivos alcançados, os quais estão listados nas Seções 1.2 e 1.3. Na Seção 1.4, são apresentadas as hipóteses a serem testadas. Por fim, a estrutura deste documento é apresentada na Seção 1.5.

1.1 Motivação/Contextualização da Pesquisa

A Renderização Não Fotorrealística (*NPR – Non photorealistic Rendering*) é uma área de pesquisa que reside na interseção das áreas da Computação Gráfica, do Processamento Digital de Imagens e da Arte. A partir de imagens e vídeos capturados do mundo real, busca-se gerar representações estilizadas e abstratas da realidade, com o propósito de comunicar visualmente ao observador uma ideia ou conceito. Exemplos de NPR variam desde a geração de cartuns, a partir de imagens e vídeos (e.g., KANG, 2015; WENHUA, 2015), até representações nas quais a imagem ou o vídeo é convertido em estilos artísticos, tais como aquarela e óleo sobre tela, dentre outros (e.g., ZHANG, 2013).

Quando aplicada a vídeos, a NPR pode desencadear alguns problemas na visualização do resultado, tendo em vista a forte correlação entre os quadros de um vídeo. Os maiores desafios nesse sentido são (i) problemas de cintilação e (ii) perda da coerência espaço-temporal,

conforme discutido por Chen, Turk e Macintyre (2012).

Há ferramentas comerciais, tais como o **AKVIS Sketch**¹, que se propõem a gerar representações não fotorrealísticas a partir de imagens fotorrealísticas. O processamento de imagens independentes entre si evita a necessidade de atenção a problemas de cintilação e possível perda da coerência espaço-temporal. Os resultados finais produzidos são satisfatórios, mas apenas para imagens independentes pois, ao se utilizar uma sequência de imagens (quadros) provenientes de um vídeo, com o intuito de exibí-los em sequência, ocorrem os problemas supracitados.

Outras ferramentas comerciais, tais como **Animationfrommovie**² e **MovieCartoonizer**³, destinam-se à renderização não fotorrealística de vídeos. Entretanto, é possível verificar que estas ferramentas produzem resultados insatisfatórios, tanto no tocante à renderização dos quadros quanto no que diz respeito à aparência de cartum de seu conteúdo. Os problemas discutidos por Chen, Turk e Macintyre (2012) são facilmente observáveis nos resultados finais destas ferramentas. Os problemas supramencionados são ilustrados nas Figuras 1 e 2.

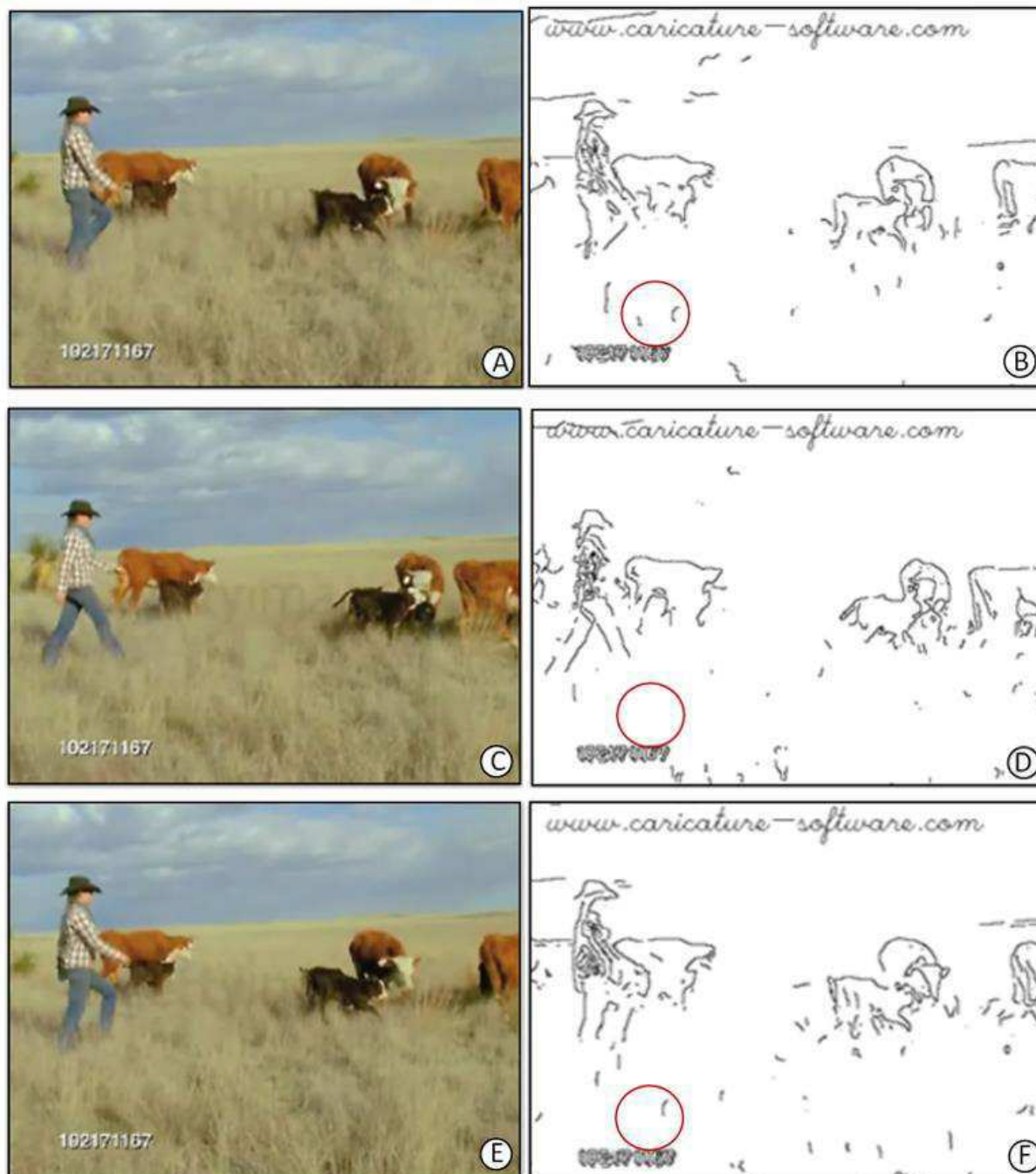
A Figura 1 contém uma sequência de 3 quadros gerados a partir da ferramenta **Animationfrommovie**. A área delimitada pelo círculo vermelho no primeiro quadro contém dois segmentos, os quais não estão presentes no segundo quadro, enquanto no terceiro quadro um novo segmento surge. Ao exibir os quadros em sequência, ocorrerá o efeito de cintilação, em que os segmentos contidos nas áreas delimitadas pelos círculos vermelhos aparecerão cintilando para o observador.

¹ <http://akvis.com/pt/sketch/index.php>

² <http://animation-from-movie.en.softonic.com/>

³ <http://movie-cartoonizer.software.informer.com/>

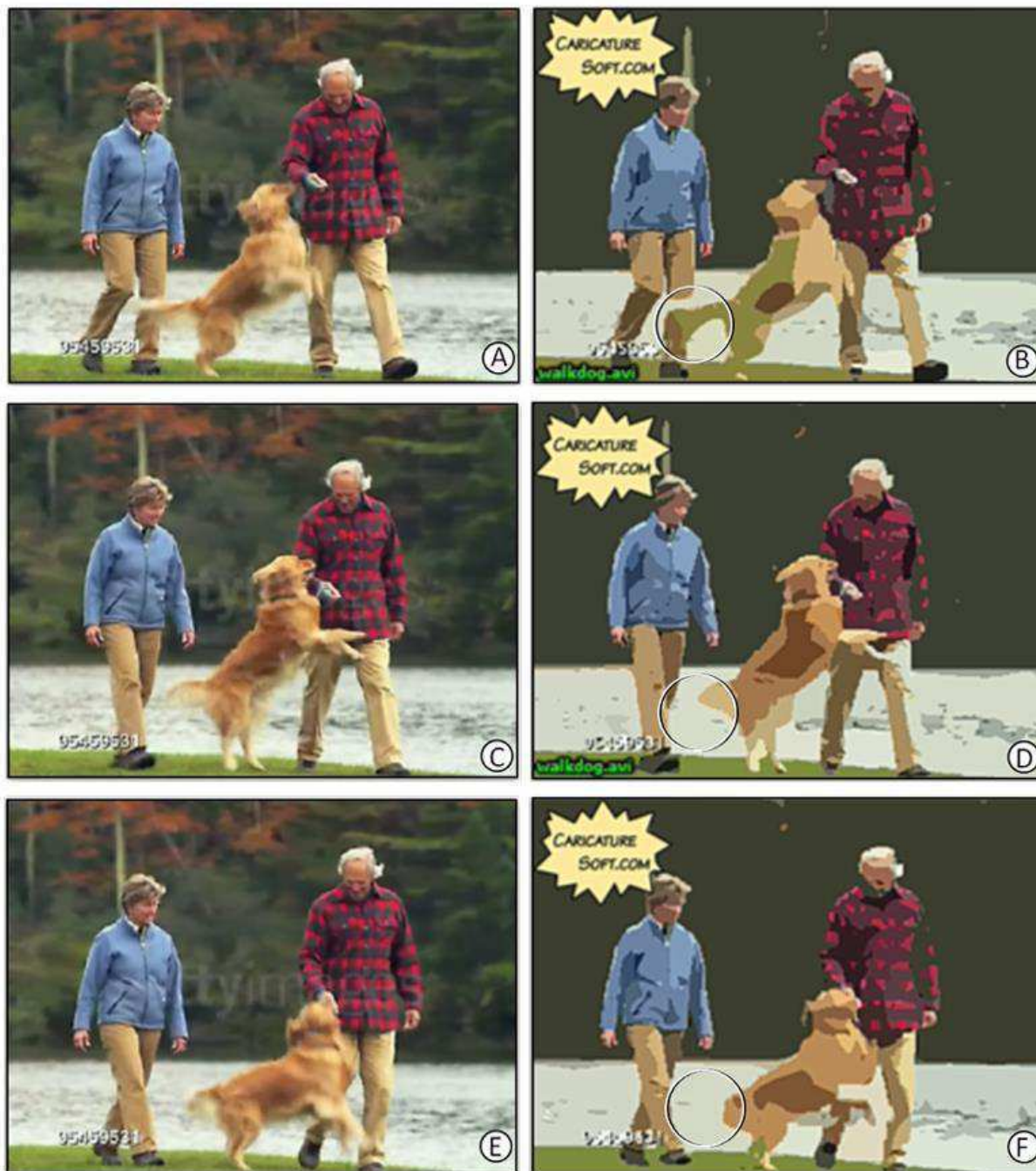
Figura 1 - Sequência de quadros ilustrando problemas de cintilação:
(A) Quadro 1 original; (B) Quadro 1 bordas marcadas; (C) Quadro 2 original; (D) Quadro 2 bordas marcadas; (E) Quadro 3 original e (F) Quadro 3 bordas marcadas.



Fonte: Autoria própria.

Na Figura 2, está contida uma sequência de 3 quadros, gerados a partir da ferramenta **MovieCartoonizer**.

Figura 2 - Sequência de quadros ilustrando problemas de coerência espaço-temporal: (A) Quadro 1 original; (B) Quadro 1 estilo cartum; (C) Quadro 2 original; (D) Quadro 2 estilo cartum; (E) Quadro 3 original e (F) Quadro 3 estilo cartum.



Fonte: Autoria própria.

No primeiro quadro, a área delimitada pelo círculo branco contém um objeto que é a cauda do cachorro. No segundo quadro, parte dessa área não está mais presente. Finalmente, no terceiro quadro, não há sinal algum dessa área. Exibindo-se os quadros em sequência, pode-se

observar a perda de coerência espaço-temporal entre eles. Observando-se individualmente os quadros 2 e 3, pode-se observar a perda de coerência espacial entre eles, pois a cauda do cachorro desapareceu. Vistos em sequência, pode-se observar a perda de coerência temporal entre os quadros, pois a cauda do cachorro, que deveria estar presente em todos os quadros, é segmentada diferentemente em cada quadro.

No campo da NPR, existem duas categorias de abordagens para o processamento de imagens e vídeos, a saber: (i) a abstração de cenas; e (ii) a segmentação de regiões de cenas. Na abstração de cenas, utilizam-se filtros para processar as regiões da imagem de forma homogênea (e.g., WINNEMÖLLER, KYPRIANIDIS e OLSEN, 2012; KYPRIANIDIS, KANG e DÖLLNER, 2009), razão pela qual esta categoria é pouco susceptível aos problemas supramencionados. Por outro lado, quando se adota a segmentação de regiões de cenas, a imagem é dividida em subregiões que serão processadas individualmente (e.g., KLETTE, 2014), o que torna esta categoria fortemente susceptível a problemas de cintilação e perda de coerência espaço-temporal.

Apesar de a abordagem fundamentada na abstração dos quadros de vídeo parecer mais vantajosa, a segmentação de quadros proporciona duas grandes vantagens, conforme enfatiza Bénard, Thollot e Collomosse (2012). A primeira é que os parâmetros de renderização podem ser aplicados de forma consistente por regiões do quadro renderizado. A segunda é que marcações de movimento podem ser fixadas em quadros de referência para cada região, de acordo com a movimentação delas ao longo do tempo, o que contribui com a tarefa de manutenção da coerência de movimento.

Em vista das considerações supracitadas, a presente pesquisa foi desenvolvida com ênfase no emprego de segmentação aplicada à renderização de vídeos não fotorrealísticos. Para tanto, foi necessária a condução de um levantamento do estado da arte, cujos resultados se

encontram no Capítulo 2 (Revisão Bibliográfica).

As próximas seções contêm uma formulação dos objetivos da pesquisa.

1.2 Objetivo Geral

Esta pesquisa tem como objetivo geral a proposição e a validação de uma abordagem para a geração de vídeos ao estilo de cartuns, fundamentada no uso de segmentação.

1.3 Objetivos Específicos

Para a consecução do objetivo geral desta pesquisa, foram formulados os seguintes objetivos específicos:

1. Desenvolvimento de um protótipo de software para a geração de vídeos ao estilo de cartuns, a fim de validar os resultados da abordagem proposta; e
2. Desenvolvimento de uma validação experimental quantitativa e qualitativa, visando à comparação de resultados obtidos a partir da abordagem proposta e resultados produzidos por abordagens do estado da arte que disponibilizaram código executável ou resultados passíveis de comparação.

1.4 Hipóteses

No intuito de atingir os objetivos apresentados nesta dissertação, foram formuladas as seguintes hipóteses nulas:

H_{01} : A abordagem proposta não gera resultados que são mais bem avaliados por um conjunto de participantes, em relação às abordagens concorrentes;

H_{02} : Os resultados gerados pela abordagem proposta não contêm artefatos espúrios ou cintilação, quando comparados àqueles produzidos por abordagens semelhantes;

H0₃: A abordagem proposta não produz resultados que apresentam melhor definição de bordas e de estilização, quando comparada a abordagens semelhantes.

1.5 Estrutura do Documento

O conteúdo desta dissertação foi estruturado como segue.

No Capítulo 2 (**Revisão Bibliográfica**), está contida a revisão bibliográfica do estado da arte na área. Tal revisão permite evidenciar os problemas decorrentes no processo de geração de vídeos ao estilo de cartuns, as estratégias adotadas para contornar/minimizar tais problemas e as pesquisas que serviram de base para o desenvolvimento da abordagem proposta no Capítulo 4 (**Abordagem Proposta**).

O Capítulo 3 (**Fundamentação Teórica**) contém a fundamentação teórica da pesquisa, além de conceitos utilizados no desenvolvimento da abordagem proposta.

Os módulos presentes na abordagem, assim como as técnicas utilizadas para instanciar cada um dos módulos, são apresentados e discutidos no Capítulo 4.

No Capítulo 5 (**Validação da Abordagem Proposta**), é apresentado o planejamento experimental completo, contemplando desde a preparação do material de teste até condução das sessões de teste. A validação da abordagem proposta nesta dissertação também é objeto do Capítulo 5, a qual foi desenvolvida por meio de um experimento com participantes humanos. Os resultados do experimento são apresentados e discutidos, incluindo variações de parâmetros e técnicas da abordagem proposta.

Por fim, no Capítulo 6 (**Considerações Finais**), são tecidas as considerações finais desta pesquisa. As conclusões advindas dos resultados apresentados e discutidos no Capítulo 5 (**Validação da Abordagem Proposta**) são formuladas e as contribuições, destacadas. Por fim, são propostas investigações futuras para dar continuidade à linha de pesquisa abordada nesta dissertação.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, é feita uma revisão bibliográfica sistemática de pesquisas do estado da arte na geração de vídeos ao estilo de cartuns. Nas próximas seções, são apresentadas informações referentes à descrição da condução da metodologia utilizada, o processo de revisão sistemática propriamente dito e as publicações classificadas conforme uma taxonomia proposta, além de conclusões acerca do resultado.

2.1 Descrição do Processo de Revisão Bibliográfica

A revisão bibliografia ocorreu com o auxílio da ferramenta de revisão sistemática StArt (ZAMBONI *et al.*, 2010), a qual permite, de forma semiautomática, a catalogação das publicações, a adição de critérios de aceitação e de exclusão de publicações que permitem atribuir uma pontuação às publicações, objetivando representar a possível relevância daquela publicação para a pesquisa.

Após o cálculo da pontuação, é possível filtrar manualmente a lista de publicações, a partir dos títulos e dos resumos, no intuito de decidir sobre a importância de cada publicação para a pesquisa.

As bases de publicações selecionadas para a pesquisa foram ACM⁴, IEEE⁵ e Springer⁶, em função da quantidade de publicações relativas ao processamento de imagens contidas nestas bases, assim como por serem bases em que os autores mais expressivos da área costumam publicar suas pesquisas.

A próxima etapa consistiu da formulação e uso, em cada uma das bases, da chave de pesquisa (**"npr" OR "non photorealistic rendering" OR "non-photorealistic rendering" OR "artistic rendering" OR**

⁴ <https://dl.acm.org>

⁵ <https://ieeexplore.ieee.org/Xplore/home.jsp>

⁶ <https://link.springer.com>

"stylization" OR "stylized rendering" OR "cartoonization") AND ("video" OR "image" OR "scene") AND ((("temporal" OR "spatial" OR "temporal and spatial" OR "spatial and temporal") AND ("coherence" OR "consistency")) OR "flickering"). Algumas alterações sintáticas foram consideradas, a fim de respeitar peculiaridades existentes nos campos de pesquisa de cada base.

Os resultados obtidos a partir da chave de pesquisa em cada base foram inseridos no aplicativo StArt, tendo totalizado 1.341 publicações recuperadas. A próxima etapa foi a leitura dos títulos e *abstracts* de cada uma das 1.341 publicações, adotando-se um dos seguintes critérios: (i) a aceitação da publicação, caso tratasse de segmentação, coerência espaço-temporal e/ou cintilação em quadros de vídeo, aplicados à NPR; ou (ii) a rejeição da publicação, caso nenhum dos tópicos anteriores fosse ali tratado.

Ao final da primeira filtragem manual, restaram 83 publicações que, caso apresentassem ideias que englobassem algum dos tópicos supracitados, seriam aceitas como publicações relacionadas à pesquisa e, em caso contrário, seriam rejeitadas. Das 83 publicações, vale ressaltar que 8 delas não puderam ser obtidas *online* ou *offline* nem via os respectivos autores.

Após a finalização desse processo, foram classificadas 21 publicações, as quais serviram de fundamentação para a pesquisa.

2.2 Taxonomia Adotada na Classificação das Publicações Revisadas

A taxonomia descrita a seguir fundamentou-se no levantamento do estado da arte conduzido por Amarnadh, Hussai e Raju (2014).

As publicações foram divididas inicialmente em dois grandes grupos, a saber: (i) abordagens que utilizam informação espacial durante o processo (e.g., abordagens que detectam formas e objetos em um quadro de vídeo e o estilizam, como fizeram Ben e Zelnik (2015)); e (ii) abordagens que utilizam informação temporal durante o processo (e.g., para a detecção da posição de um dado objeto ao longo de uma sequência

de quadros de um vídeo, como fez Zhang (2013)).

Após a classificação inicial das publicações obtidas da pesquisa em um dos grupos supracitados, foi feita uma nova classificação, visando à inseri-la em uma das seguintes categorias:

I Método fundamentado em pixels (*Pixel-based method*)

Utilização de comparações par a par entre pixels, levando-se em conta as características de cada um deles, e.g., cor, diferença de intensidade e/ou textura.

II Método fundamentado em blocos (*Patch-based method*)

Utilização de agrupamentos de regiões de pixels ou de áreas presentes em um quadro que possuam intensidades parecidas, de modo a compor um bloco homogêneo que pode ser utilizado, por exemplo, na comparação com outros blocos ou na detecção de objetos.

III Método fundamentado em movimento (*Motion-based method*)

Utilização de dados de movimento, usualmente vetores de movimento, fluxo óptico ou frequência de correlação entre bordas de regiões.

IV Método fundamentado em características (*Feature-based method*)

Utilização de características tais como transições de cores, invariantes de movimento, modelos escondidos de Markov, bordas, aplicações bayesianas, aplicações estatísticas, etc.

V Método fundamentado em histogramas (*Histogram-based method*)

Utilização da diferença, intersecção, normalização e também similaridade entre histogramas.

VI Combinação de métodos (*Combination of methods*)

Assim como o nome desta seção sugere, as publicações inseridas nesta categoria fundamentam-se na combinação de dois ou mais dos métodos previamente citados.

VII Outros (*Peculiarity-based strategy*)

Utilização de uma abordagem que não pertence a nenhuma das categorias supracitadas.

As 21 publicações classificadas como relacionadas à pesquisa desenvolvida estão contidas nos dois quadros. O Quadro 1 contém as publicações obtidas via revisão sistemática que utilizam informação espacial e a categoria a que pertencem.

Quadro 1 - Publicações que utilizam informação espacial.

Publicação	Categoria	Descrição
Ben e Zelnik (2015)	II	Rastreamento de objetos por correspondência de blocos de pixels
Wang <i>et al.</i> (2013)	VII	Extração de linhas utilizando filtro de diferença de gaussianas
Su, Luo e Artusi (2013)	VII	Decomposição de textura preservando as bordas
Rosin e Lai (2013)	VII	Redução de cores por classes
Lu <i>et al.</i> (2013)	VII	Filtragem bilateral
Xue <i>et al.</i> (2014)	VII	Deteção de segmentos com um detector com controle de falsa deteção
Chen, Turk e Macintyre (2012)	VII	Utilização de pontos âncora
Song e Xu (2012)	VII	Mapa de saliência
Gangopadhyay, Pachori e Raman (2016)	VII	Segmentação <i>top-down</i>
Kolivand <i>et al.</i> (2018)	VII	<i>Survey</i>

Fonte: Autoria própria.

No Quadro 2, apresentam-se as publicações advindas da revisão sistemática que utilizam informações de caráter temporal, assim como a categoria a que pertencem. Ambos os quadros fundamentam-se na taxonomia definida nesta seção.

Quadro 2 - Publicações que utilizam informação temporal.

Publicação	Categoria	Descrição
Zhang <i>et al.</i> (2013)	I	Transferência de estilo via correspondência de pixels em quadros adjacentes
Lin, Lin e Wang (2012)	III	Estilização via correspondência de camadas de movimento em quadros adjacentes
Cao <i>et al.</i> (2011)	III	Estilização de vídeos guiada por movimento
Chen, Turk e MacIntyre (2011)	IV	Estilização guiada por rastreamento de características
Lu <i>et al.</i> (2015)	VI	Estilização guiada por correspondência de pixels e histórico de movimento
Kang <i>et al.</i> (2015)	VI	Transferência de estilo pixel a pixel guiada por fluxo óptico
Bénard, Thollot e Collomosse (2012)	VI	Discute as várias técnicas comumente utilizadas na NPR
Yoon, Lee e Kang (2012)	VII	Operador de Sobel 3D
Bénard <i>et al.</i> (2012)	VII	Grupos independentes de curvas
Klette (2014)	VII	Segmentação por crescimento de semente
Jiang <i>et al.</i> (2015)	VII	Estrutura de movimento

Fonte: Autoria própria.

2.3 Discussão das Publicações Selecionadas

Nesta seção, são discutidas as publicações relacionadas com a pesquisa, as quais foram apresentadas nos Quadros 01 e 02, da Seção 2.2 (**Taxonomia Adotada na Classificação das Publicações Revisadas**).

Uma consideração importante quando se adota a segmentação de vídeos é a existência de forte coerência temporal entre quadros adjacentes, conforme ressaltado por Ben e Zelnik (2015). A comparação entre blocos (*patches*) extraídos do quadro atual e grupos de blocos previamente processados é utilizada pelos autores para a detecção e o rastreamento de objetos presentes em vídeos. Os problemas enfrentados pelos autores ocorrem quando o vídeo é composto por múltiplas cenas ou possui alguma transição abrupta, porque a coerência que deveria existir

entre os quadros adjacentes é perdida. Um exemplo de resultado da abordagem de Ben e Zelnik (2015) pode ser visto na Figura 3.

**Figura 3 - Exemplo de resultado da abordagem de Ben e Zelnik:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Ben e Zelnik (2015).

A pesquisa realizada por Wang *et al.* (2013) tratou da utilização do filtro de diferença de gaussianas para a extração de linhas. Apesar de ser uma técnica relacionada à abstração, esta utilizou uma ideia de realizar a detecção de regiões com cores diferentes. Tal ideia, permite a distinção de regiões próximas, divergentes apenas na cor, que seriam segmentadas como uma única região ao se ignorar as diferenças existentes entre as cores. Um exemplo da extração de linhas de Wang *et al.* (2013) pode ser visto na Figura 4.

**Figura 4 - Exemplo do método de extração de linhas de Wang *et al.*:
(A) Quadro original; (B) Quadro com extração de linhas.**

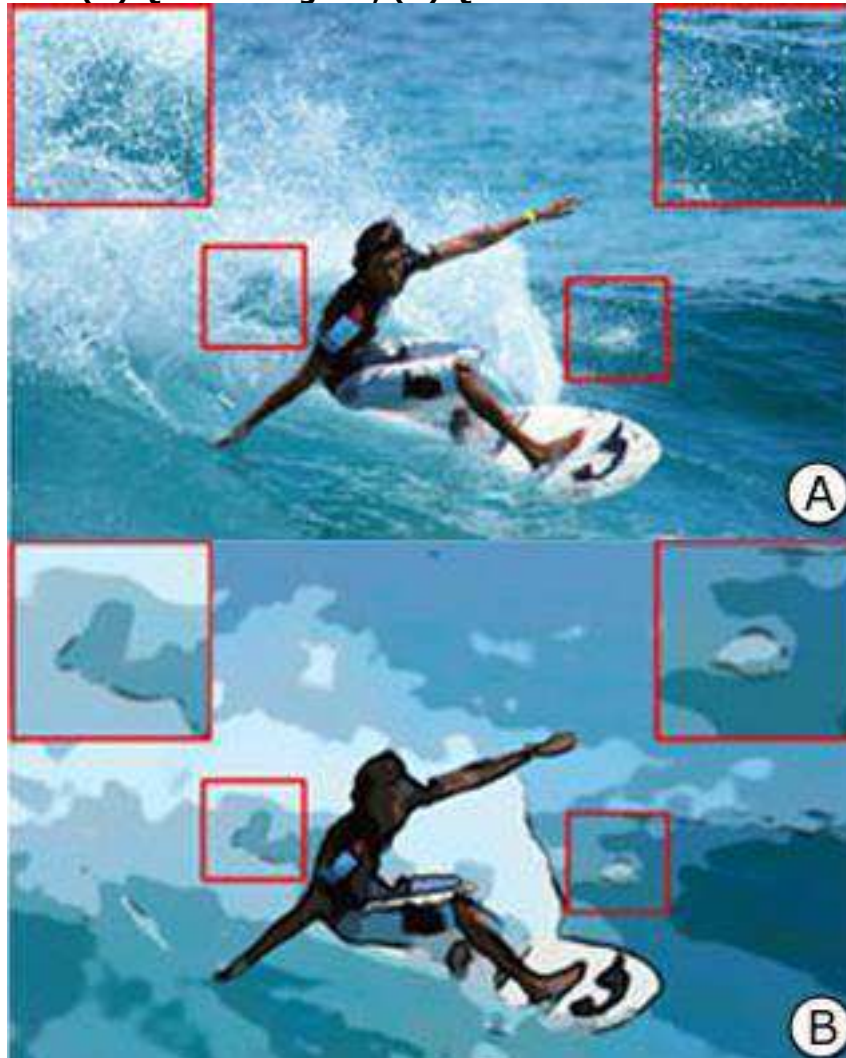


Fonte: Wang *et al.* (2013).

Outra pesquisa envolvendo abstração de imagens foi conduzida por Su, Luo e Artusi (2013), na qual os autores realizaram a decomposição da imagem em textura e estrutura, por meio do uso de um filtro não linear LTV (*Local Total Variation*). O uso do filtro LTV em uma imagem pode melhorar a extração de bordas, se combinado à segmentação, além de poder ser utilizado como um auxílio em regiões em que o processo de segmentação tende a falhar. O maior problema é que o uso do LTV em imagens que possuem alto grau de textura gera resíduos de textura próximos às bordas, o que demanda o cuidado redobrado com o processamento, de modo que esses resíduos não venham a ocasionar cintilação e perda de coerência temporal. Um exemplo de resultado da

abordagem de Su, Luo e Artusi (2013) pode ser visto na Figura 5.

**Figura 5 - Exemplo de resultado da abordagem de Su, Luo e Artusi:
(A) Quadro original; (B) Quadro estilizado.**

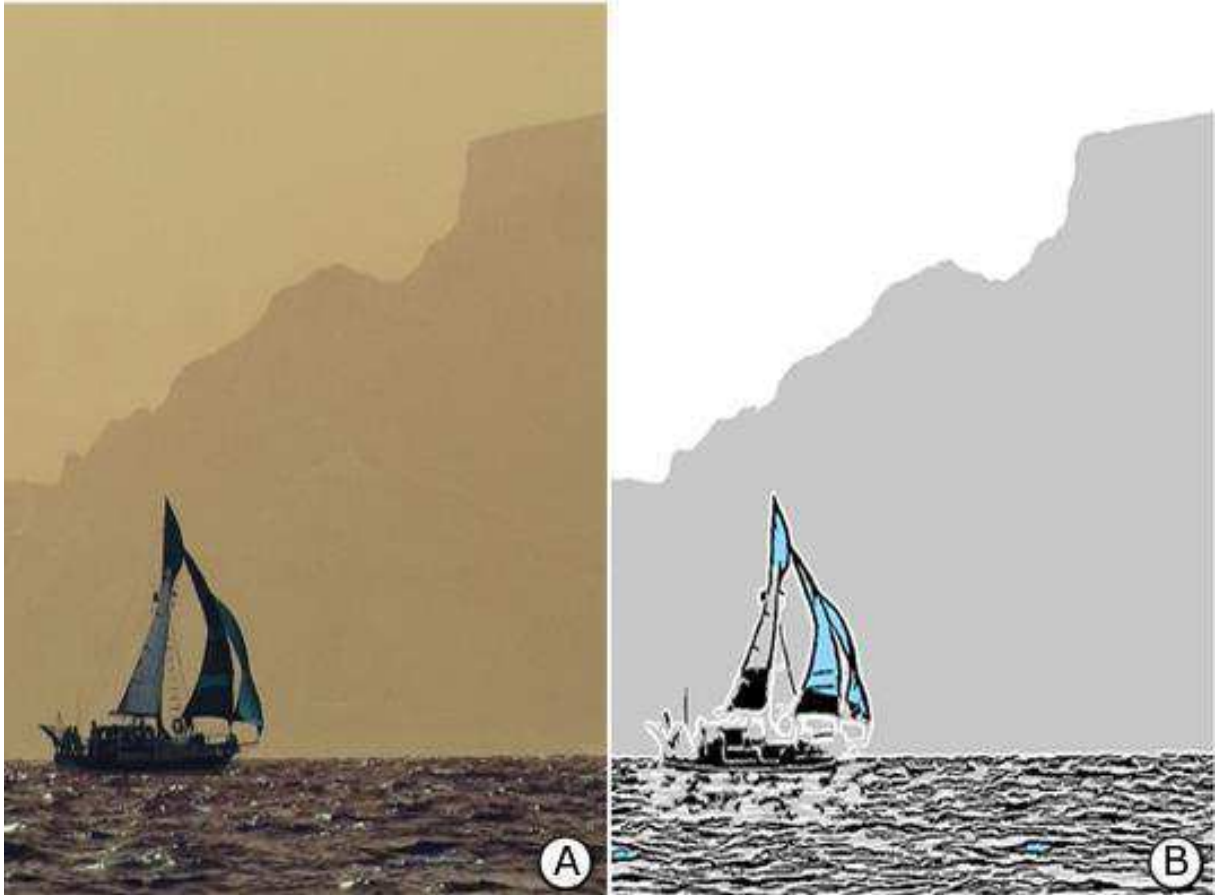


Fonte: Su, Luo e Artusi (2013).

Outra abordagem adotada em processos de segmentação de quadros de vídeo é classificar os pixels em branco, preto ou cinza, segundo limiares pré-estabelecidos e selecionar, em seguida, qual das três regiões é a mais apropriada para receber a colorização, conforme proposto por Rosin e Lai (2013). Tal abordagem permite distinção entre o primeiro plano e o plano de fundo do quadro, ao calcular a média da magnitude das bordas, a partir de um detector de Sobel. Este processo contribui para a manutenção da coerência temporal, por facilitar a identificação dos objetos que devem ser rastreados no decorrer dos quadros de um vídeo.

Apesar de promissora, a abordagem é dependente de limiares. A Figura 6 contém um exemplo da abordagem de Rosin e Lai (2013).

Figura 6 - Exemplo da abordagem de Rosin e Lai: (A) Quadro original; (B) Quadro estilizado.



Fonte: Rosin e Lai (2013).

Lu *et al.* (2013) propuseram a geração de representações finais em diversos estilos para imagens estáticas. Partindo de uma imagem de entrada, os autores geraram uma abstração e uma representação quantizada. Ao final do procedimento, os dois artefatos foram combinados em uma única imagem de saída. Ao se substituir a abstração por segmentação e a representação quantizada por um grupo de possíveis estilos que podem ser selecionados facilmente, pode-se conceber um arcabouço que gere múltiplos estilos de saída. Um exemplo de resultado da abordagem de Lu *et al.* (2013) é apresentado na Figura 7.

Figura 7 - Exemplo de resultado da abordagem de Lu et al. 2013: (A)Quadro original; (B) Quadro estilizado.



Fonte: Lu et al. (2013).

A extração de segmentos em uma imagem pode ser melhorada a partir da inclusão de um detector de segmentos com um controle de falsa detecção *LSD* (*Line Segment Detector*), conforme proposto por Xue et al. (2014). O *LSD* é combinado com o filtro de diferença de gaussianas para gerar uma imagem final com bordas mais nítidas. Ao substituir o filtro por segmentação ou utilizar o *LSD* como um passo da segmentação, é possível gerar um resultado final com segmentos mais definidos. O problema é que a adição do *LSD* pode causar super segmentação, prejudicando o processamento como um todo e propiciando o surgimento de cintilação. Um exemplo da extração de segmentos por Xue et al. (2014) pode ser visto na Figura 8.

**Figura 8 - Exemplo da extração de segmentos de Xue *et al.* (2014):
(A) Quadro original; (B) Quadro com extração de linhas.**



Fonte: Xue *et al.* (2014).

Independentemente do estilo utilizado em uma representação não fotorrealística, é desejável que o resultado tenha o aspecto mais próximo possível do que seria produzido por um artista real e, para tanto, as pinceladas são um ponto-chave. O uso de vetores para representar linhas que serão traçadas por um pincel é algo comum, mas ao se utilizar tensores, conforme demonstrado em Chen, Turk e MacIntyre (2012), é possível traçar linhas com mais de uma orientação, aumentando-se o realismo do estilo e tornando o resultado mais agradável para um observador. Um exemplo de resultado da abordagem proposta por Chen, Turk e MacIntyre (2012) pode ser visto na Figura 9.

Figura 9 - Exemplo de resultado da abordagem de Chen, Turk e MacIntyre de 2012.



Fonte: Chen, Turk e MacIntyre (2012).

Uma abordagem diferente das demais supramencionadas utiliza o filtro Kuwahara. Proposta por Song e Xu (2012), tal abordagem considera que existe alta probabilidade de o objeto de interesse estar próximo ao centro da imagem permitindo, assim, a utilização de diferentes tratamentos para o primeiro plano e o plano de fundo. Os problemas que surgem com a adoção desta abordagem em vídeos é que o objeto de interesse pode se encontrar em regiões diferentes do centro ou em que os objetos podem se mover para locais diferentes do centro ou transitar por ele. Desta maneira, os tratamentos serão aplicados de maneira errônea, resultando em vídeos incoerentes com o esperado. Um exemplo de resultado da abordagem de Song e Xu (2012) pode ser visto na Figura 10.

**Figura 10 - Exemplo de resultado da abordagem de Song e Xu:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Song e Xu (2012).

A transferência de estilo é algo empregado muitas vezes na NPR, conforme pode ser visto em Zhang *et al.* (2013). Os autores propuseram a abstração de um vídeo em um dado estilo ao transferi-lo de uma imagem alvo. A transferência de estilo ocorre no primeiro quadro e, por correspondência pixel-a-pixel, é feita a propagação do estilo para os quadros adjacentes. O maior problema que ocorre é a falta de correspondência entre os pixels, ao longo da sequência de quadros considerada, causando assim a perda da coerência temporal ao longo do vídeo. Um exemplo de resultado produzido pela abordagem de Zhang *et al.* (2013) pode ser visto na Figura 11.

**Figura 11 - Exemplo de resultado da abordagem de Zhang *et al.*:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Zhang *et al.* (2013).

Em NPR, existem várias pesquisas que utilizam a detecção de movimento, conforme empregada por Lin, Lin e Wang (2012), em que o vídeo de entrada é particionado em múltiplas camadas de movimento, por meio da correspondência entre pares de quadros. O próximo passo é a correspondência das camadas de movimento, utilizando uma função de similaridade. A partir deste ponto, qualquer estilo pode ser aplicado e propagado por todo o vídeo. Um exemplo de resultado da abordagem de Lin, Lin e Wang (2012) pode ser visto na Figura 12.

**Figura 12 - Exemplo de resultado da abordagem de Lin, Lin e Wang:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Lin, Lin e Wang (2012).

A transferência de estilo entre imagens é utilizada por Cao *et al.* (2011). Grupos de pixels são transferidos de uma imagem de referência para uma imagem de entrada, até que o estilo de ambas se torne semelhante. Em vídeos, o processo é aplicado ao primeiro quadro e o resultado é propagado para os outros quadros via fluxo de movimento. A coerência temporal é preservada por meio de um algoritmo de fluxo óptico para encontrar a correspondência entre quadros subsequentes. O que há de correspondente entre um quadro e seu sucessor é copiado para o quadro sucessor, na mesma posição na qual se encontra no predecessor. Um exemplo de resultado da abordagem de Cao *et al.* (2011) pode ser visto na Figura 13.

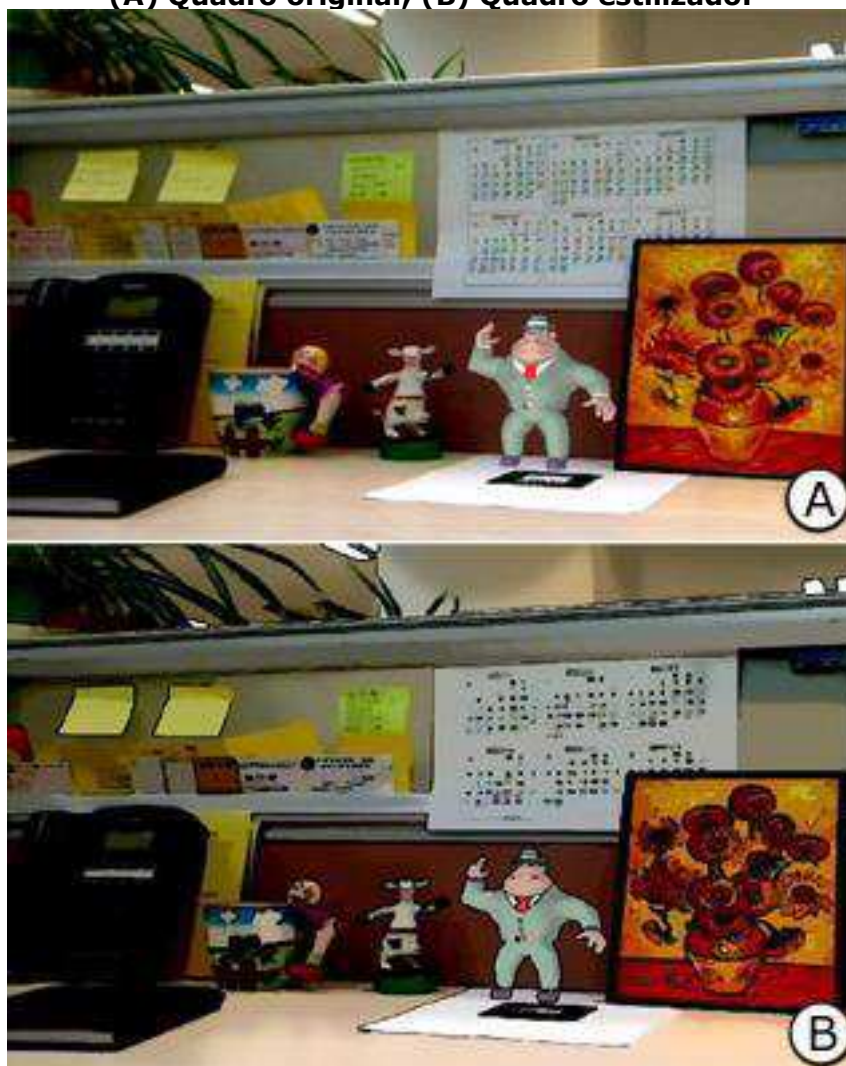
**Figura 13 - Exemplo de resultado da abordagem de Cao *et al.*:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Cao *et al.* (2011).

Lu *et al.* (2015) fizeram o uso de filtragem temporal para manter a coerência espacial. Primeiramente, os autores utilizaram um *buffer* para armazenar quadros consecutivos pertencentes à mesma cena. O próximo passo foi calcular a correspondência entre os pixels de quadros vizinhos, por meio de um histórico de movimento. Por fim, os autores aplicaram um filtro gaussiano unidimensional para realizar a extração das bordas. Um vídeo composto de múltiplas cenas, incorretamente divididas, acarretará transições abruptas entre quadros adjacentes comprometendo, assim, todas as etapas seguintes do processo. Um exemplo de resultado da abordagem proposta por Lu *et al.* (2015) pode ser visto na Figura 14.

**Figura 14 - Exemplo de resultado da abordagem de Lu et al. 2015:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Lu et al. (2015).

A transferência de textura de uma imagem para um vídeo foi realizada por Kang et al. (2015). Pixels aleatórios são selecionados de uma imagem de referência e transferidos para o primeiro quadro do vídeo de interesse. O passo seguinte é a transferência de pixels do quadro atual para o seguinte, levando-se em consideração os pixels que melhor representam a imagem de referência. A transferência de pixels é feita utilizando o fluxo óptico. Os próprios autores levantaram como ponto fraco de sua abordagem a falta de precisão na estimação do movimento, que pode causar o surgimento de padrões de textura incorretos. Um exemplo de resultado da abordagem de Kang et al. (2015) pode ser visto na Figura 15.

**Figura 15 - Exemplo de resultado da abordagem de Kang et al.:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Kang et al. (2015).

Vetores gradiente tridimensionais foram empregados por Yoon, Lee e Kang (2012), os quais foram gerados a partir de um operador de Sobel tridimensional com uma máscara cúbica. Tal abordagem permite a representação da vizinhança de pixels de maneira espacial e temporal, simultaneamente. O maior problema relacionado ao operador de Sobel é sua instabilidade na presença de ruídos. Um exemplo de resultado da abordagem de Yoon, Lee e Kang (2012) pode ser visto na Figura 16.

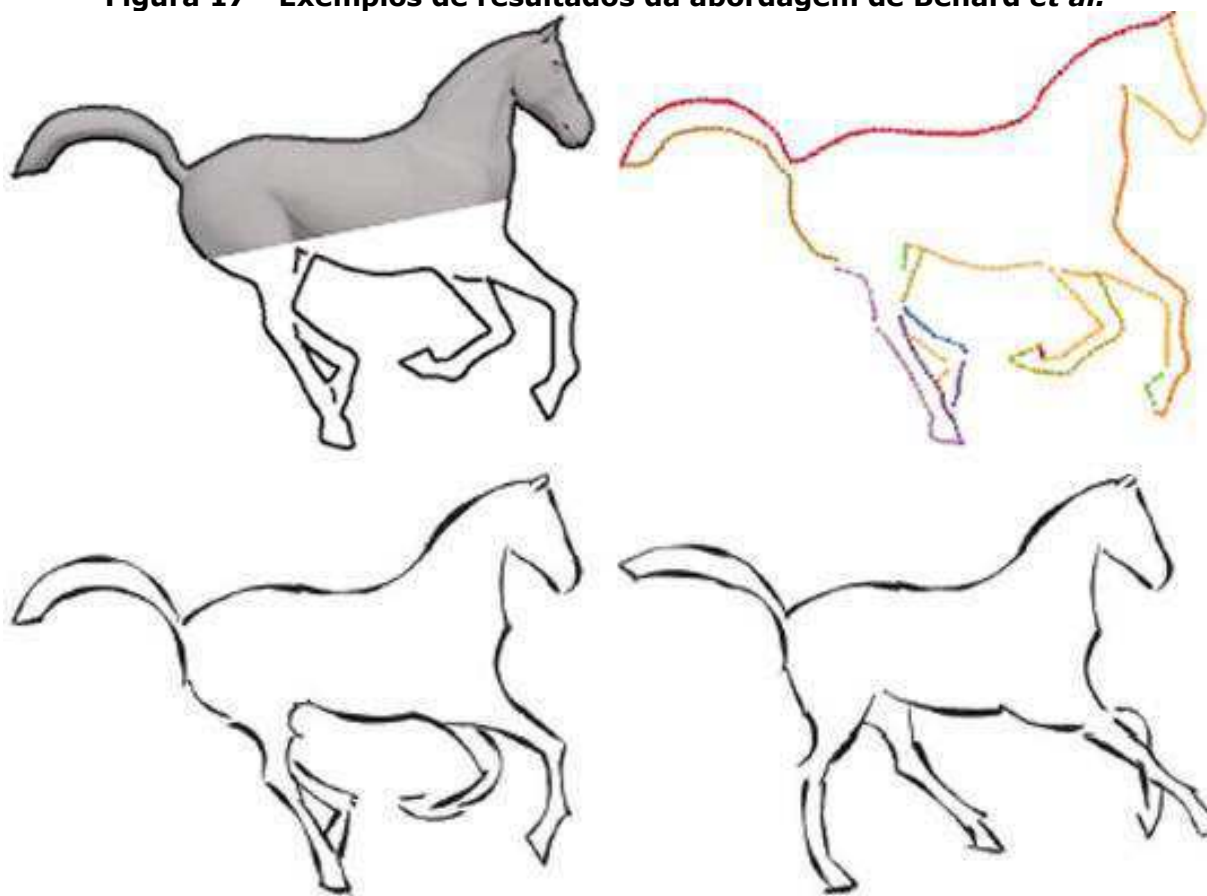
**Figura 16 - Exemplo de resultado da abordagem de Yoon, Lee e Kang:
(A) Quadro original; (B) Quadro estilizado.**



Fonte: Yoon, Lee e Kang (2012).

Linhas ativas (*active strokes*), utilizadas por Bénard *et al.* (2012), são uma forma de rastrear e organizar linhas, obtidas por meio de um processo de detecção de bordas que pode prover maior coerência temporal durante o processo de estilização de imagens. As linhas ativas utilizam contornos ativos (*snakes*), com o intuito de monitorar e vetorizar as informações de entrada. Linhas ativas também contêm informações de grupos de trilhas de pinceis para propagar as alterações de forma coerente, além dos detalhes das linhas. Os autores afirmaram que uma limitação de sua abordagem é que decisões sobre os contornos ativos ou caminhos de pinceis são feitas apenas localmente. Exemplos de resultados da abordagem proposta por de Bénard *et al.* (2012) podem ser vistos na Figura 17.

Figura 17 - Exemplos de resultados da abordagem de Bénard *et al.*



Fonte: Bénard *et al.* (2012).

A técnica de segmentação por crescimento de semente (*seed growing*), empregada por Klette (2014), tem como base a seleção de um pixel semente. O próximo passo é a integração de pixels adjacentes ao pixel semente, de forma recursiva, obedecendo a um critério de similaridade pré-estabelecido. O processo é repetido até que todos os pixels da imagem estejam contidos em algum dos segmentos gerados. A maior dificuldade associada à abordagem de Klette (2014) é a definição de um critério de similaridade representativo. Um critério pouco representativo fará com que pixels incompatíveis sejam colocados em um mesmo segmento, enquanto pixels compatíveis serão postos em segmentos diferentes. Um exemplo de resultado da abordagem proposta por Klette (2014) pode ser visto na Figura 18.

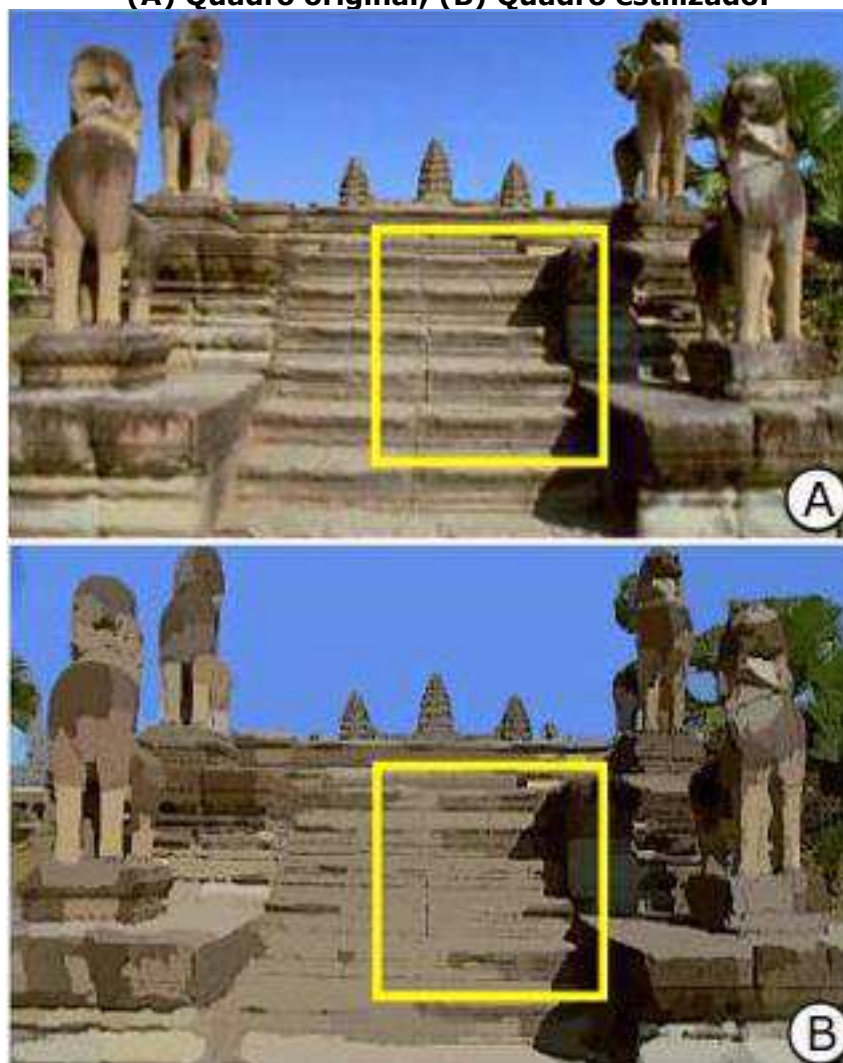
Figura 18 - Exemplo de resultado da abordagem de Klette: (A) Quadro original; (B) Quadro estilizado.



Fonte: Klette (2014).

Uma maneira diferenciada de realizar a segmentação foi empregada por Jiang *et al.* (2015) consistindo, inicialmente, na construção de um mapa de segmentos para cada quadro pertencente ao vídeo de entrada. Os segmentos presentes nos mapas são projetados para os quadros adjacentes e, então, é feita a otimização dos segmentos, por meio de estatísticas coletadas do vídeo, que representam as probabilidades de possíveis bordas. Um exemplo de resultado da abordagem de Jiang *et al.* (2015) pode ser visto na Figura 19.

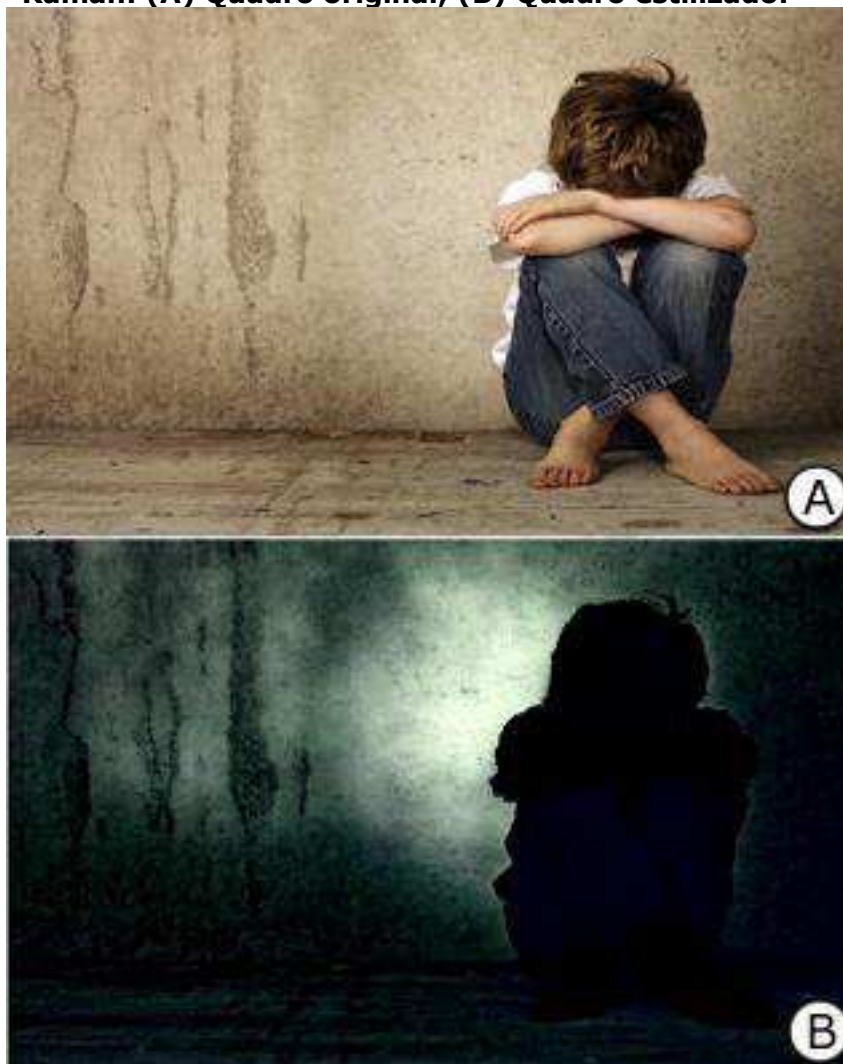
Figura 19 - Exemplo de resultado da abordagem de Jiang et al.:
(A) Quadro original; (B) Quadro estilizado.



Fonte: Jiang et al. (2015).

Gangopadhyay, Pachori e Raman (2016) utilizaram em suas pesquisas um mapa de saliências para realizar a detecção dos objetos presentes no primeiro plano, e assim, poder distingui-los do plano de fundo. O mapa de saliências assim como o *spot color*, proposto por Rosin e Lai (2013) produzem resultados visualmente semelhantes. Um exemplo de resultado da abordagem de Gangopadhyay, Pachori e Raman (2016) pode ser visto na Figura 20.

Figura 20 - Exemplo de resultado da abordagem de Gangopadhyay, Pachori e Raman: (A) Quadro original; (B) Quadro estilizado.



Fonte: Gangopadhyay, Pachori e Raman (2016).

Os resultados apresentados na literatura revisada são, em geral, promissores. Como exemplos, podem ser citados Xue *et al.* (2014), em que são observados traços mais nítidos e Lin, Lin e Wang (2012), em que a detecção de movimento ocorre de forma precisa. Contudo, os problemas de cintilação e perda de coerência espaço-temporal continuam não totalmente solucionados no campo da segmentação, conforme se pode observar em Chen, Turk e Macintyre (2012).

A presente revisão sistemática serviu de base para a condução dos experimentos envolvendo conceitos, técnicas e abordagens contidas no estado da arte revisado, além de possibilitar combinações adequadas das técnicas adotadas pelos autores revisados.

O uso do filtro LTV por Su, Luo e Artusi (2013) foi uma abordagem não selecionada para utilização nesta pesquisa porque são gerados resíduos de textura próximos às bordas de algumas imagens. Os resíduos geram cintilação, prejudicando o resultado final.

Rosin e Lai (2013) conseguiram realizar uma discriminação satisfatória entre primeiro plano e plano de fundo da imagem. Entretanto, o uso de limiares tornou-se um obstáculo para a automatização do método e a obtenção de resultados satisfatórios com vídeos de diferentes categorias.

A variedade de estilos gerados pôde ser utilizada conforme escrita em Lu *et al.* (2013). A segmentação é realizada normalmente e os passos seguintes diferem apenas no estilo de saída escolhido, o que possibilita a criação de um arcabouço de estilos baseado em segmentação.

Outra abordagem não utilizada foi aquela proposta por Song e Xu (2012), segundo a qual o objeto de interesse se encontra próximo ao centro da imagem. Tal abordagem não consegue obter resultados satisfatórios em vídeos que não obedeçam esta restrição.

A transferência de estilos também não foi o foco das investigações desta dissertação, tendo em vista a necessidade de criação de um arcabouço contendo múltiplos estilos, conforme supramencionado, o que estava fora do escopo desta pesquisa.

Vetores gradientes tridimensionais empregados por Yoon, Lee e Kang (2012) permitem a representação espacial e temporal da vizinhança de pixels, mas sua instabilidade em presença de ruído dificultou sua utilização na geração de uma abordagem robusta.

Active strokes e *seed growing* são duas técnicas que tampouco foram utilizadas nesta pesquisa. A primeira apresenta soluções locais, o que prejudica a coerência temporal ao longo dos quadros. A segunda necessita de limiares e medidas de similaridade, o que prejudica a variedade de vídeos que podem ser tratados de maneira satisfatória.

Ao final da revisão bibliográfica, constatou-se uma escassez de pesquisas na área de renderização não fotorrealística fundamentada na segmentação de imagens. Tal escassez é facilmente observável nos Quadros 01 e 02, os quais contêm pouquíssimas pesquisas recentes (GANGOPADHYAY, PACHORI, RAMAN, 2016; JIANG *et al.*, 2015) fundamentadas em segmentação. Por outro lado, a abstração de imagens se afigura mais comum e pode ser observada em diversas pesquisas da última década, mesmo mais recentes como aquelas realizadas por Wang *et al.* (2017) e Hassan *et al.* (2017).

Apesar de suas limitações, a presente pesquisa, afigura-se uma contribuição para a área de NPR, no sentido de explorar a vertente da segmentação de imagens. Adicionalmente, a investigação da segmentação por quantização de cores também pode ser vista como outra contribuição, tendo em vista não terem sido encontradas na literatura da área pesquisas que adotem essa estratégia.

2.4 Considerações Finais

Esta revisão bibliográfica sistemática fundamentou-se na taxonomia proposta por Amarnadh, Hussai e Raju (2014). A taxonomia adotada neste documento inclui as categorias Combinação de Métodos e Outros.

As publicações classificadas utilizam uma diversidade de técnicas. Algumas fazem filtragem espacial. Em sua maioria, é utilizada uma técnica que pode ser classificada como Outros (ver Seção 2.2) para detectar os objetos e realizar todo o processamento. Outras publicações utilizam informações temporais e fazem uso mais frequente de técnicas fundamentadas no movimento (*motion based*) ou em características (*feature based*).

No intuito de alcançar os objetivos formulados na Seção 1.3 desta dissertação, as publicações classificadas foram examinadas, buscando-se formas de combiná-las para a geração de uma abordagem automática de geração de vídeos ao estilo de cartuns. A abordagem concebida nesta pesquisa é detalhada e discutida no próximo capítulo.

Capítulo 3

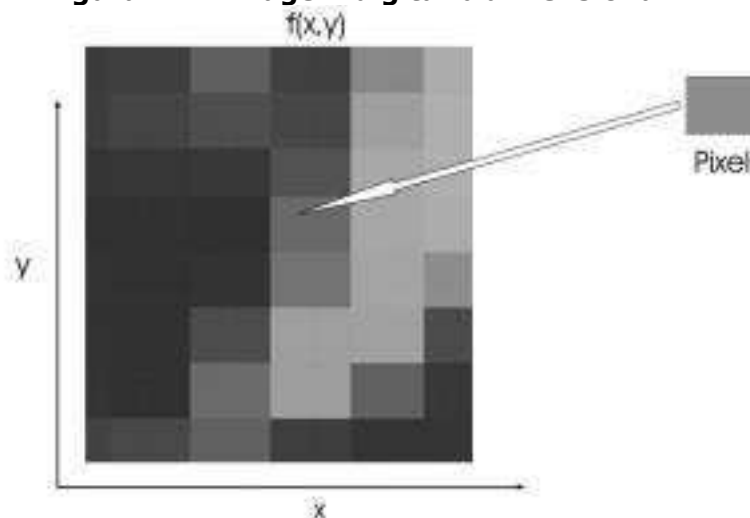
Fundamentação Teórica

No presente capítulo, descrevem-se os conceitos empregados na composição da abordagem proposta. Na Seção 3.1, são descritas as técnicas mais comuns de filtragem espacial. Na Seção 3.2, são descritas técnicas para a detecção de bordas. Na Seção 3.3, são discutidos conceitos relacionados ao processamento de imagens coloridas e espaços cromáticos. Na Seção 3.4, é discutida a rotulação de componentes conectados e, por fim, são formuladas as considerações finais na Seção 3.5.

3.1 Filtragem Espacial

A representação de uma imagem pode ser feita utilizando-se duas ou mais dimensões. Nesta pesquisa, foi considerada uma imagem digital como um arranjo bidimensional de pixels representada por uma função discreta $f(x,y)$, em que x e y são as coordenadas espaciais e a amplitude de f é a intensidade ou nível de cinza associado ao pixel (Ver Figura 21).

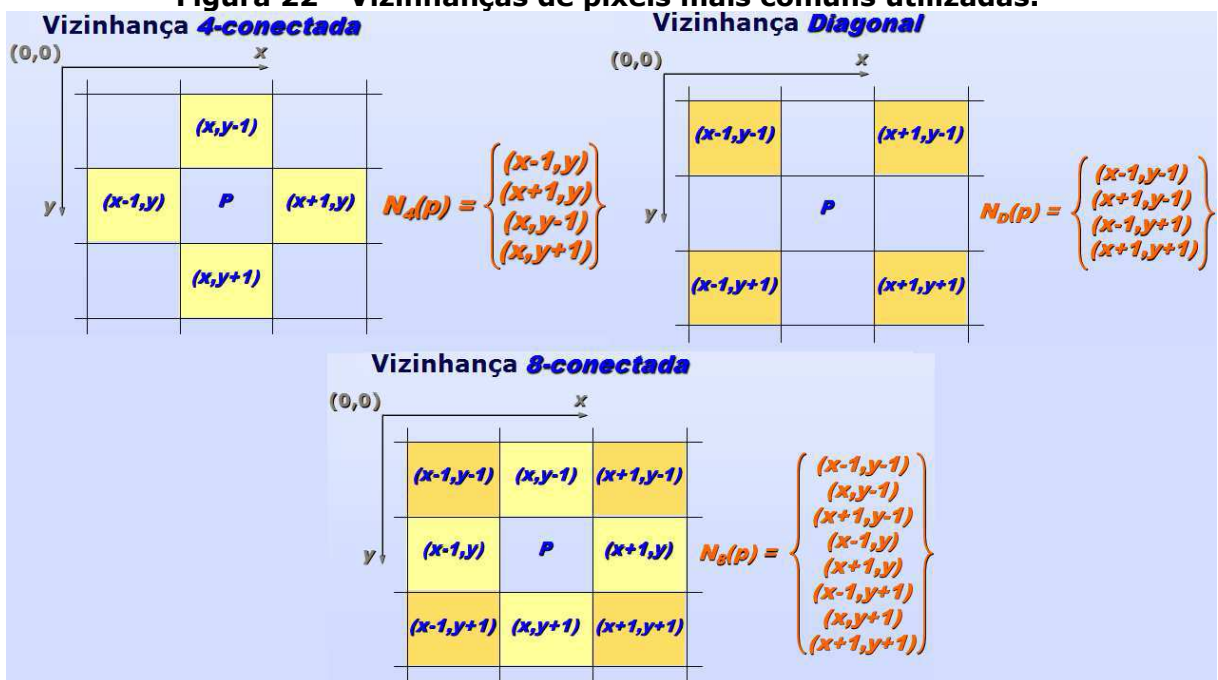
Figura 21 - Imagem digital bidimensional.



Fonte: <http://www.tecnologiaradiologica.com/digital.htm>.

A filtragem espacial foi realizada nesta pesquisa com o uso de operações locais, as quais consistem na manipulação dos valores dos pixels de uma determinada vizinhança da imagem. O valor de saída em uma determinada coordenada (x,y) depende tipicamente do valor de entrada e de uma vizinhança $(x+1, y)$, $(x, y+1)$, $(x-1, y)$, $(x, y-1)$, $(x+1,y+1)$, $(x-1, y+1)$, $(x+1, y-1)$ e $(x-1, y-1)$. Na Figura 22, são apresentadas as vizinhanças mais comuns utilizadas.

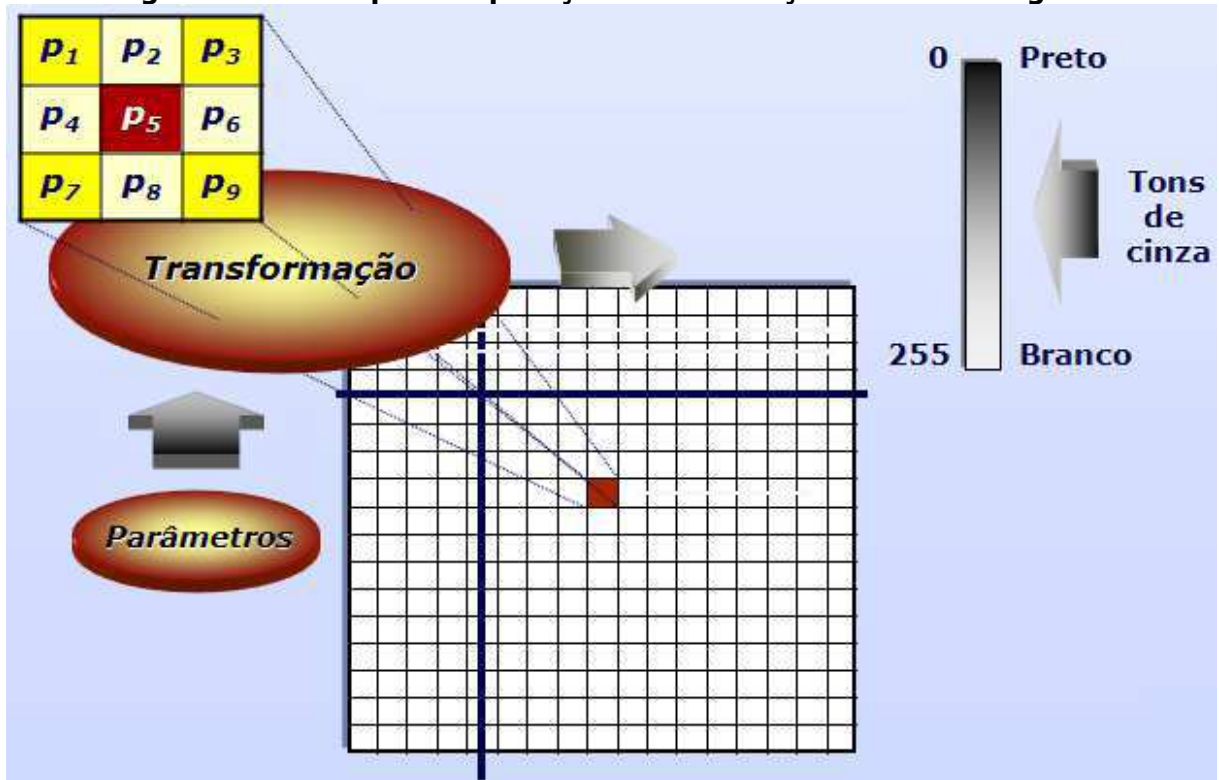
Figura 22 - Vizinhanças de pixels mais comuns utilizadas.



Fonte: <http://www.dsc.ufcg.edu.br/~rangel/pdi/NOTAS06.ppt>.

Na filtragem espacial realiza-se uma operação de convolução de uma máscara com uma imagem digital. A máscara é uma matriz de dimensões inferiores a imagem filtrada, normalmente apresenta a forma quadrada e os valores definidos na máscara representam pesos para a geração dos novos valores dos pixels. A convolução é executada sequencialmente nos pixels da imagem. Na Figura 23, é ilustrado o processo de convolução de uma máscara.

Figura 23 - Exemplo de aplicação de convolução em uma imagem.

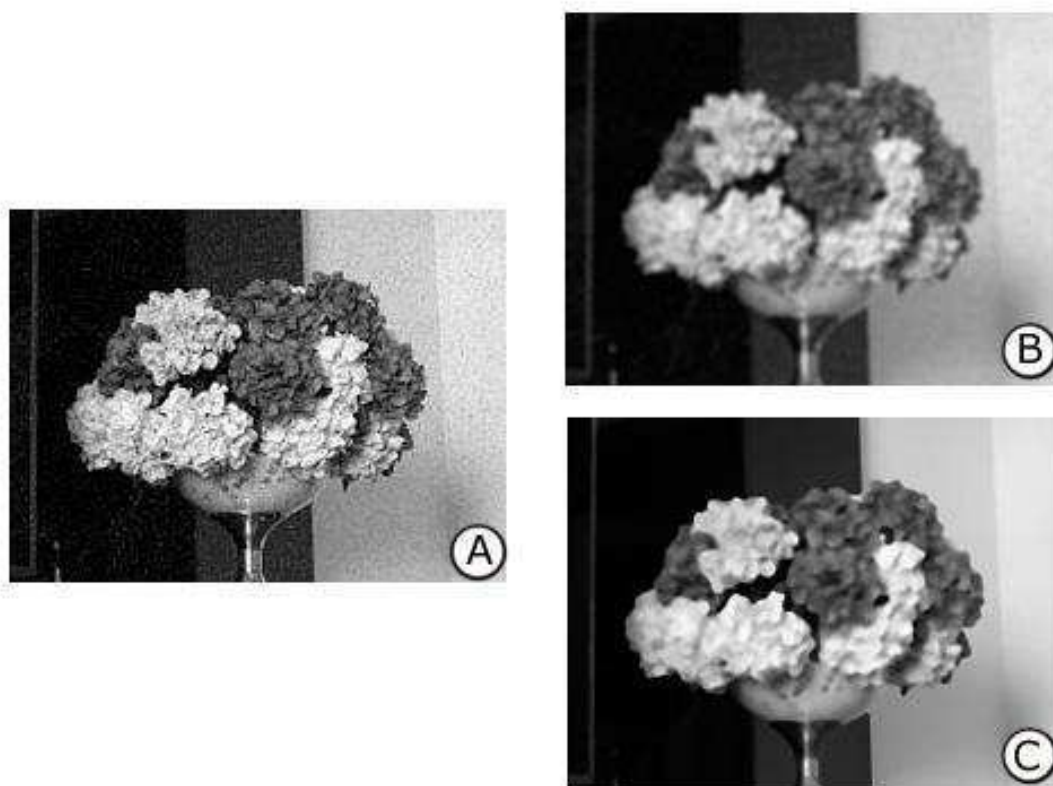


Fonte: <http://www.dsc.ufcg.edu.br/~rangel/pdi/NOTAS06.ppt>.

Na filtragem espacial é muito comum o uso dos filtros da média, mediana e moda para reduzir ruídos e borramento.

O filtro da média considera os valores dos pixels da vizinhança e faz uma média aritmética com os valores dos pixels incluindo o pixel central. Utilizando uma máscara de dimensões 3x3, cada pixel terá o peso de 1/9 no novo valor associado ao pixel central. O filtro da mediana e moda também consideram os valores de uma vizinhança de pixel, mas em vez de utilizar a média dos valores utilizam uma operação diferente. O filtro da mediana escolhe o valor correspondente a mediana dos valores da vizinhança e o filtro da moda escolhe o valor mais frequente. Um exemplo pode ser observado na Figura 24.

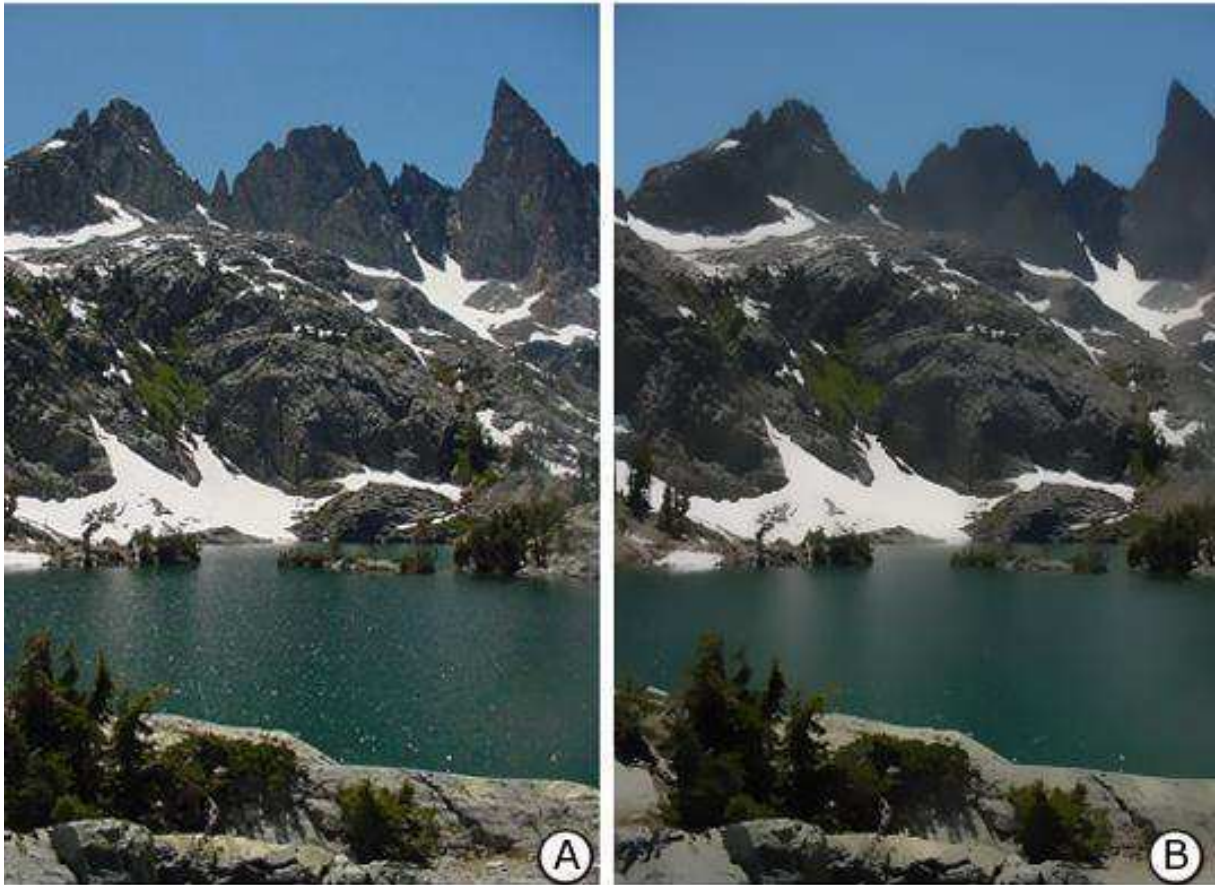
Figura 24 - Exemplo de aplicação de filtros espaciais: (A) Quadro original; (B) Filtro da média 9x9; (C) Filtro da mediana 9x9.



Fonte: https://pt.wikipedia.org/wiki/Filtragem_no_dom%C3%ADnio_espacial.

Filtros espaciais também podem ser utilizados para homogeneização de regiões com preservação das bordas, a exemplo do filtro bilateral. O filtro bilateral redefine a intensidade de cada pixel pela média ponderada das intensidades de pixels da vizinhança, os pesos podem ser determinados, por exemplo, utilizando-se uma distribuição gaussiana. O FABL é uma modificação do filtro bilateral conforme descrito na Seção 4.2. Um exemplo da aplicação do filtro bilateral pode ser observado na Figura 25.

**Figura 25 - Exemplo de aplicação do filtro bilateral:
(A) Quadro original; (B) Filtro bilateral.**



Fonte: https://en.wikipedia.org/wiki/Bilateral_filter.

3.2 Detecção de Bordas

Diferente dos filtros discutidos na seção anterior, o gradiente é uma técnica que se utiliza da primeira derivada para aguçar as bordas de uma imagem. Ao se considerar o pixel central da máscara de convolução de dimensões 3x3, como z_5 , tem-se o gradiente aproximado pela Equação 1.

$$\nabla f \approx |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)| \quad (1)$$

O filtro de Sobel utiliza a técnica do gradiente para representar as máscaras para o cálculo das derivadas da imagem em relação a x e y , dx e dy , conforme Equação 2.

$$dx = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \text{ e } dy = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (2)$$

Utilizando Sobel são calculadas duas derivadas: uma em relação a x e outra em relação a y para cada pixel. Calcula-se então, o quadrado de

cada uma das duas derivadas que são então somadas e por fim é aplicada a raiz quadrada. A Equação 3 formaliza este processo.

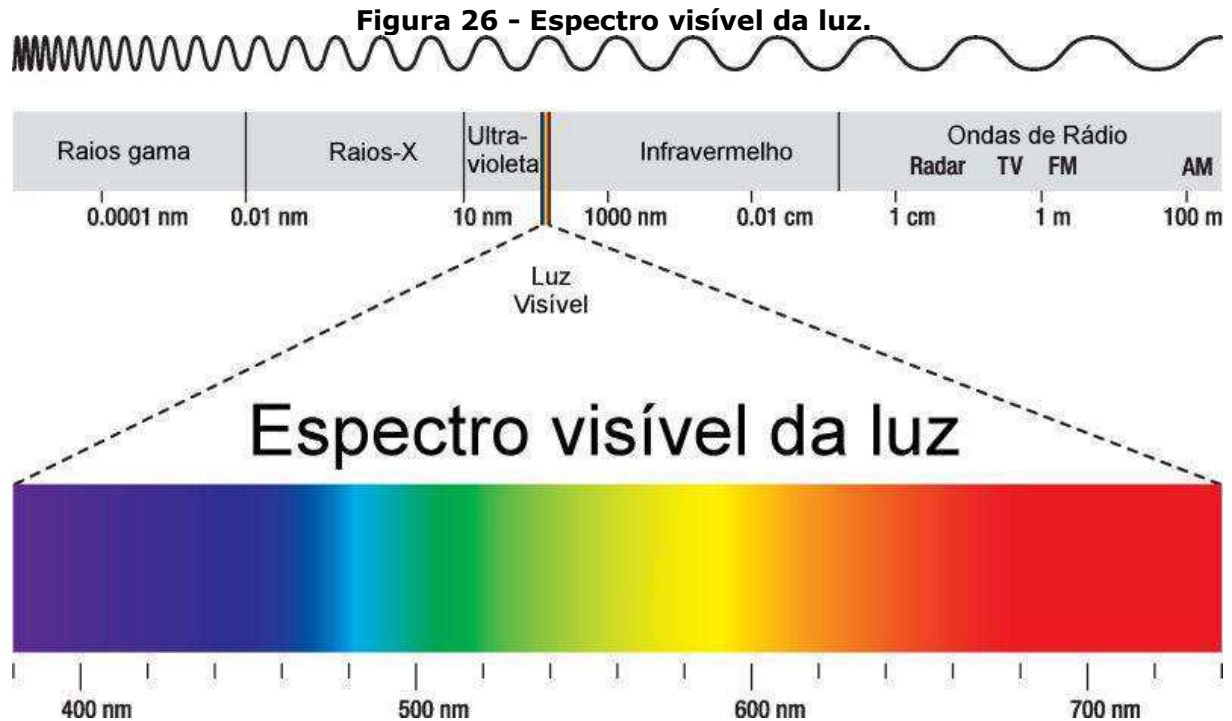
$$p = \sqrt{dx^2 + dy^2}, \quad (3)$$

em que p é o pixel atual, dx a derivada em relação a x do pixel p e dy a derivada em relação a y do pixel p .

3.3 Processamento de Imagens Coloridas

O fenômeno de percepção da cor não é algo totalmente compreendido, mas a natureza física da cor pode ser expressada de maneira formal por meio de experimentos e resultados teóricos.

As cores de um objeto para os seres humanos são resultantes da natureza da luz refletida pelo objeto observado. Essa luz é composta de várias frequências de ondas pertencentes ao espectro eletromagnético. Na Figura 26, o espectro de luz visível pode ser observado.



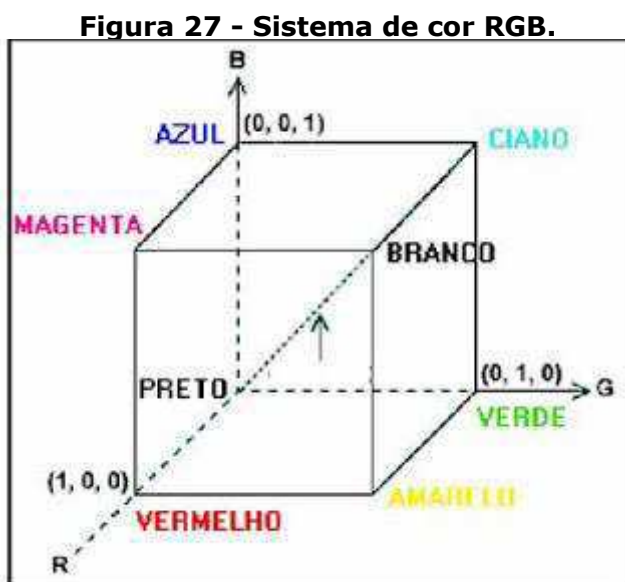
Fonte: <https://www.infoescola.com/fisica/espectro-eletromagnetico/>.

A luz visível emitida por objetos é resultante de um processo aditivo da soma das cores espectrais emitidas. O olho humano registra as cores como uma combinação de verde, vermelho e azul, sendo assim,

denominadas cores primárias. O processo de geração de cores do olho humano serviu de base para a criação do modelo cromático RGB (Red, Green e Blue).

Modelos cromáticos, também chamados de sistemas de cor ou espaço de cor, foram criados com o objetivo de facilitar a especificação de cores por associarem cores a pontos e coordenadas existentes nos modelos. Os modelos de cores utilizados nesta pesquisa foram o RGB e o HSV.

No modelo RGB a quantidade de bits utilizada para representar cada pixel é chamada de profundidade do pixel. A Figura 27 contém uma representação do sistema de cor RGB.



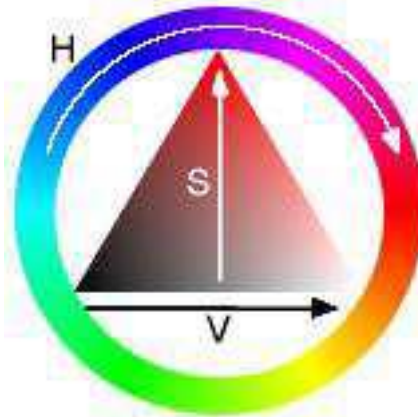
Fonte: <http://api12.blogspot.com/2012/03/rbg-cmyk-hsv-yuv-e-cores-em-html.html>.

A combinação de duas cores de mesma intensidade no sistema RGB produz as cores secundárias magenta, ciano e amarelo, conforme visto na Figura 27.

O HSV é um sistema de cor formado pela combinação das componentes matiz, saturação e valor. A matiz é a tonalidade da cor utilizada, ela abrange as cores do espectro desde o vermelho até o violeta mais o magenta. Os valores que podem ser assumidos pela matiz vão de 0 a 360. Saturação é referente a adição de branco a cor definida. Os valores de saturação podem variar de 0 a 1. O valor define o brilho da cor,

podendo variar de 0 a 1. O modelo HSV é utilizado para a realização de misturas de cores por sua tonicidade ser definida com a combinação de elementos de brilho e saturação, em vez de combinações de vermelho, verde e azul (RGB). A Figura 28 ilustra o sistema HSV.

Figura 28 - Sistema de cor HSV.



Fonte: <https://pt.wikipedia.org/wiki/HSV>.

Utilizando o modelo RGB, foi feita a quantização de cores dos quadros de entrada da abordagem proposta. A quantização de cores é o processo de redução da quantidade de cores presentes em uma imagem. A abordagem proposta fez uso da quantização de cores para segmentar conjuntos de pixels de cores próximas em cubos de níveis equidistantes de valores de brilho quantizados, conforme descrito na Seção 4.3, posteriormente utilizados para reduzir a quantidade de cores totais existentes.

A segmentação de cores é a subdivisão de uma imagem em regiões menores de acordo com as cores dos pixels presentes na imagem. Além da quantização de cores utilizada existem outras técnicas comumente utilizadas como o uso de limiares entre os valores de brilho dos pixels, o agrupamento de pixels utilizando o algoritmo k-means e etc.

Um exemplo de redução de cores pode ser observado na Figura 29 e um exemplo de segmentação de cores na Figura 30.

Figura 29 - Exemplo de redução de cores.



Fonte: <https://i.stack.imgur.com/I9hHy.jpg>.

Figura 30 - Exemplo de segmentação de cores: (A) Quadro original; (B) Quadro segmentado.



Fonte: <https://au.mathworks.com/matlabcentral/fileexchange/25257-color-image-segmentation>.

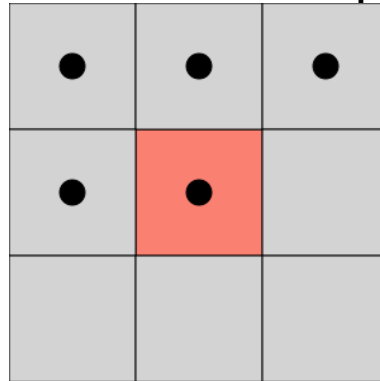
Utilizando o modelo HSV, foi feito o ajuste na saturação dos quadros do vídeo estilizado para a obtenção de uma maior similaridade com o gênero de cartuns tradicionais desejado.

3.4 Componentes Conectados

A rotulação de componentes conectados é um algoritmo de teoria dos grafos, em que subgrupos de componentes conectados são rotulados de maneira única.

Um pixel que tenha vizinho(s) próximo(s) e/ou vizinho(s) distante(s) com valores de intensidade próximos são ditos conectados. Na Figura 31, é ilustrado um pixel central conectado a outros 4 pixels adjacentes.

Figura 31 - Pixel central conectado a 4 pixels adjacentes.



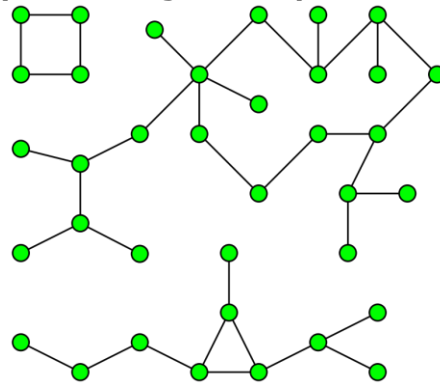
Fonte: https://en.wikipedia.org/wiki/Connected-component_labeling.

Na abordagem proposta, as imagens foram transformadas em imagens binárias para que apenas pixels de mesma intensidade fossem conectados.

O algoritmo utilizado executa uma varredura em uma imagem pixel a pixel para identificar regiões de pixels conectados. Regiões de pixels adjacentes que compartilham o mesmo valor de intensidade são rotulados com o mesmo rótulo. O processo de rotulação se repete até que todos os pixels da imagem sejam rotulados. É possível executar o algoritmo tanto em imagens binárias quanto em imagens em tons de cinza.

Na Figura 32, pode-se observar 3 regiões de pixels conectadas distintas.

Figura 32 - Exemplo de 3 regiões de pixels conectadas distintas.



Fonte: [https://en.wikipedia.org/wiki/Connected_component_\(graph_theory\)](https://en.wikipedia.org/wiki/Connected_component_(graph_theory)).

3.5 Considerações Finais

Neste capítulo, foram apresentadas as principais técnicas e conceitos utilizados no desenvolvimento e implementação da abordagem proposta.

Foram apresentados os filtros mais comuns utilizados na área, assim como o filtro bilateral que serviu como base para o desenvolvimento do FABL.

Os filtros utilizados para detecção de bordas, ao qual foi utilizado o filtro de Sobel que tem como base o gradiente.

Os conceitos utilizados para o processamento de imagens coloridas, assim como os espaços de cor utilizados no processo de segmentação (RGB) e no ajuste de saturação (HSV).

E por fim, discutiu-se a conexão de componentes realizada pelo algoritmo de rotulação existente em teoria dos grafos.

Capítulo 4

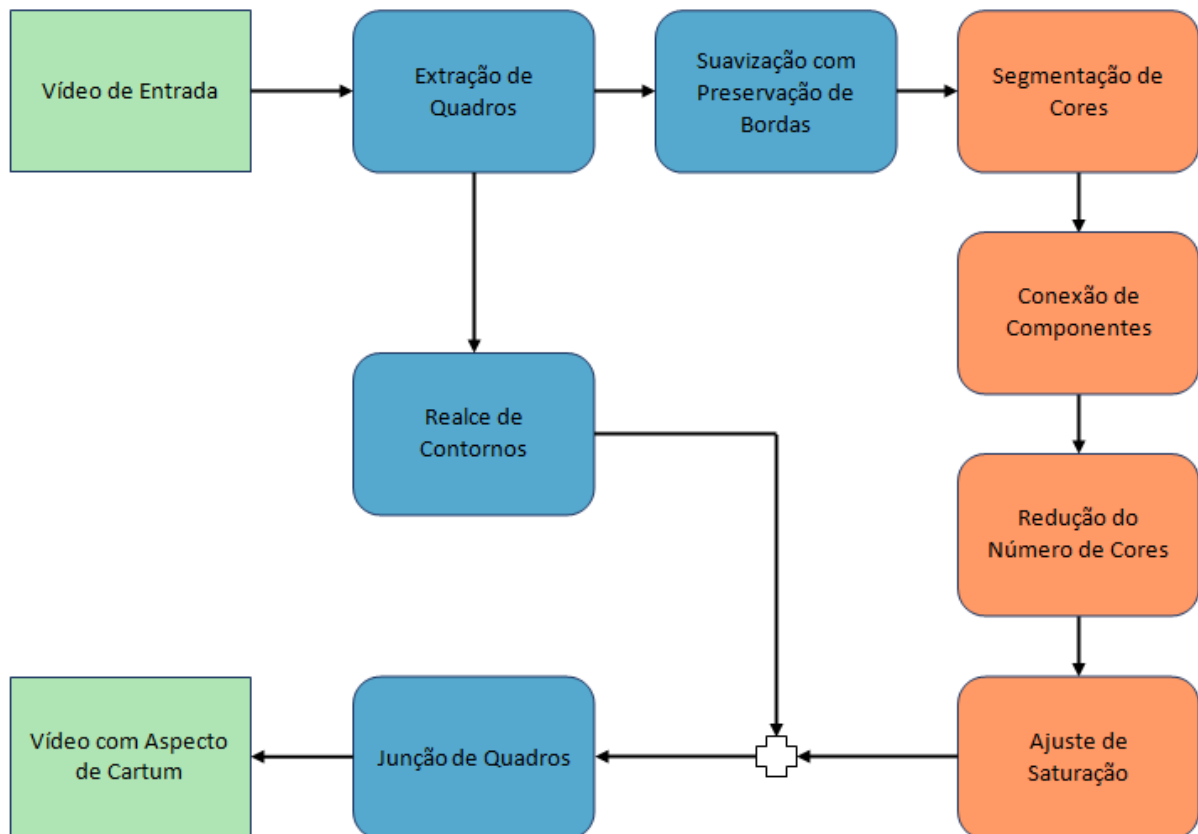
Abordagem Proposta

No presente capítulo, descreve-se a abordagem proposta nesta pesquisa. Nas Seções 4.1 à 4.7, são apresentados e comentados os módulos da abordagem proposta e também detalhadas as técnicas utilizadas para instanciar cada módulo. Na Seção 4.8, são discutidas as limitações da abordagem. Por fim, na Seção 4.9 são dispostas as considerações finais.

4.1 Descrição da Abordagem

A Figura 33 contém um diagrama que integra os módulos da abordagem proposta.

Figura 33 – Diagrama dos módulos da abordagem proposta.



Fonte: Autoria própria.

Um vídeo fotorrealístico é recebido como entrada, sendo processado conforme o fluxo indicado na Figura 33. Inicialmente, o vídeo de entrada é pré-processado, visando à extração dos quadros. Os demais módulos contidos na abordagem proposta são detalhados nas seções a seguir.

4.2 Suavização com Preservação de Bordas

O módulo de Suavização com Preservação de Bordas foi instanciado da pesquisa realizada por Alencar Jr. (2010), mais especificamente, fundamentou-se no filtro FABL (Filtro Bilateral Adaptativo Baseado em Fluxo) ali proposto. Tal módulo desempenha a função de preparar o vídeo para as etapas seguintes. Assim sendo, FABL é utilizado para suavizar as regiões presentes nos quadros do vídeo, visando a reduzir a presença de artefatos no vídeo (e.g., variações de movimento ou de iluminação). Um exemplo da saída do módulo pode ser observado na Figura 34.

Figura 34 – Exemplo da aplicação do FABL no módulo: (A) Quadro original; (B) Quadro resultante do FABL.



Fonte: Autoria própria.

O FABL é resultante de uma alteração no filtro FBL na filtragem na direção do gradiente das bordas. As funções dos deslocamentos e os sigmas, $\zeta(p)$ e $\sigma^2(p)$, respectivamente, da gaussiana no FBL foram substituídas no FABL pelas Equações 4 e 5, em que p é um pixel, s é um pixel da vizinhança de p , LoG é o operador Laplaciano da Gaussiana, MAX_{sigma} e MIN_{sigma} delimitam a faixa de valores em que a função da gaussiana é deslocada e m é uma constante.

$$\zeta(p) = m \cdot LoG(p) \quad (4)$$

$$\sigma^2(p) = \left(e^{\left(\frac{(LoG(p))^2}{2\sigma^2 s} \right)} (MAX_{sigma} - MIN_{sigma}) \right) + MIN_{sigma} \quad (5)$$

4.3 Segmentação de Cores

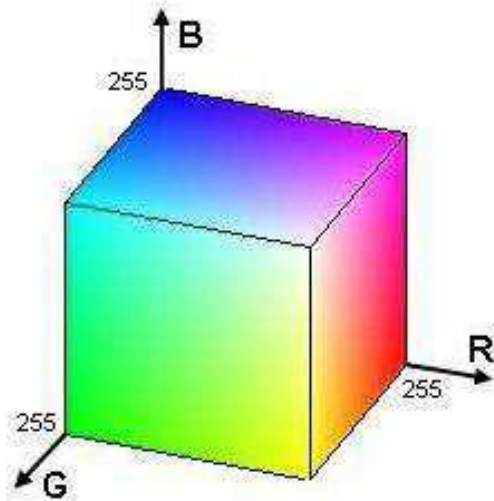
Por sua vez, o módulo de Segmentação de Cores recebe como entrada o resultado do FABL. A ideia principal deste módulo é segmentar cada quadro do vídeo em um conjunto de regiões, levando-se em consideração as cores, i.e., os componentes vermelho, verde e azul (RGB - *Red Green Blue*) de cada pixel.

A segmentação supramencionada é realizada por meio da quantização das cores dos pixels dos quadros de vídeo em cubos de níveis de quantização equidistantes de valores de brilho. Níveis de quantização pequenos resultam em maior quantidade de cubos, o que implica maior quantidade de cores da imagem quantizada.

Quanto maior a quantidade de cores contidas no quadro, mais semelhante ao quadro fotorrealístico original será o quadro estilizado. Sendo assim, níveis de quantização maiores reduzirão a quantidade total de cores no quadro resultante, tornando o quadro estilizado menos semelhante ao quadro fotorrealístico original.

Na Figura 35, observa-se o cubo RGB que contém 256x256x256 cores.

Figura 35 - Ilustração das possíveis cores no sistema RGB em forma de cubo ao utilizar 256 níveis de quantização por canal: (A) Quadro original; (B) Quadro Quantizado com único cubo de 256x256x256 níveis.

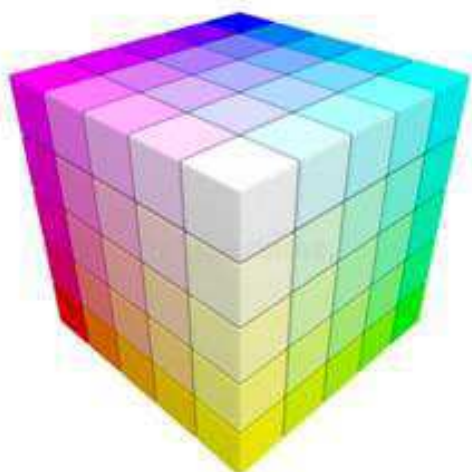


Fonte: <http://sidigicor.blogspot.com>.

A quantização do cubo RGB em um único nível converte o quadro num retângulo de cor única.

Na Figura 36, pode-se observar o cubo RGB fatiado em cubos de 52x52x52 níveis de cinza, o que possibilita a redução do número de cores do quadro a 125.

Figura 36 - Ilustração das possíveis cores no sistema RGB fatiado em cubos de 52x52x52 níveis de cinza: (A) Quadro original; (B) Quadro Quantizado com cubos de 52x52x52 níveis.



Fonte: <https://pt.dreamstime.com/imagem-de-stock-cubo-da-cor-do-rgb-amp-do-cmyk-image27461991>.

Utilizando-se estes cubos de 52x52x52 níveis de cinza é possível reduzir as cores do quadro a 125 cores. Na Figura 37, pode-se observar o cubo RGB fatiado em cubos de 32x32x32 níveis de cinza.

Figura 37 - Ilustração das possíveis cores no sistema RGB fatiado em cubos de 32x32x32 níveis de cinza: (A) Quadro original; (B) Quadro Quantizado com cubos de 32x32x32 níveis.



Fonte: <https://maalencar.wordpress.com/2012/03/29/color-models/>.

Utilizando-se cubos de 32x32x32 níveis de cinza, possibilita a redução das cores do quadro a 512 cores.

A definição de cubos de níveis de quantização adequados foi objeto de estudo antes da realização do experimento. A utilização de níveis de quantização pequenos demonstrou-se ineficiente no tocante ao processamento. O alto número de cubos gerados aumentava exponencialmente o tempo de execução do módulo. O uso de níveis de quantização grandes resultava na distorção do assunto do vídeo. Mais detalhes são apresentados na Seção 5.3 (**Experimentos Envolvendo a Abordagem Proposta**). A Figura 38 contém um exemplo de saída do módulo de segmentação de cores.

Figura 38 - Exemplo de saída do módulo de segmentação de cores.

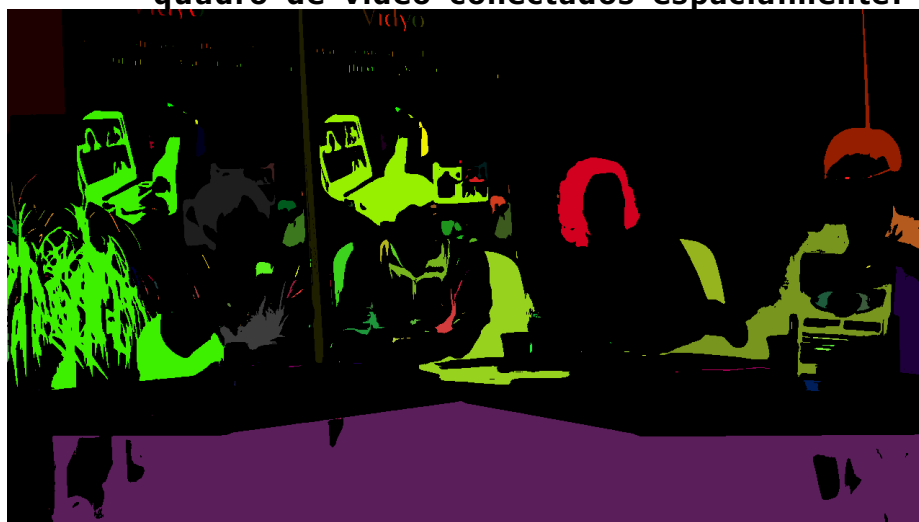


Fonte: Autoria própria.

4.4 Conexão de Componentes

Após o processamento no módulo de Segmentação de Cores, o módulo de Conexão de Componentes recebe como entrada os pixels que pertencem a um determinado nível de quantização. A próxima etapa é identificar quais desses pixels estão espacialmente conectados e, assim, marcá-los com o mesmo identificador (ver Figura 39). Isto possibilita distinguir pixels de mesma cor que estão espacialmente distanciados, indicando que possivelmente são referentes a objetos distintos presentes no mesmo quadro de vídeo.

Figura 39 - Pixels pertencentes a um nível de quantização de um quadro de vídeo conectados espacialmente.

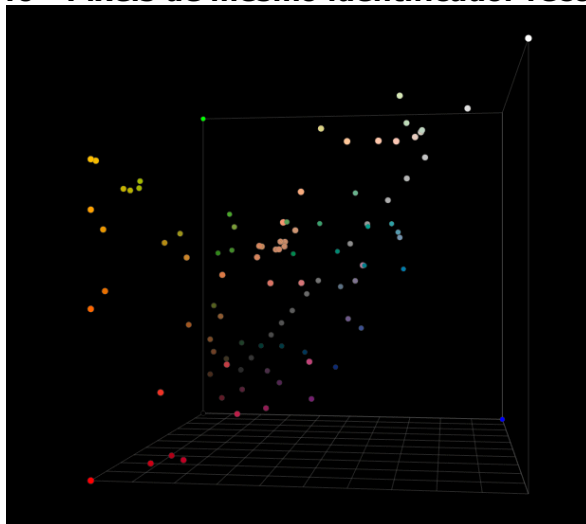


Fonte: Autoria própria.

4.5 Posterização

O módulo de Redução do Número de Cores executa uma varredura em cada um dos pixels com o mesmo identificador. Os valores correspondentes às cores originais daqueles pixels são somados e, então, divididos pelo total de pixels pertencentes àquele identificador, a fim de encontrar a cor central correspondente ao centro daquele grupo de pixels. Estes pixels de mesmo identificador tem suas cores modificadas para a cor central encontrada (ver Figura 40). O processo se repete até que todos os pixels de todos os níveis quantizados tenham suas cores redefinidas.

Figura 40 - Pixels de mesmo identificador recoloridos.



Fonte: <http://www.displaycalibrationtools.com/display-calibration-guides-and-workflows/advantage-of-custom-color-patch-sets-in-display-calibration/>.

4.6 Ajuste de Saturação

O módulo de Ajuste de Saturação realiza a tarefa de aumentar a saturação das cores presentes no vídeo em uma determinada porcentagem em relação ao original. O ajuste de saturação é feito seguindo a Equação 6.

$$S_f = S_i + a, \quad (6)$$

na qual S_f representa a saturação final do pixel, S_i a saturação inicial do pixel e a o ajuste de saturação realizado, o qual pode variar de 0 até 1, conforme o HSV. Caso o valor da S_f seja maior que 1, o valor da S_f é ajustado para o valor máximo 1.

Na Figura 41, é ilustrado um quadro com a saturação original aumentada em 50% ($a = 0,5$).

Figura 41 - Aumento da saturação original de um quadro de vídeo em 50%: (A) Quadro original; (B) Quadro ajustado.



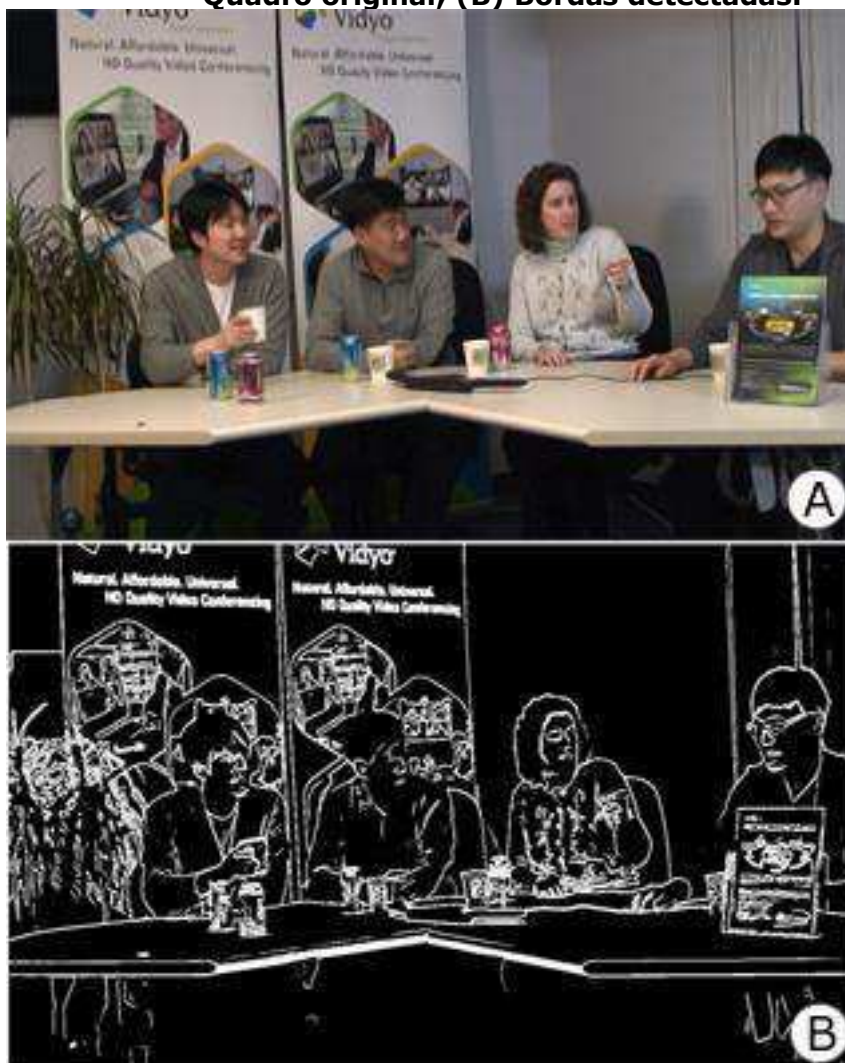
Fonte: Autoria própria.

4.7 Realce de Contornos

O módulo de Realce de Contornos foi instanciado a partir do filtro de Sobel, conforme descrito na Seção 3.2.

Ao final do processo, sobrepõem-se os contornos gerados pelo filtro de Sobel ao quadro com as cores reduzidas do qual as bordas foram extraídas. Um exemplo de saída do módulo pode ser observado na Figura 42.

Figura 42 - Detecção de bordas de um quadro de vídeo: (A) Quadro original; (B) Bordas detectadas.



Fonte: Autoria própria.

A composição do vídeo ao estilo de cartuns se dá por meio da integração sequencial dos quadros processados a partir da abordagem proposta. Ao final desse processo, tem-se como saída um vídeo não fotorrealístico.

Na Figura 43, é ilustrado um quadro resultante após o processamento por todos os módulos da abordagem proposta.

Figura 43 - Quadro exemplo do processamento realizado pela abordagem proposta: (A) Quadro original; (B) Quadro estilizado.



Fonte: Autoria própria.

Os módulos de Conexão de Componentes e Redução do Número de Cores são responsáveis pela redução da cintilação. A utilização de pixels espacialmente conectados evita que pixels de objetos diferentes, mas que possuem cores similares, influenciem na cor resultante escolhida. A Redução do Número de Cores altera a cor original de um pixel para nova cor similar reduzindo-se, assim, o impacto causado pela mudança de iluminação ao longo dos quadros.

4.8 Limitações da Abordagem Proposta

A abordagem proposta nesta dissertação, apesar de poder ser aplicada, em teoria, a qualquer vídeo fotorrealístico, possui suas limitações.

Uma variação brusca de iluminação em um vídeo durante a aplicação da abordagem pode resultar no surgimento de artefatos indesejáveis. O

módulo de Separação de Cores é fortemente afetado pela variação de iluminação.

Caso haja variação de iluminação durante o processo de seleção do nível de quantização à qual um determinado pixel pertence, este será mapeado para um nível de quantização diferente daquele para a qual ele deveria ser mapeado. Assim sendo, além de influenciar na definição do novo centro que será calculado para o nível de quantização que esse pixel foi alocado, também terá sua cor modificada incorretamente.

A limitação supramencionada ocorre quando existem pixels que transitam entre as fronteiras dos níveis quantizados, entre quadros adjacentes. O resultado desta mudança de nível resulta em cintilações ao renderizar uma sequência de quadros.

O método de segmentação baseia-se na quantização de cores, sendo assim, efeitos externos que causem sua alteração impactarão diretamente na precisão da segmentação. A limitação supracitada pode ser observada na Figura 44.

Figura 44 - Cintilação provocada pela mudança de nível de quantização de alguns pixels: (A) Quadro 1; (B) Quadro 2.



Fonte: Autoria própria.

4.9 Considerações Finais do Capítulo

A abordagem presente nesta dissertação foi descrita nas seções anteriores. Um aspecto a ressaltar são os problemas encontrados ao se utilizarem ideias advindas das pesquisas escolhidas como base para o desenvolvimento da abordagem proposta. A indisponibilidade de versões implementadas, *online* ou *offline*, que pudessem ser utilizadas inviabilizou a utilização de abordagens, a exemplo daquela proposta por Ben e Zelnik (2015).

Por fim, a abordagem proposta apresenta limitações, as quais foram apontadas na Seção 4.8.

Capítulo 5

Validação da Abordagem Proposta

Neste capítulo, é apresentada a metodologia utilizada para a validação da abordagem proposta e os resultados obtidos a partir de sua administração. Na Seção 5.1, é apresentado o planejamento experimental. Na Seção 5.2, são apresentados detalhes relacionados à análise dos resultados experimentais. Na Seção 5.3, são discutidos os parâmetros utilizados para produzir as versões NPR dos vídeos de teste. Na Seção 5.4, é apresentado o delineamento do perfil dos participantes do experimento. Na Seção 5.5, são sumarizados os dados obtidos durante o experimento de validação. Na Seção 5.6, são processados e discutidos os resultados obtidos na avaliação das versões NPR dos vídeos de teste da abordagem proposta e das abordagens concorrentes. Na Seção 5.7, são descritos detalhes de implementação. Por fim, na Seção 5.8 são dispostas as considerações finais.

5.1 Planejamento Experimental

O experimento conduzido consistiu da sondagem da opinião de observadores humanos acerca de versões NPR de vídeos, para cuja à avaliação constata-se a escassez de metodologias. Foi adotada a metodologia proposta por Arruda (2014), fundamentada no padrão ITU-R BT.500 (2012), publicado pela União Internacional de Telecomunicações para a normatização de avaliações subjetivas da qualidade de imagens de TV.

Foram considerados os mesmos critérios adotados por Arruda (2014), advindos das características presentes em cartuns, conforme Ianeva, Vries e Rohrig (2003), Medhi, Patabandhige e Toyama (2007) e Alencar, Queiroz e Gomes (2013).

Os critérios supramencionados são:

(I) **Simplicidade**: Uniformidade das áreas coloridas com contornos mais escuros;

(II) **Preservação do conteúdo**: Inexistência de partes faltantes ou falhas que descaracterizem o tema do vídeo;

(III) **Semelhança ao conteúdo original**: O vídeo com aspecto de cartum deve apresentar o mesmo conteúdo semântico que o vídeo original, sendo facilmente reconhecível;

(IV) **Nitidez**: O vídeo com aspecto de cartum deve ter o tema apresentado de forma nítida;

(V) **Coerência Temporal**: Ausência de cintilação ou artefatos indesejados no vídeo com aspecto de cartum;

(VI) **Satisfação Subjetiva**: O vídeo com aspecto de cartum deve ter um aspecto visual que satisfaça ao participante.

Os 6 critérios adotados para a avaliação dos vídeos foram mapeados diretamente em 6 questões. Assim sendo, cada participante votou em 6 escalas para cada um dos 45 vídeos (duração média de 11s cada), totalizando 270 votações por teste (6 escalas × 45 vídeos).

As questões que sumarizam todos os critérios são apresentadas a seguir:

1. O vídeo ao estilo de cartuns apresenta áreas coloridas homogêneas com contornos escuros;
2. O vídeo ao estilo de cartuns não apresenta falhas (partes faltantes) em alguns objetos;

3. A ideia principal transmitida pelo vídeo original foi preservada no vídeo ao estilo de cartuns;
4. Os objetos e as cenas retratadas no vídeo ao estilo de cartuns são nítidas e de fácil percepção;
5. O vídeo ao estilo de cartuns não apresenta variações frequentes de brilho e/ou cor (efeito pisca-pisca);
6. O vídeo ao estilo de cartuns apresenta um aspecto visual agradável.

A avaliação realizada pelos participantes utilizou uma escala de Likert de 5 pontos, em que o participante deveria selecionar uma das seguintes âncoras semânticas: **Discordo totalmente**, **Discordo parcialmente**, **Não concordo nem discordo**, **Concordo parcialmente** e **Concordo totalmente**. Utilizou-se um universo amostral de 60 participantes, a fim de se ter uma amostra minimamente representativa (HOGG, TANIS, 1993) e, assim, ser possível a administração de testes estatísticos com resultados mais robustos.

Os vídeos de teste foram escolhidos da coletânea de vídeos de teste **derf's collection**⁷, disponibilizada pela Xiph.Org Foundation⁸, uma das bases de maior destaque na utilização por projetos relacionados à área de multimídia.

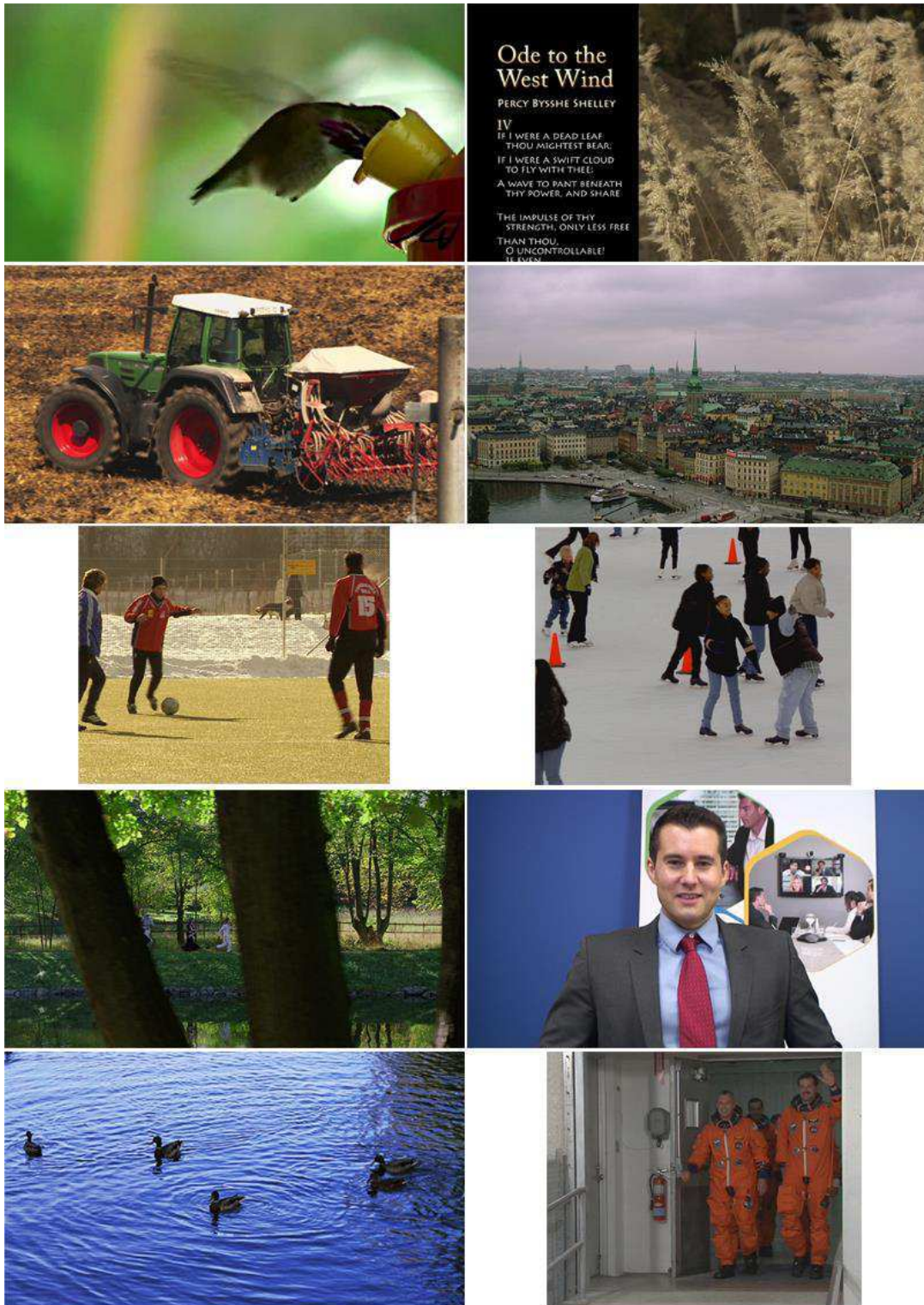
A **derf's collection** totaliza 78 vídeos, dos quais foram utilizados para a avaliação da abordagem proposta vídeos com temas variados e duração de pelo menos 5 segundos a não mais que 50 segundos. Tal seleção ocorreu para que o participante do experimento não fosse submetido a vídeos muito curtos que poderiam atrapalhar sua avaliação ou vídeos muito longos que tornariam as sessões muito extensas.

A Figura 45 contém o primeiro quadro de cada um dos 10 vídeos finais selecionados.

⁷ <https://media.xiph.org/video/derf/>

⁸ <https://www.xiph.org>

Figura 45 - Quadro inicial de cada um dos 10 vídeos selecionados.



Fonte: Autoria própria.

As abordagens concorrentes escolhidas para a validação reportada neste capítulo foram: (i) Filtro Anisotrópico Kuwahara (*Anisotropic Kuwahara Filter*) (KYPRIANIDIS, KANG, DÖLLNER, 2009); (ii) Filtro de

Aprimoramento Coerente (*Coherence-Enhancing Filtering*) (KYPRIANIDIS, KANG, 2011) e (iii) Filtro Estendido de Diferença de Gaussianas (*XDoG Filter*) (WINNEMÖLLER, 2012). Tais abordagens foram selecionadas por terem a mesma motivação desta pesquisa de mestrado, que é a geração de vídeos não fotorrealísticos ao estilo de cartuns. A configuração dos parâmetros das abordagens utilizadas ocorreu conforme recomendado pelos autores (ver Apêndice C).

As abordagens concorrentes utilizam a abstração como recurso para gerar suas estilizações, enquanto na abordagem proposta é utilizada a segmentação de regiões. O uso de segmentação em detrimento à abstração acarreta os problemas anteriormente discutidos na Seção 1.1. Entretanto, o uso das diferentes abordagens se fez necessário dada a escassez de abordagens fundamentadas em segmentação disponíveis no estado da arte.

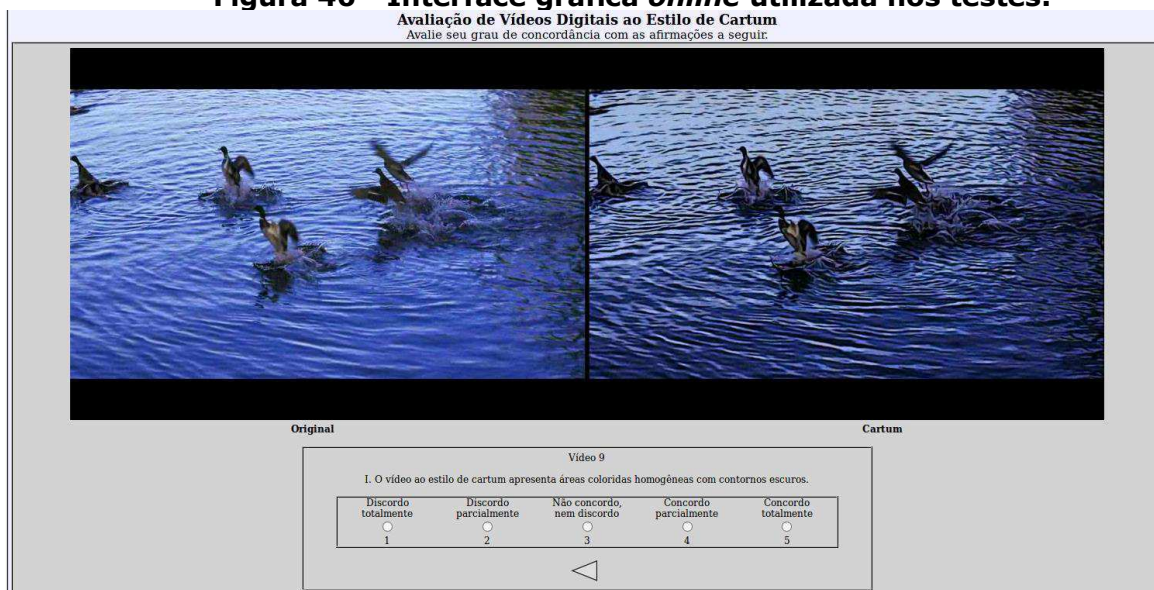
A população-alvo dos testes contemplou indivíduos com faixa etária superior a 18 anos, no intuito de facilitar o processo de recrutamento de voluntários em um ambiente de educação superior. Optou-se por não estratificar a amostra considerada devido à ampla disseminação do acesso/uso de vídeos dos mais variados gêneros disponíveis na Web por indivíduos de praticamente todas as faixas etárias, camadas sociais e níveis de instrução.

As sessões do experimento foram iniciadas com explicações acerca da confidencialidade da votação, da finalidade do experimento e do anonimato do participante. Em seguida, foi apresentado o Termo de Consentimento Livre e Esclarecido (TCLE), a ser assinado pelo participante.

Após a assinatura do TCLE foi administrado um questionário para delinear o perfil do participante do experimento (ver Apêndice A). Na etapa seguinte, o participante opinou sobre os vídeos selecionados para o teste.

Foi concebida uma interface Web *online*, inspirada na pesquisa de Arruda (2014) (ver Figura 46), na qual foram visualizados os vídeos par-a-par e a escala de Likert utilizada pelo respondente. Os dados de identificação do vídeo e da abordagem utilizada foram removidos, afim de que o participante do experimento não pudesse identificar o vídeo nem a abordagem que o produziu.

Figura 46 - Interface gráfica *online* utilizada nos testes.



Fonte: Autoria própria.

Cada participante votou 45 vezes, comparando os vídeos par-a-par. Assim sendo, a sessão de teste compreendeu a visualização de 45 pares de vídeos, com duração média de 10 segundos. A distribuição dos vídeos aos participantes foi realizada de forma equitativa. A transição entre os pares de vídeos apresentados só ocorria após a votação de todos os critérios supramencionados. O áudio de todos os vídeos foi suprimido, afim de evitar distrações aos participantes e manter o foco na avaliação visual dos vídeos.

Na Figura 47, é apresentada uma visão geral do ambiente no qual foram realizados os testes com os participantes do experimento.

Figura 47 - Foto do ambiente de testes.



Fonte: Autoria própria.

Ao final de todas as sessões de votação, foi feita a contabilização dos dados e a sua análise.

5.2 Aspectos Relativos à Análise dos Resultados Experimentais

Por ser de caráter visual e estritamente pessoal, a avaliação dos vídeos pelos participantes está sujeita a ambiguidades e subjetividades. Assim sendo, buscou-se no processamento estatístico dos dados coletados o suporte ferramental para elucidar as questões levantadas acerca de cada vídeo.

Levando-se em consideração as questões apresentadas na seção anterior, a etapa inicial da avaliação individual de cada vídeo por um determinado participante resultou na aquisição de dados, cuja normalidade foi testada em seguida, a fim de que o teste estatístico inferencial fosse escolhido.

O método estatístico escolhido foi o Kruskal-Wallis (KRUSKAL, WALLIS, 1952), aplicado nos dados coletados para cada critério a fim de verificar a existência de diferenças entre as abordagens. Verificada a

existência de diferenças entre as avaliações, fez-se necessária a aplicação do teste de Wilcoxon (WILCOXON, 1945) entre pares de abordagens com o intuito de determinar qual delas foi mais bem avaliada em relação a cada um dos critérios.

Primeiramente, buscou-se validar os parâmetros da abordagem proposta, comparando-se as diferentes configurações para identificar a melhor delas. Após a identificação da melhor configuração, foi necessário compará-la com as abordagens concorrentes, afim de verificar qual delas é preferida pelos participantes. Por fim, com os resultados estatísticos em mãos, decidiu-se sobre a aceitação ou rejeição das hipóteses nulas.

5.3 Experimentos Envolvendo a Abordagem Proposta

Os parâmetros que necessitavam serem definidos na abordagem proposta, conforme descrito no Capítulo 4, totalizavam 3, a saber: (i) Largura do nível de quantização; (ii) Número de aplicações do FABL; e (iii) Nível de saturação.

A largura do nível de quantização pode variar discretamente entre 1 a 256 pixels, em potências de 2. Estes valores garantem que os cubos resultantes terão mesmas dimensões pois, ao se utilizarem outros valores, os cubos próximos às extremidades (0 e 255) teriam larguras, alturas e profundidades inferiores àquelas dos demais. Tal redução, afeta diretamente o processo de mapeamento dos pixels aos níveis de quantização correspondentes, assim, pixels seriam mapeados para cores incompatíveis com aquelas presentes no quadro de entrada.

Ao se testar experimentalmente os níveis de quantização supracitados, notou-se que valores inferiores a 64 resultava em uma quantidade de cubos que elevava consideravelmente o tempo de execução do módulo de Segmentação de Cores, tornando o processo inviável. Em contrapartida, ao se utilizarem níveis de quantização superiores a 64, observou-se que a baixa quantidade de cubos associada resultava em perda acentuada dos detalhes presentes nos quadros. Sendo assim,

optou-se por se utilizar cubos com largura igual a 64 pixels, pois o tempo de processamento era aceitável, além de se evitar perda excessiva de detalhes nos quadros renderizados.

O número de aplicações sucessivas do FABL foi objeto de estudo experimental. Foi possível se observar que, ao se iterar o FABL mais de 3 vezes a um mesmo quadro, as regiões eram excessivamente homogeneizadas, o que prejudicava os demais módulos da abordagem proposta. Portanto, decidiu-se limitar a faixa de iterações do FABL de 1 a 3 em cada quadro. Entretanto, devido à diferença visual ser perceptível apenas com 1 e 3 iterações do FABL, foram selecionados os valores de 1 e 3.

O nível de saturação pode variar de 0 a 1, sendo 0 correspondente à cor com adição máxima de branco e 1 à cor com mínima adição de branco. Entretanto, como não seria possível testar todas as possibilidades, foram selecionados os valores 0, 0,3 e 0,6.

O número de combinações possíveis dos parâmetros da abordagem proposta para cada vídeo totalizou 6 (2 [número de aplicações do FABL possíveis] × 3 [ajustes de saturação possíveis]), o que resultou no teste de 60 vídeos (6 × 10 vídeos escolhidos da base de dados) renderizados a partir da abordagem proposta.

A Tabela 1 contém os parâmetros correspondentes a cada uma das 6 configurações possíveis da abordagem proposta, utilizadas durante os experimentos com os participantes.

Tabela 1 - Parâmetros utilizados em cada uma das configurações da abordagem proposta.

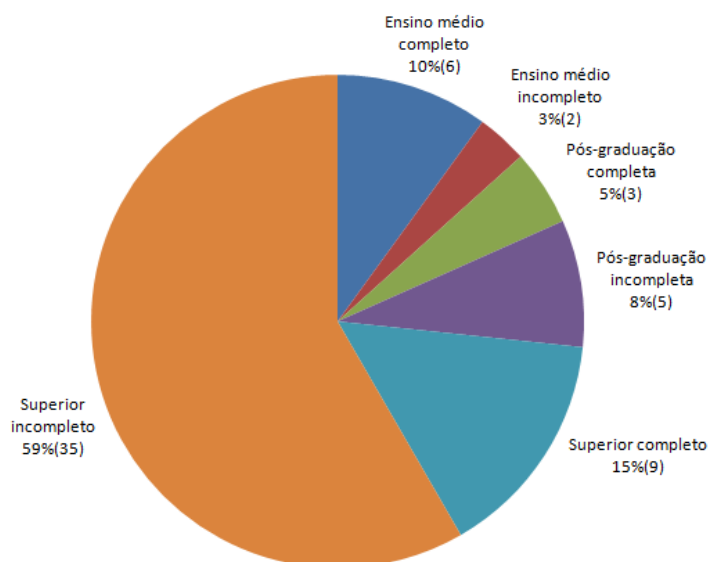
Configuração	Aplicações do FABL	Ajuste de Saturação
1	1	0
2	1	0,3
3	1	0,6
4	3	0
5	3	0,3
6	3	0,6

Fonte: Autoria própria.

5.4 Delineamento do Perfil dos Participantes do Experimento

Os participantes do experimento preencheram um formulário sociocultural, conforme descrito na Seção 5.1. A partir desses formulários preenchidos, foi feito o delineamento do perfil dos participantes. A Figura 48 contém um gráfico referente ao grau de instrução dos participantes do experimento.

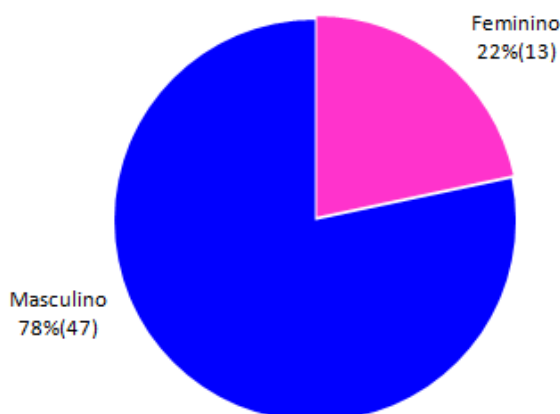
Figura 48 - Gráfico do grau de instrução dos participantes.
Grau de instrução



Fonte: Autoria própria.

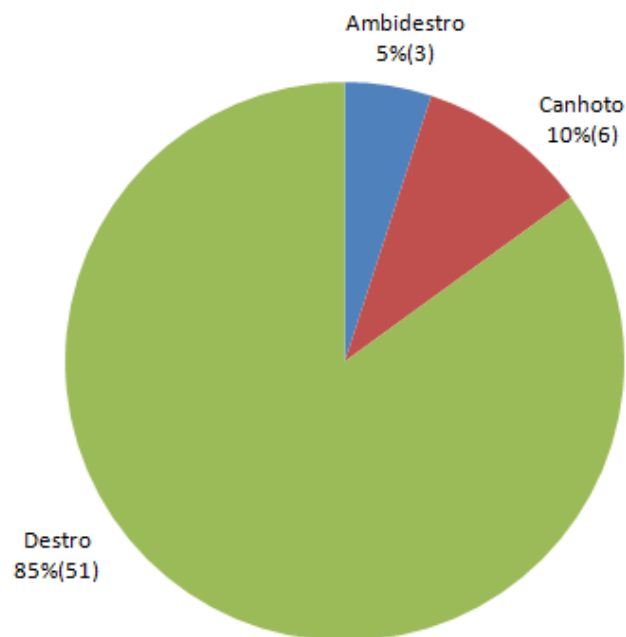
Outras informações acerca dos participantes podem ser observadas nas Figuras 49 a 57.

Figura 49 - Gráfico referente ao sexo dos participantes.
Sexo



Fonte: Autoria própria

**Figura 50 - Gráfico referente a escrita dos participantes.
Escrita**



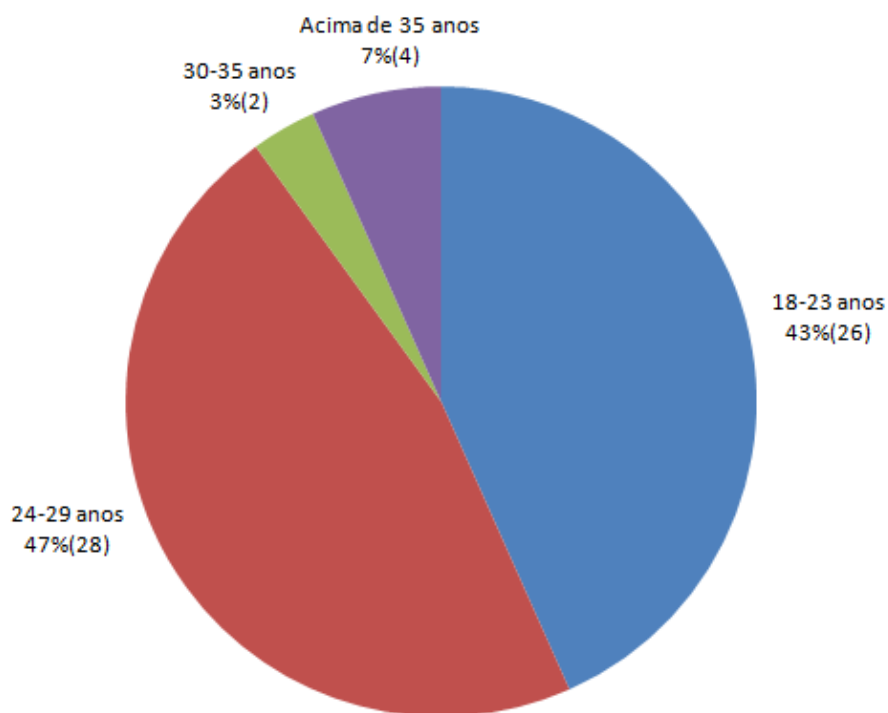
Fonte: Autoria própria.

**Figura 51 - Gráfico referente ao uso de óculos e lentes dos participantes.
Uso de óculos ou lentes**



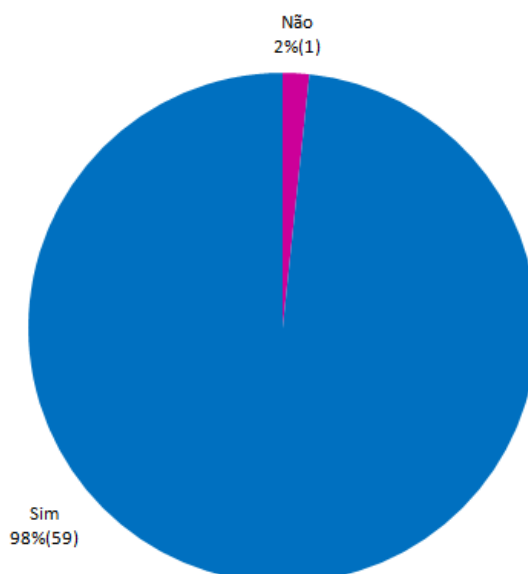
Fonte: Autoria própria.

**Figura 52 - Gráfico referente a faixa etária dos participantes.
Faixa etária**



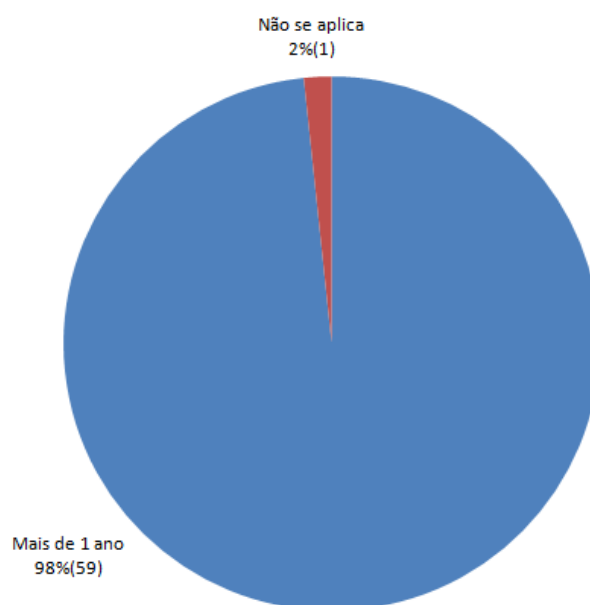
Fonte: Autoria própria.

**Figura 53 - Gráfico referente a experiência prévia com sistemas computacionais dos participantes.
Experiência prévia com sistemas computacionais**



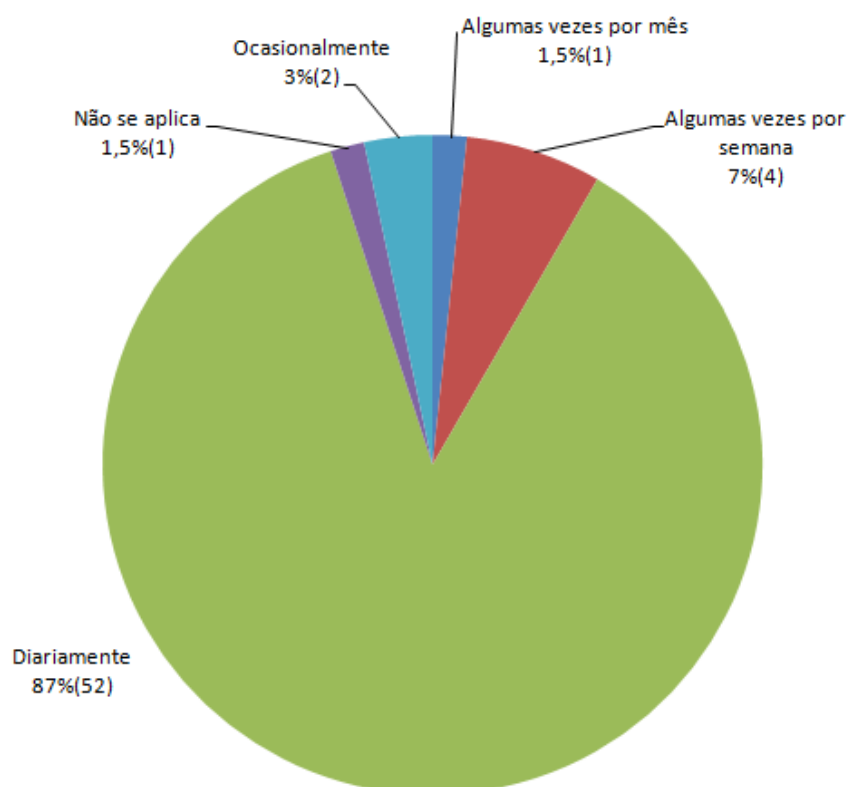
Fonte: Autoria própria.

Figura 54 - Gráfico referente ao tempo de uso de sistemas computacionais dos participantes.
Tempo de uso de sistemas computacionais



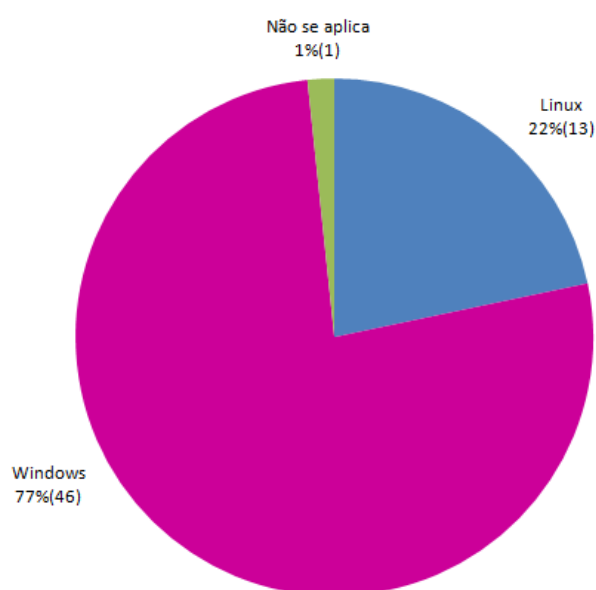
Fonte: Autoria própria.

Figura 55 - Gráfico referente a frequência do uso de sistemas computacionais dos participantes.
Frequência do uso de sistemas computacionais



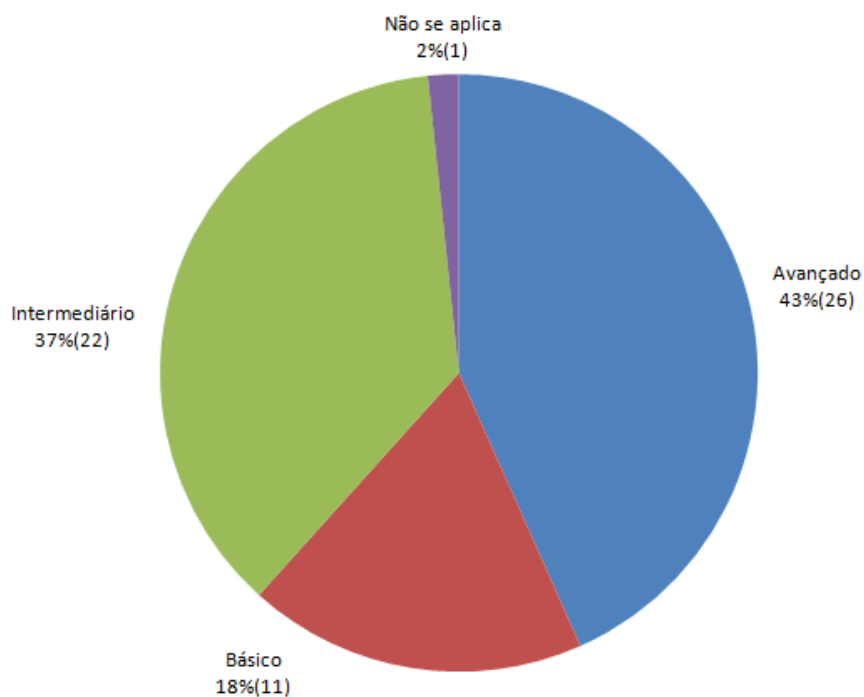
Fonte: Autoria própria.

**Figura 56 - Gráfico referente à plataforma mais utilizada pelos participantes.
Plataforma computacional utilizada**



Fonte: Autoria própria.

**Figura 57 - Gráfico referente ao nível de conhecimento em informática dos participantes.
Nível de conhecimento em informática**



Fonte: Autoria própria.

O participante médio do experimento é universitário, do sexo masculino, destro, usa óculos ou lentes, tem entre 24 e 29 anos, tem experiência com sistemas computacionais, usa sistemas computacionais

há mais de um ano, usa o computador diariamente, utiliza o sistema operacional Windows e é um usuário com conhecimentos avançados em informática.

Após as sessões de votação, foi solicitado também aos participantes que sugerissem contextos para o processo de renderização apresentado. A Figura 58 contém um gráfico com os contextos sugeridos pelos participantes. Os participantes que não quiseram sugerir contexto algum estão contabilizados no contexto **Nenhum**.

Figura 58 - Gráfico referente aos contextos de utilização dos vídeos renderizados sugeridos pelos participantes.
Total de participantes que sugeriram o contexto de uso



Fonte: Autoria própria.

Após a finalização de todas as sessões do experimento, o qual teve a duração de 3 semanas e sessões com média de 40 minutos de duração, passou-se à análise dos resultados obtidos.

5.5 Resultados da Avaliação do Processo de Renderização de Vídeos

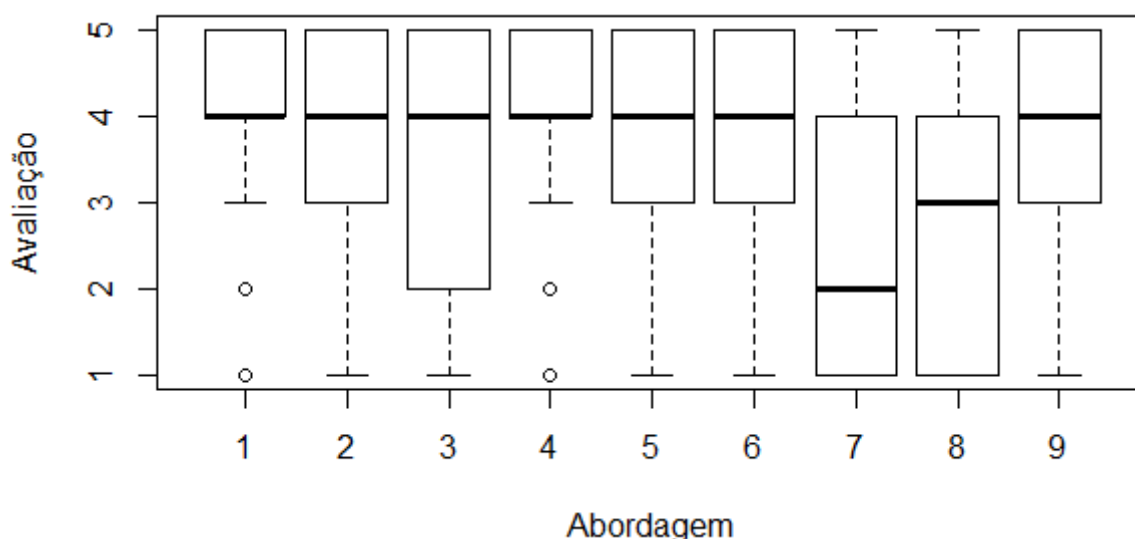
No tocante às abordagens concorrentes, foram renderizados 30 vídeos adicionais (3 concorrentes × 10 vídeos). O total de vídeos gerados, considerando-se a abordagem proposta e as abordagens concorrentes, foi de 90 vídeos, divididos em dois grupos de 45, sendo o primeiro grupo de vídeos votados pelos 30 primeiros participantes do experimento e o segundo grupo votado pelos 30 demais participantes.

Os votos obtidos após a execução das 60 sessões de teste estão sumarizados nas Figuras 59 a 64.

O eixo x de cada figura contém as abordagens de 1 à 9, correspondendo, respectivamente, às 6 configurações de parâmetros possíveis da abordagem proposta, i.e., Filtro Anisotrópico de Kuwahara, Filtro de Aprimoramento Coerente e Filtro Estendido de Diferença de Gaussianas.

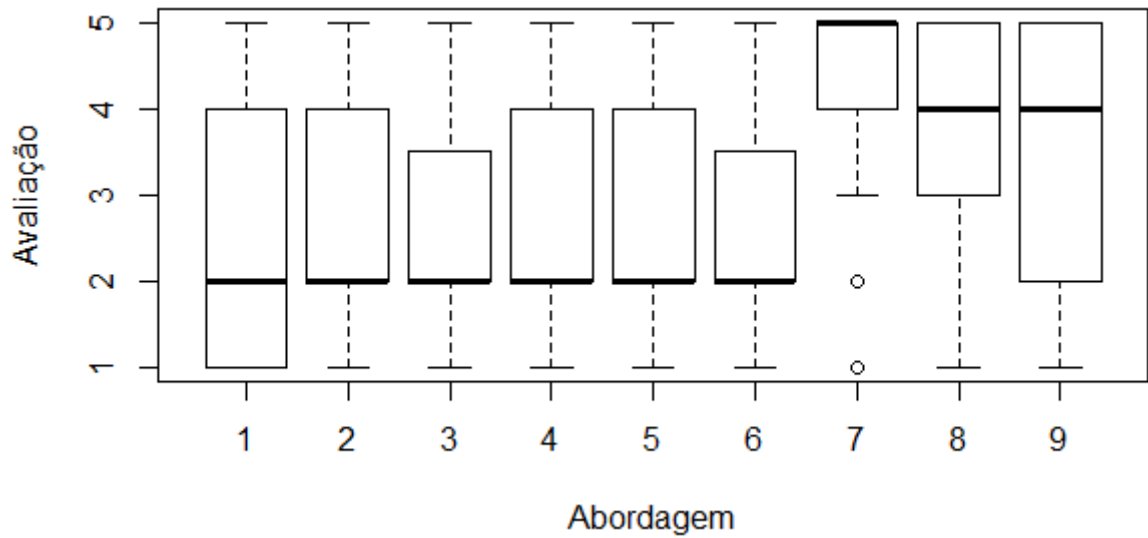
O eixo y de cada figura corresponde ao grau de concordância (representado na escala de *Likert*) dos participantes em relação ao critério avaliado, em que 1 corresponde a Discordo Totalmente e 5 a Concordo Totalmente.

Figura 59 - Boxplot dos votos dos participantes para o critério 1.



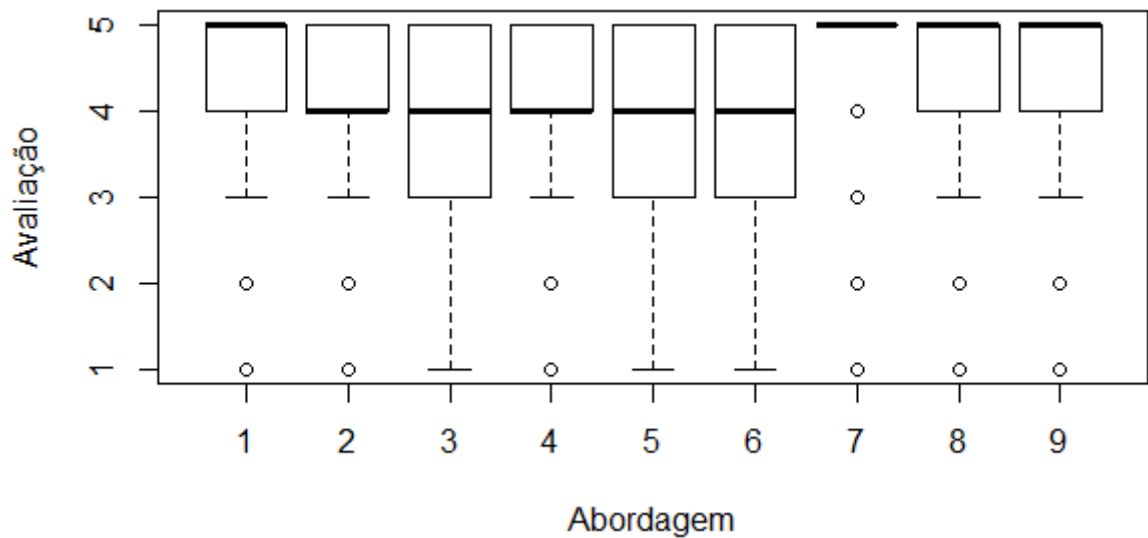
Fonte: Autoria própria.

Figura 60 - Boxplot dos votos dos participantes para o critério 2.



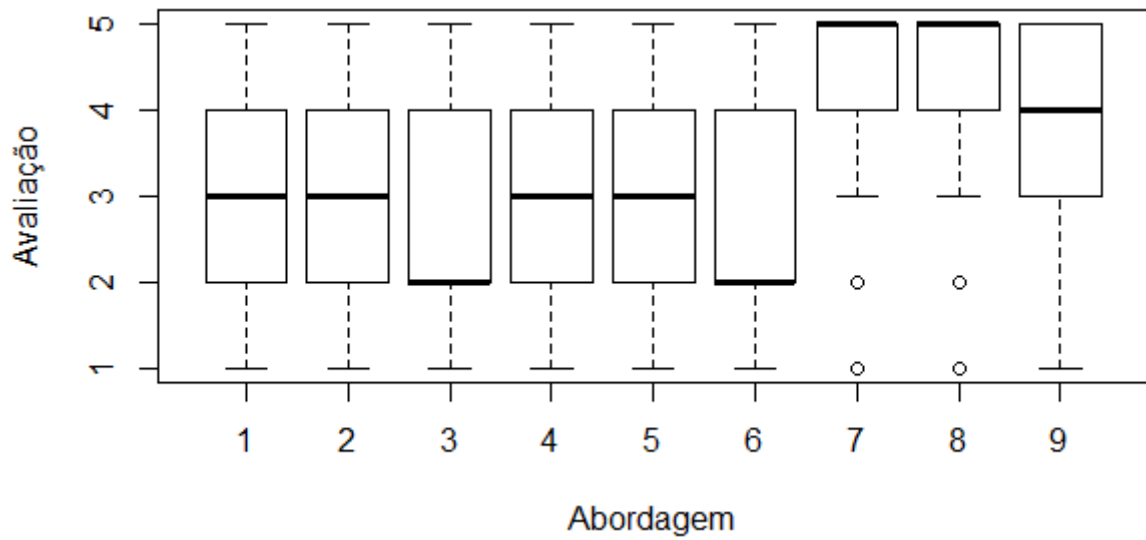
Fonte: Autoria própria.

Figura 61 - Boxplot dos votos dos participantes para o critério 3.



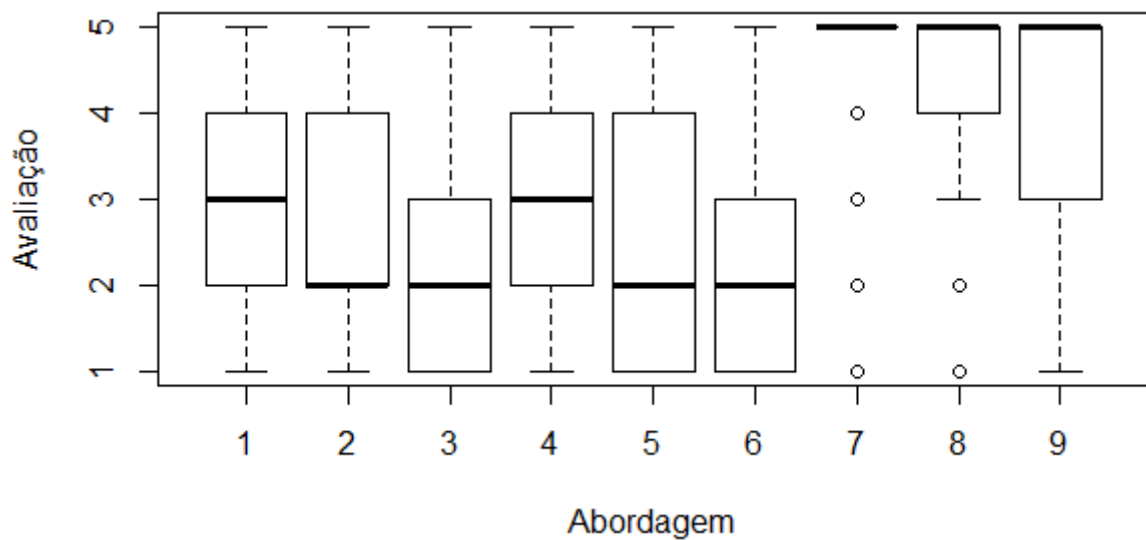
Fonte: Autoria própria.

Figura 62 - Boxplot dos votos dos participantes para o critério 4.



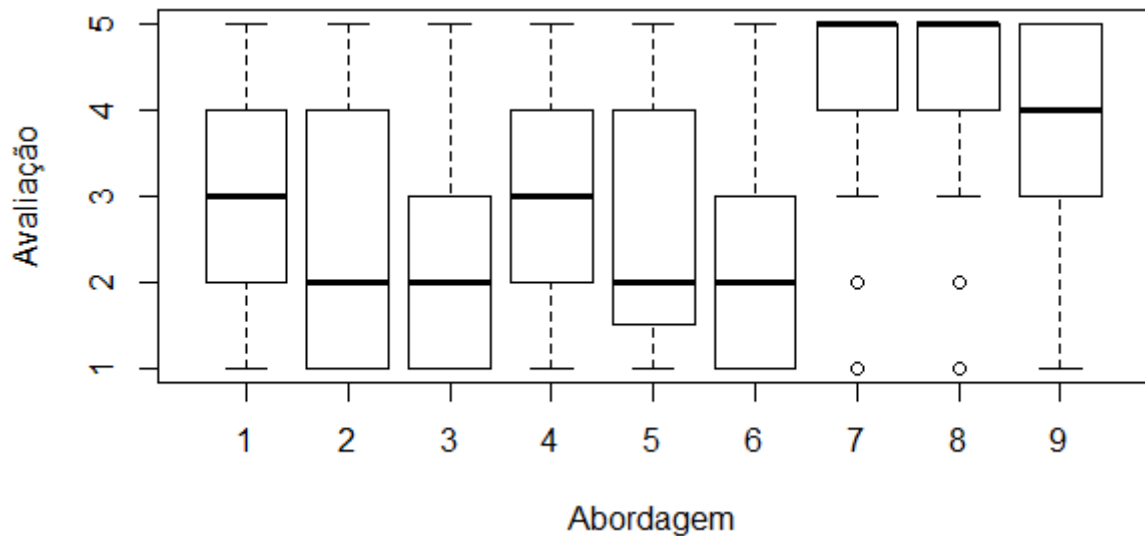
Fonte: Autoria própria.

Figura 63 - Boxplot dos votos dos participantes para o critério 5.



Fonte: Autoria própria.

Figura 64 - Boxplot dos votos dos participantes para o critério 6.



Fonte: Autoria própria.

5.6 Resultados da Avaliação Comparativa entre Abordagens

Os dados obtidos foram processados seguindo a mesma abordagem adotada por Arruda (2014). Destaca-se que os parâmetros utilizados nas abordagens concorrentes seguiram as recomendações dos autores, eliminando-se assim a necessidade de calibração dos parâmetros dos concorrentes.

Primeiramente, foi aplicado o teste de Shapiro-Wilk para cada um dos critérios no intuito de verificar a normalidade dos dados. O *p-value* para cada critério foi menor que $2,2 \times 10^{-16}$, sendo assim, nenhuma das distribuições dos dados de cada um dos critérios apresenta distribuição normal, considerando-se um grau de confiança de 99%.

A fim de elucidar a existência de possíveis diferenças estatísticas entre as abordagens, quando consideradas as 9 abordagens comparadas, foram conduzidos 6 testes Kruskal-Wallis (1 teste para cada critério, 9 abordagens comparadas). O *p-value* resultante de cada um dos testes, considerando um grau de confiança de 99%, foi menor que $2,2 \times 10^{-16}$. O que significa que **há diferenças** estatísticas significativas entre os valores associados aos critérios das abordagens.

O passo seguinte foi definir qual configuração de parâmetros da abordagem proposta obteve melhores resultados, assim sendo, foi aplicado o teste de Wilcoxon entre pares de configurações de parâmetros para detectar quais configurações tinham resultados estatisticamente diferentes umas das outras. Os *p-values* resultantes da comparação para cada critério podem ser observadas nas Tabelas 2 a 7.

Tabela 2 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 1.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,23944	-	-	-	-
Config. 3	0,00191	0,10309	-	-	-
Config. 4	0,46796	0,66461	0,02914	-	-
Config. 5	0,16587	0,89952	0,12252	0,59487	-
Config. 6	0,00895	0,23944	0,62390	0,09626	0,29853

Fonte: Autoria própria.

Tabela 3 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 2.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,0824	-	-	-	-
Config. 3	0,5863	0,2006	-	-	-
Config. 4	0,0765	0,9540	0,2025	-	-
Config. 5	0,1211	0,7280	0,3286	0,7479	-
Config. 6	0,2904	0,4079	0,6197	0,4302	0,6197

Fonte: Autoria própria.

Tabela 4 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 3.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,11071	-	-	-	-
Config. 3	0,00022	0,04088	-	-	-
Config. 4	0,08560	0,91797	0,04688	-	-
Config. 5	0,00509	0,26749	0,32672	0,28639	-
Config. 6	$1,5 \times 10^{-6}$	0,00165	0,26749	0,00180	0,03533

Fonte: Autoria própria.

Tabela 5 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 4.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,55798	-	-	-	-
Config. 3	0,01293	0,03850	-	-	-
Config. 4	0,47226	0,17901	0,00056	-	-
Config. 5	0,19831	0,46153	0,18076	0,03190	-
Config. 6	0,00962	0,03179	0,96522	0,00036	0,16534

Fonte: Autoria própria.

Tabela 6 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 5.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,06544	-	-	-	-
Config. 3	$7,4 \times 10^{-9}$	$1,7 \times 10^{-5}$	-	-	-
Config. 4	0,87831	0,03825	$7,0 \times 10^{-10}$	-	-
Config. 5	0,00031	0,05832	0,00897	$7,7 \times 10^{-5}$	-
Config. 6	$9,8 \times 10^{-9}$	$3,2 \times 10^{-5}$	0,75155	$8,6 \times 10^{-10}$	0,01785

Fonte: Autoria própria.

Tabela 7 - Teste para verificação de diferenças estatísticas entre as configurações da abordagem proposta para o critério 6.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config. 2	0,09819	-	-	-	-
Config. 3	$1,0 \times 10^{-4}$	0,01797	-	-	-
Config. 4	0,10407	0,00068	$1,2 \times 10^{-8}$	-	-
Config. 5	0,05919	0,81514	0,02877	0,00025	-
Config. 6	$3,9 \times 10^{-6}$	0,00138	0,36498	$1,7 \times 10^{-10}$	0,00233

Fonte: Autoria própria.

Os *p-values* em negrito indicam que **existe diferença** entre as configurações dos parâmetros da abordagem proposta, em relação ao respectivo critério, considerando um grau de confiança de 99%. Restava então, aplicar o teste de Wilcoxon, entre os pares de configurações dos parâmetros da abordagem proposta em que existe diferença, para

identificar qual configuração é a melhor.

Os resultados da aplicação do Wilcoxon para identificar a possível melhor configuração estão dispostos nas Tabelas 8 a 12.

Tabela 8 - Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 1.

X	Config. 1
Config. 3	0,000425
Config. 6	0,002112

Fonte: Autoria própria.

Tabela 9 - Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 3.

X	Config. 1	Config. 2	Config. 4
Config. 3	$5,84 \times 10^{-5}$	-	-
Config. 5	0,001627	-	-
Config. 6	$3,36 \times 10^{-7}$	0,000482	0,000549

Fonte: Autoria própria.

Tabela 10 - Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 4.

X	Config. 1	Config. 3	Config. 4
Config. 4	-	0,9998	-
Config. 6	0,003207	-	0,000106

Fonte: Autoria própria.

Tabela 11 - Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 5.

X	Config. 1	Config. 2	Config. 3	Config. 4
Config. 3	$2,15 \times 10^{-9}$	$5,61 \times 10^{-6}$	-	-
Config. 4	-	-	1	-
Config. 5	0,000116	-	0,9964	$2,76 \times 10^{-5}$
Config. 6	$3,00 \times 10^{-9}$	$1,1 \times 10^{-5}$	-	$2,37 \times 10^{-10}$

Fonte: Autoria própria.

Tabela 12 - Teste para verificação da mais bem avaliada entre as configurações da abordagem proposta para o critério 6.

X	Config. 1	Config. 2	Config. 3	Config. 4	Config. 5
Config.3	$3,45 \times 10^{-5}$	-	-	-	-
Config. 4	-	0,9997	1	-	-
Config. 5	-	-	-	$8,94 \times 10^{-5}$	-
Config. 6	$1,20 \times 10^{-6}$	0,000536	-	$4,69 \times 10^{-11}$	0,000938

Fonte: Autoria própria.

Novamente, considerando um grau de confiança de 99%, observa-se que apesar de algumas configurações se mostrarem mais bem avaliadas que outras em alguns critérios, **não há uma configuração que tenha sido mais bem avaliada na maioria dos critérios**. Dada a similaridade das configurações, será escolhida a **configuração de parâmetros 1**, pois em 5 dos critérios ela foi mais bem avaliada do que alguma outra configuração.

Segue-se, então, a comparação da abordagem proposta com a configuração de parâmetros 1 e as abordagens concorrentes, utilizando-se o teste de Wilcoxon, entre os pares de abordagens, para identificar a existência de diferenças estatísticas entre os valores associados aos critérios. Os *p-values* resultantes da comparação para cada critério podem ser observadas nas Tabelas 13 a 18.

Tabela 13 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 1.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	$6,4 \times 10^{-14}$	0,00049	-
Abordagem 9	0,52284	2×10^{-16}	$1,2 \times 10^{-10}$

Fonte: Autoria própria.

Tabela 14 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 2.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	2×10^{-16}	0,0539	-
Abordagem 9	2×10^{-16}	$1,6 \times 10^{-9}$	0,0001

Fonte: Autoria própria.

Tabela 15 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 3.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	$7,4 \times 10^{-9}$	-	-
Abordagem 8	$2,8 \times 10^{-7}$	0,43737	-
Abordagem 9	0,00625	0,00091	0,01014

Fonte: Autoria própria.

Tabela 16 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 4.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	2×10^{-16}	0,00036	-
Abordagem 9	$1,3 \times 10^{-15}$	$2,6 \times 10^{-13}$	0,0001

Fonte: Autoria própria.

Tabela 17 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 5.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	2×10^{-16}	0,00342	-
Abordagem 9	2×10^{-16}	$1,9 \times 10^{-7}$	0,01785

Fonte: Autoria própria.

Tabela 18 - Teste para verificação de diferenças estatísticas entre as abordagens para o critério 6.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	2×10^{-16}	$4,4 \times 10^{-6}$	-
Abordagem 9	2×10^{-16}	2×10^{-16}	$4,9 \times 10^{-6}$

Fonte: Autoria própria.

Os *p-values* em negrito indicam que, estatisticamente, **existem diferenças** entre as abordagens. Segue-se então, a aplicação do teste de Wilcoxon, entre pares de abordagens que apresentaram diferenças, para determinar qual delas foi mais bem avaliada, considerando-se 99% de grau de confiança.

Na Tabela 19, estão contidos os *p-values* da comparação entre abordagens para o critério 1.

Tabela 19 - Teste para verificação da mais bem avaliada entre as abordagens para o critério 1.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	2×10^{-16}	-	-
Abordagem 8	$7,165 \times 10^{-15}$	0,9999	-
Abordagem 9	-	1	1

Fonte: Autoria própria.

O teste indica que, para o critério 1, as abordagens 1 e 9 são as mais bem avaliadas estatisticamente, seguidas pela abordagem 8 e por último a abordagem 7. Na Tabela 20, estão contidos os *p-values* da comparação entre abordagens para o critério 2.

Tabela 20 - Teste para verificação da mais bem avaliada entre as abordagem para o critério 2.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	1	-	-
Abordagem 8	1	-	-
Abordagem 9	1	$4,329 \times 10^{-10}$	$2,876 \times 10^{-5}$

Fonte: Autoria própria.

O teste indica que, para o critério 2, as abordagens 7 e 8 são mais bem avaliadas estatisticamente, seguidas da abordagem 9 e por último a abordagem 1. Na Tabela 21, estão contidos os *p-values* da comparação entre abordagens para o critério 3.

Tabela 21 - Teste para verificação da mais bem avaliada entre as abordagem para o critério 3.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	1	-	-
Abordagem 8	1	-	-
Abordagem 9	0,9979	0,0002537	-

Fonte: Autoria própria.

O teste indica que, para o critério 3, as abordagens 7, 8 e 9 são mais bem avaliadas estatisticamente que a abordagem 1. Na Tabela 22, estão contidos os *p-values* da comparação entre abordagens para o critério 4.

Tabela 22 - Teste para verificação da mais bem avaliada entre as abordagem para o critério 4.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	1	-	-
Abordagem 8	1	0,000109	-
Abordagem 9	1	$6,867 \times 10^{-14}$	$2,855 \times 10^{-5}$

Fonte: Autoria própria.

O teste indica que, para o critério 4, a abordagem 7 é a mais bem avaliada estatisticamente, seguida pela abordagem 8, por penúltima a abordagem 9 e por fim a abordagem 1. Na Tabela 23, estão contidos os *p-values* da comparação entre abordagens para o critério 5.

Tabela 23 - Teste para verificação da mais bem avaliada entre as abordagens para o critério 5.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	1	-	-
Abordagem 8	1	0,00133	-
Abordagem 9	1	$6,021 \times 10^{-8}$	-

Fonte: Autoria própria.

O teste indica que, para o critério 5, a abordagem 7 é a mais bem avaliada estatisticamente, seguida pelas abordagens 8 e 9 e em último a abordagem 1. Na Tabela 24, estão contidos os *p-values* da comparação entre abordagens para o critério 6.

Tabela 24 - Teste para verificação da mais bem avaliada entre as abordagens para o critério 6.

X	Abordagem 1	Abordagem 7	Abordagem 8
Abordagem 7	1	-	-
Abordagem 8	1	$1,399 \times 10^{-6}$	-
Abordagem 9	1	$2,2 \times 10^{-16}$	$1,622 \times 10^{-6}$

Fonte: Autoria própria.

O teste indica que, para o critério 6, a abordagem 7 é a mais bem avaliada estatisticamente, seguida pela abordagem 8, em penúltimo a 9 e, por fim, a abordagem 1.

Ao observar os resultados supramencionados, considerando-se um grau de confiança de 99%, a abordagem proposta e a abordagem 9 obtiveram os melhores resultados para o critério 1. Em relação aos demais critérios, a abordagem proposta obteve os piores resultados em comparação com as abordagens concorrentes. Ao comparar as abordagens, é possível concluir que há maior preferência visual pelo processo de renderização das abordagens concorrentes, as quais são fundamentadas em abstração.

5.7 Detalhes de Implementação

A máquina utilizada para a geração dos vídeos referentes à abordagem proposta possuía o sistema operacional Linux, processador Intel i5, 8Gb de memória RAM e um disco rígido de 565Gb.

As abordagens concorrentes tiveram seus vídeos gerados em uma máquina com sistema operacional Windows, processador Intel i7, 8Gb de memória RAM e um disco rígido de 1Tb.

Foram processados 10 quadros de um vídeo utilizando cada uma das abordagens nas máquinas supracitadas e o tempo de execução de cada abordagem pode ser observado na Tabela 25.

Uma vez que o propósito inicial desta pesquisa não estava focado na eficiência computacional da abordagem, é importante salientar que o tempo de processamento exibido na Tabela 25, referente à abordagem proposta, ainda pode ser reduzido em função de otimizações que venham a ser feitas.

Tabela 25 - Tempo de processamento de 10 quadros de um vídeo.

X	Proposta	Kuwahara	Coerente	XDoG
Tempo (s)	488	10	8	15

Fonte: Autoria própria.

5.8 Considerações Finais do Capítulo

Neste capítulo, foram apresentados os resultados da investigação experimental relacionada à pesquisa desenvolvida, cujo propósito foi a validação da abordagem proposta no Capítulo 4. Após a realização dos experimentos, pode-se constatar que:

1. Os diferentes valores dos parâmetros de configuração da abordagem proposta não exerceram influência suficiente para que ele se destacasse em relação às demais. A configuração 1 de parâmetros foi a escolhida para representar a abordagem proposta, na comparação com as abordagens concorrentes, pois

foi a mais bem avaliada na maioria dos critérios; e

2. O experimento de validação, realizado com participantes humanos, evidenciou que a abordagem proposta resultou em resultados mais bem avaliados relacionados ao critério 1. Entretanto, a abordagem foi preterida pelos participantes nos demais critérios.

Resultados referentes à abordagem proposta, inclusive os dados completos tabulados, podem ser visualizados em <https://github.com/albertofagner/Cartoon>.

Capítulo 6

Considerações Finais

O presente capítulo contém considerações gerais desta dissertação. Na Seção 6.1, são retomadas hipóteses formuladas na Seção 1.4 (**Hipóteses**), às quais são analisadas, à luz dos resultados dos testes estatísticos da Seção 5.6 (**Resultados da Avaliação Comparativa entre Abordagens**), para a formulação das conclusões. Investigações futuras são sugeridas na Seção 6.2.

6.1 Conclusões e Contribuições

A pesquisa realizada teve como objetivo a proposição e validação de uma abordagem totalmente automática para a geração de vídeos digitais ao estilo de cartuns. A validação foi realizada com o desenvolvimento de um protótipo, no qual foram implementados todos os 8 módulos descritos no Capítulo 4 (**Abordagem Proposta**).

Abordagens fundamentadas em segmentação espaço-temporal de quadros, no tocante à geração de vídeos NPR, afiguram-se como um tópico de pesquisa desafiador, conforme corroborado por Arruda (2014). Este segmento de pesquisa é um problema ainda não solucionado para o caso geral e os resultados demonstrados no Capítulo 5 (**Validação da Abordagem Proposta**), reforçam as afirmações supramencionadas.

Ao serem retomadas as hipóteses apresentadas na Seção 1.4, é possível finalmente testá-las utilizando ferramental estatístico, analisando os votos dos participantes.

As hipóteses apresentadas na seção supracitada são confrontadas

com a análise dos resultados obtidos:

1. A abordagem proposta não gera resultados que são mais bem avaliados por um conjunto de participantes, em relação às abordagens concorrentes;

Hipótese **aceita**. Considerando a preferência dos participantes, com 99% de confiança, à luz dos testes de Wilcoxon realizados (vide Tabelas 18 a 23). Visto o insucesso em 5 dos 6 critérios utilizados.

2. Os resultados gerados pela abordagem proposta não contêm artefatos espúrios ou cintilação, quando comparados àqueles produzidos por abordagens semelhantes;

Hipótese **rejeitada**. Em vista das opiniões dos participantes referentes aos critérios 2 e 5, com 99% de confiança, os testes de Wilcoxon aplicados resultam na rejeição desta hipótese (vide Tabelas 19 e 22).

3. A abordagem proposta não produz resultados que apresentam melhor definição de bordas e de estilização, quando comparada a abordagens semelhantes;

Hipótese **aceita**. Considerando a prevalência da votação dos critérios 1, 3, 4 e 6 (vide Tabelas 18, 20, 21 e 23), com 99% de confiança, à luz dos testes de Wilcoxon realizados.

Sendo assim, a partir da proposição da abordagem proposta fundamentada na segmentação de cores, da coleta de dados do experimento relatado e da análise dos dados utilizando as ferramentas estatísticas. Conclui-se que a abordagem proposta nesta dissertação mostrou-se ineficaz no processo de renderização de vídeos não fotorrealísticos digitais, ao estilo de cartuns, em detrimento as abordagens disponíveis.

A abordagem para a geração de versões NPR de vídeos ao estilo de cartuns, ilustrada no Capítulo 4 (**Abordagem Proposta**), afigura-se como

uma contribuição desta dissertação. A escassez de pesquisas envolvendo segmentação na geração de representações NPR, conforme discutido no Capítulo 2 (**Revisão Bibliográfica**), reforça a importância da pesquisa desenvolvida.

Adicionalmente, podem ser elencadas outras contribuições:

- O uso de quantização de cores empregada na geração de vídeos digitais ao estilo de cartuns.
- A validação estatística realizada. Em geral, os pesquisadores utilizam amostras reduzidas (menos de 30 participantes) ou nem utilizam processamento estatístico para a avaliação de suas abordagens (e.g., WANG *et al.*, 2013; KANG *et al.*, 2015). Neste aspecto, a literatura da área considera uma grande amostra ao conter 30 ou mais unidades amostrais, conforme Costa (2017), Gravetter e Wallnau (2016), Hogg *et al.* (2015) e Martin e Bridgmon (2012). Oposto as pesquisas iniciais supramencionadas, Arruda (2014) utilizou uma grande amostra de 30 participantes. Esta pesquisa, teve o emprego de **60** participantes, o que permitiu maior rigor durante a aplicação dos testes estatísticos.

6.2 Proposições para Trabalhos Futuros

Apesar das contribuições citadas na Seção 6.1, verifica-se a necessidade de realizar investigações futuras, a fim de mitigar os problemas encontrados durante esta pesquisa e melhorar o desempenho da abordagens nos critérios, aos quais as abordagens concorrentes obtiveram melhores resultados.

A adição de um mecanismo de controle de cintilação mais eficiente para reduzi-la ao mesmo ou menor nível que o apresentado pelas soluções existentes. A manutenção da coerência espaço-temporal, discutida no Capítulo 1, configura-se com um dos principais desafios na utilização da segmentação.

A experimentação de outras maneiras de realizar a segmentação, substituindo a quantização de cores por outras formas, por exemplo, a segmentação de objetos na cena.

A utilização de outros detectores de bordas diferentes de Sobel que permitam a geração de linhas mais suaves, mas que preservem as curvas e contornos.

Por fim, propõe-se o estudo de critérios que permitam a avaliação de processos de renderização de imagens ao estilo de cartuns modernos, i.e., aqueles que apresentam um número maior de cores e detalhes, além da ausência de contornos, conforme ilustrado na Figura 65.

Figura 65 - Exemplo de cartuns modernos.



Fonte: <https://br.pinterest.com/chasesmiley21/comic/>.

Referências Bibliográficas

Alencar, J. D.; Queiroz, J. D.; Gomes, H. An approach for non-photorealistic rendering that is appealing to human viewers. In: 26th SIBGRAPI - Conference on Graphics, Patterns and Images. **Proceedings**. [S.l.: s.n.], 2013. p. 242-249.

Alencar Jr., J.B. de O. **Renderização não fotorrealística: Uma abordagem para abstração de imagens coloridas digitais**. 2010. 135 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Campina Grande, 2010.

Amarnadh, Mada; Hussain, S. Asif; Raju, M. Janardhana. A Review of Video Segmentation Techniques. *Global Journal of Advanced Engineering Technologies*. **Proceedings**. 2014.

Arruda, F. **Uma Abordagem Híbrida para a Renderização Não Fotorrealística de Vídeos Digitais**. 2014. 234 f. Tese (Doutorado em Ciência da Computação) - Universidade Federal de Campina Grande, Campina Grande, 2014.

Baugh, Gary; Kokaram, Anil. Semi-automatic motion based segmentation using long term motion trajectories. In: IMAGE PROCESSING (ICIP), 2010 17TH IEEE INTERNATIONAL CONFERENCE ON. **Proceedings**. IEEE, 2010. p. 3009-3012.

Ben, Nir; Zelnik, Lih. Approximate nearest neighbor fields in video. In: IEEE conference on computer vision and pattern recognition. **Proceedings**. 2015. p. 5233-5242.

Benard, Pierre; Bousseau, Adrien; Thollot, Joelle. State-of-the-Art report on temporal coherence for stylized animations. **Computer graphics forum**. Blackwell Publishing Ltd, 2011. p. 2367-2386.

Bénard, Pierre et al. Active strokes: Coherent line stylization for animated 3d models. In: Symposium on non-photorealistic animation and rendering. **Proceedings**. Eurographics Association, 2012. p. 37-46.

Bénard, Pierre; Thollot, Joëlle; Collomosse, John. Temporally coherent video stylization. In: Rosin, Paul; Collomosse, John. **Image and video-based artistic stylisation**. London: Springer-Verlag, 2012. cap. 13.

Borawski, Mariusz. Non-photorealistic rendering with the use of short segments of straight lines in a vector space of increments. **Control&Cybernetics**, v. 43, n. 1, p. 161-178, 2014.

Cao, Chen et al. Automatic motion-guided video stylization and personalization. In: The 19th ACM international conference on multimedia. **Proceedings**. ACM, 2011. p. 1041-1044.

Chen, Jiajian; Turk, Greg; Macintyre, Blair. A non-photorealistic rendering framework with temporal coherence for augmented reality. In: MIXED AND AUGMENTED REALITY (ISMAR), 2012 IEEE INTERNATIONAL SYMPOSIUM ON. **Proceedings**. IEEE, 2012. p. 151-160.

Chen, Jiajian; Turk, Greg; MacIntyre, Blair. Painterly rendering with coherence for augmented reality. In: VR INNOVATION (ISVRI), 2011 IEEE INTERNATIONAL SYMPOSIUM ON. **Proceedings**. IEEE, 2011. p. 103-110.

Costa, D. L. **Avaliação semântica do design de interfaces humano-computador em jogos digitais fundamentada em uma abordagem multimétodos**. 2017. 221 f. Dissertação (Mestrado em Design), Programa de Pós-Graduação em Design, Universidade Federal de Campina Grande, Campina Grande. 2017.

Farneback, Gunnar. Two-frame motion estimation based on polynomial expansion. In: **Lecture Notes in Computer Science**, 2003. p. 363-370.

Gangopadhyay, Aalok; Pachori, Shubham; Raman, Shanmugananthan. Automatic silhouette photography. In: Communication (NCC), 2016 Twenty Second National Conference on. **Proceedings**. IEEE, 2016. p. 1-6.

Gravetter, F.; Wallnau, L. **Statistics for the behavioral sciences**. Belmont: Cengage Learning, 2016.

Hassan, Eman T. et al. Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos. In: **CVPR Workshops**. 2017. p. 1333-1342.

Hogg, R. V.; Tanis, E. A. **Probability and statistical inference**. 4. ed. [S.l.]: MacMillan, 1993.

Hogg, R. V.; Tanis, E. A.; Zimmerman, D. L. **Probability and statistical inference**. 9. ed. Upper Saddle River: Pearson Education, Inc., 2015.

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 - Volume 2. **Proceedings**. Washington, DC, USA: IEEE Computer Society, 2003. (ICME '03), p. 449-452.

Jiang, Hanqing et al. Spatio-temporal video segmentation of static scenes and its applications. IEEE Transactions on Multimedia. **Proceedings**. v. 17, n. 1, p. 3-15, 2015.

Kang, Dongwann et al. Directional texture transfer for video. **Multimedia tools and applications**, v. 74, n. 1, p. 245-258, 2015.

Klette, Reinhard. Image segmentation. In: ____. **Concise computer vision**. London: Springer-Verlag, 2014. cap. 5.

Kolivand, Hoshang et al. Photorealistic rendering: a survey on evaluation. **Multimedia Tools and Applications**, p. 1-26, 2018.

Kruskal, William H.; Wallis, W. Allen. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association. **Proceedings**. v. 47, n. 260, p. 583-621, 1952.

Kung, Sun-Yuan; Tin, Yun-Ting; Chen, Yen-Kuang. Motion-based segmentation by principal singular vector (PSV) clustering method. In: Acoustics, speech, and signal processing, 1996. ICASSP-96. 1996 IEEE international conference on. **Proceedings**. IEEE, 1996. p. 3410-3413.

Kyprianidis, J. E.; Kang, H.; Döllner, J. Image and Video Abstraction by Anisotropic Kuwahara Filtering. Computer Graphics Forum. **Proceedings**. 2009.

Kyprianidis, J. E.; Kang, H. Image and Video Abstraction by Coherence-Enhancing Filtering. Computer Graphics Forum. **Proceedings**. 2011. p. 593-602.

Lawrence, C. Zitnick; Dollár, Piotr. Edge boxes: Locating object proposals from edges. **ECCV**. 2014.

Lin, Tao; Lin, Liang; Wang, Qing. Robust stroke-based video animation via layered motion and correspondence. In: The 20th ACM international conference on multimedia. **Proceedings**. ACM, 2012. p. 729-732.

Lu, Li-wen et al. A non-photorealistic rendering algorithm for cartoons. In: IMAGE AND SIGNAL PROCESSING (CISP), 2013 6TH INTERNATIONAL CONGRESS ON. **Proceedings**. IEEE, 2013. p. 680-685.

Lu, Ping et al. Image-based non-photorealistic rendering for realtime virtual sculpting. **Multimedia tools and applications**, v. 74, n. 21, p. 9697-9714, 2015.

Martin, W. E.; Bridgmon, K. D. **Quantitative and statistical research methods: From hypothesis to results**. Indianapolis: Wiley Publishing, Inc., 2012.

Medhi, I.; Patabandhige, C. R.; Toyama, K. Cartoon generation for text-free user interfaces. In: International Workshop on Computer Vision

Applications for Developing Regions. **Proceedings**. [S.l.: s.n.], 2007, p.1-9.

Rosin, Paul L.; Lai, Yu-Kun. Non-photorealistic rendering with spot colour. In: The symposium on computational aesthetics. **Proceedings**. ACM, 2013. p. 67-75.

Song, Jie; Xu, Dan. Abstraction for static and dynamic images based on visual saliency. In: DIGITAL HOME (ICDH), 2012 FOURTH INTERNATIONAL CONFERENCE ON. **Proceedings**. IEEE, 2012. p. 57-62.

Su, Zhuo; Luo, Xiaonan; Artusi, Alessandro. A novel image decomposition approach and its applications. **The visual computer**, v. 29, n. 10, p. 1011-1023, 2013.

Wang, Qiaoyu et al. An adaptive cartoon-like stylization for color video in real time. **Multimedia Tools and Applications**. v. 76, n. 15, p. 16767-16782, 2017.

Wang, ShanDong et al. Coherence-enhancing line drawing for color images. **Science china information sciences**, v. 56, n. 11, p. 1-11, 2013.

Wenhua, Qian et al. Image abstraction painting of flow-like stylization. **Tehničkivjesnik**, v. 22, n. 4, p. 837-844, 2015.

Wilcoxon, Frank. Individual comparisons by ranking methods. *Biometrics bulletin*. **Proceedings**. v. 1, n. 6, p. 80-83, 1945.

Winnemöller, H. XDoG: Advanced image stylization with eXtended Difference-of-Gaussians. Proc. Symposium on Non-Photorealistic Animation and Rendering (NPAR). **Proceedings**. 2012. p. 147-156.

Winnemöller, H.; Kyprianidis, J. E.; Olsen, S. C. XDoG: An eXtended difference-of-Gaussians compendium including advanced image stylization. **Computers & Graphics**, 36(6), pp. 740-753, 2012.

Xue, Ru et al. An extended flow-based difference-of-Gaussians method of line drawing for polyhedral image. **Optik-international journal for light and electron optics**, v. 125, n. 17, p. 4624-4628, 2014.

Yoon, Jong-Chul; Lee, In-Kwon; Kang, Henry. Video painting based on a stabilized time-varying flow field. **Visualization and computer graphics, IEEE Transactions on**, v. 18, n. 1, p. 58-67, 2012.

Zamboni, A. B. et al. StArt Uma Ferramenta Computacional de Apoio à Revisão Sistemática. In: BRAZILIAN CONFERENCE ON SOFTWARE:

THEORY AND PRACTICE - TOOLS SESSION. **Proceedings**. UFBA, 2010.

Zhang, Wei et al. Style transfer via image component analysis. **Multimedia, IEEE transactions on**, v. 15, n. 7, p. 1594-1601, 2013.

Apêndice A

Documentação Pré-Teste

Delineamento do Perfil do Participante

Avaliação da qualidade de cartoons gerados por
uma abordagem totalmente automática

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:

- 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 renderizações dos vídeos melhor avaliados durante a votação?

12 - Houve algum aspecto desagradável durante a votação? Em caso afirmativo, qual?



Apêndice B

Histórico de Alterações

As abordagens utilizadas preliminarmente como base para esta pesquisa continham módulos anteriores ao módulo de Segmentação de Cores. Entretanto, não se conseguiu obter resultados que justificassem a permanência dos referidos módulos, o que implicou sua remoção da abordagem proposta. Uma discussão sobre os módulos removidos é apresentada a seguir.

O módulo removido de Subtração de Plano de Fundo foi instanciado utilizando-se misturas de gaussianas (MOG2) (LAWRENCE, DOLLÁR, 2014). Neste módulo, era feita a subtração de plano de fundo para pares de quadros adjacentes. Um exemplo do processamento realizado por este módulo pode ser visto na Figura 66.

Figura 66 - Exemplo de Saída gerada pelo módulo de Subtração de Plano de Fundo.



Fonte: Autoria própria.

O módulo removido de Detecção de Fluxo Óptico foi instanciado utilizando-se o cálculo do fluxo óptico pelo método Farneback (FARNEBACK, 2003), que calcula o fluxo óptico piramidal para cada pixel presente em pares de quadros adjacentes. Na Figura 67, vê-se o resultado.

Figura 67 - Exemplo do cálculo do fluxo óptico Farneback de um quadro.



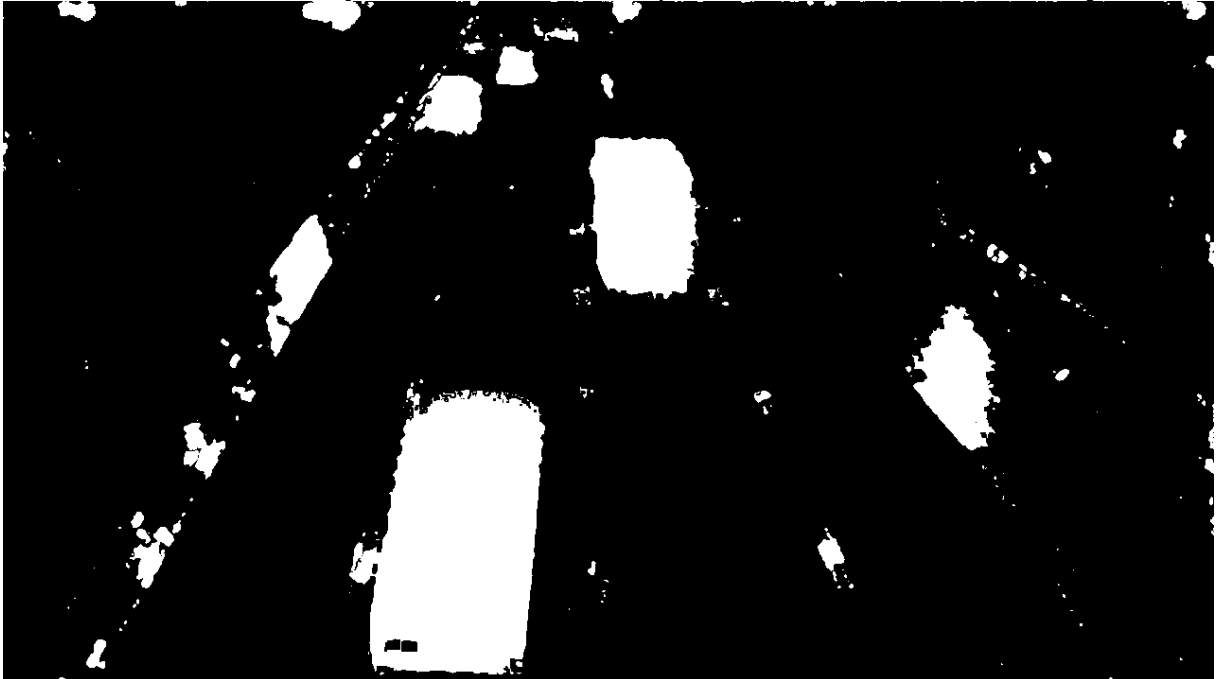
Fonte: Autoria própria.

Os parâmetros utilizados para o cálculo do fluxo óptico foram selecionados após experimentos utilizando-se uma referência (*ground truth*) marcada manualmente. Apesar dos melhores parâmetros terem sido selecionados, observava-se a presença de alto nível de ruído nos resultados.

O ruído existente no cálculo do fluxo óptico resultante era reduzido ao se calcular a mediana para cada pixel central de uma janela 3x3 e também a vizinhança correspondente ao cálculo do fluxo óptico Farneback posterior. Esta operação de filtragem espaço-temporal, reduzia parte do ruído resultante de variações de iluminação que ocorrem em vídeos. A morfologia matemática era aplicada em seguida, utilizando-se a operação de *close* para a redução de ruídos persistentes.

Os resultados das etapas supramencionadas podem ser observados, respectivamente, nas Figuras 68 e 69.

Figura 68 - Exemplo de combinação do fluxo óptico Farneback calculado em 3 quadros consecutivos, utilizando-se uma vizinhança 3x3.



Fonte: Autoria própria.

Figura 69 - Exemplo de aplicação da operação de morfologia matemática de fechamento em um quadro.

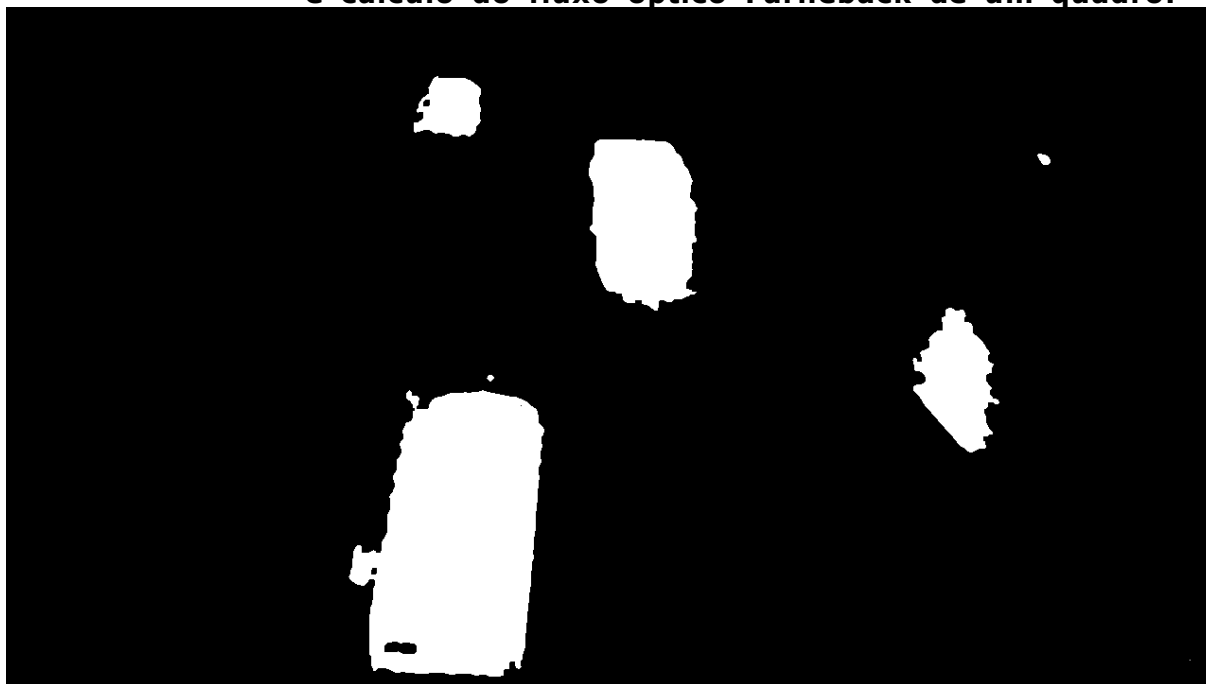


Fonte: Autoria própria.

O módulo removido de Remoção de Artefatos era resultante da combinação das saídas dos dois módulos supramencionados. Ambos

módulos tinham como saída uma máscara correspondente à detecção de movimento referente ao primeiro plano. Os pixels de movimento detectados pelo método *MOG2* eram inicialmente marcados na máscara gerada pelo cálculo do fluxo óptico. Em seguida, era realizada uma operação de preenchimento (*floodfill*) dos pixels, tendo-se como ponto de partida os pixels previamente marcados. Após o processo de preenchimento, a máscara de cálculo do fluxo óptico continha pixels marcados correspondentes ao *MOG2* e a expansão do preenchimento. Os demais pixels não marcados eram descartados, pois eram correspondentes a ruídos capturados durante o cálculo do fluxo óptico e assim os pixels marcados seriam a saída do módulo. O resultado do processo pode ser observado na Figura 70.

Figura 70 - Máscara resultante da combinação das máscaras de *MOG2* e cálculo do fluxo óptico Farneback de um quadro.



Fonte: Autoria própria.

A estilização do vídeo era feita utilizando-se o método FABL (ALENCAR JR., 2010). A partir da técnica *meanshift*, era feito o agrupamento de pixels similares. O passo seguinte era a aplicação do FABL para a suavização dos quadros com preservação das bordas, variando-se suas iterações de acordo com a máscara resultante da

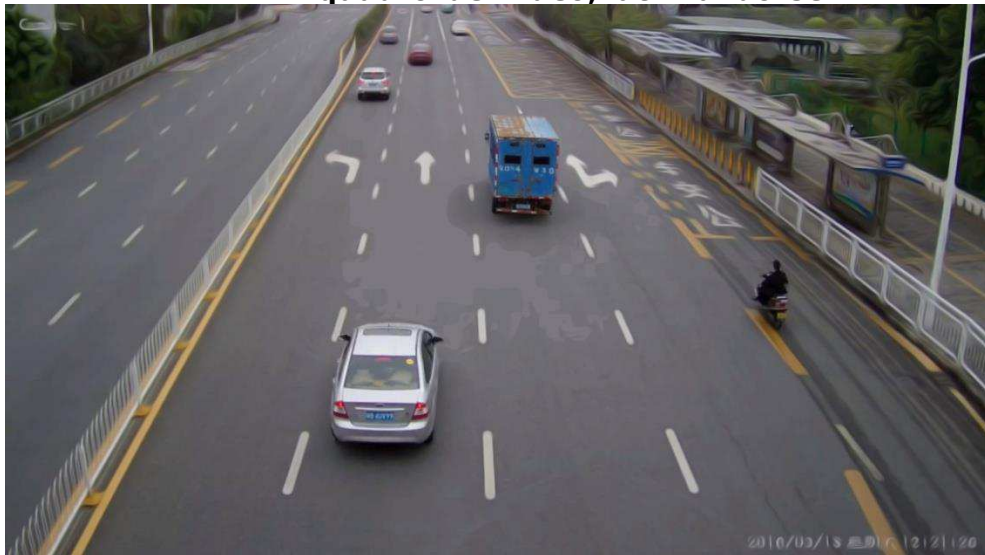
combinação entre o MOG2 e o fluxo óptico calculado. A variação das iterações do FABL resultava em alta homogeneização em pixels do plano de fundo e baixa homogeneização em pixels do primeiro plano (ver Figuras 71 e 72).

Figura 71 - Exemplo de um quadro de entrada.



Fonte: Autoria própria.

Figura 72 - Exemplo de aplicação de estilização a um quadro de vídeo, utilizando-se FABL.



Fonte: Autoria própria.

A combinação do *MOG2* com o fluxo óptico calculado eliminou parte dos ruídos. Entretanto, outro problema foi inserido. Partes dos objetos detectados desapareceram ou sofreram distorções que deixavam lacunas

nos objetos. Na tentativa de corrigir as lacunas dos objetos deixados pela combinação dos métodos, foi aplicado o preenchimento descrito anteriormente. O resultado final deste processo não era suficiente para que a redução de cores fosse aplicada de maneira consistente nos objetos presentes nos quadros. Os problemas supramencionados podem ser observados na Figura 73.

Figura 73 - Exemplos de lacunas nos objetos detectados em um quadro de vídeo processado.

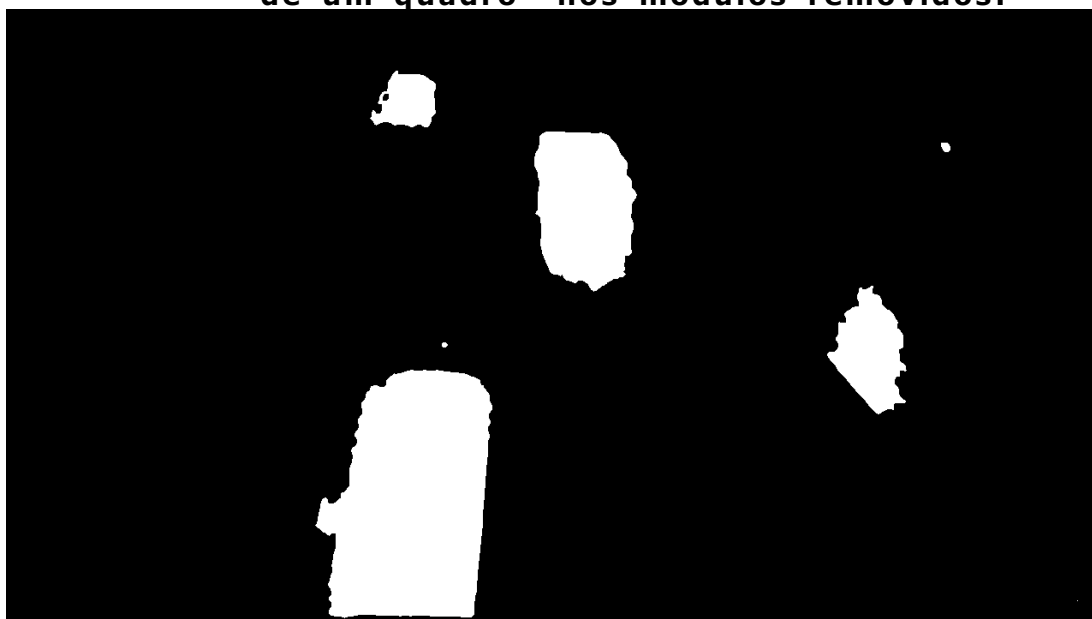


Fonte: Autoria própria.

Ao se aplicar o método FABL variante utilizando-se a máscara supramencionada como guia, foram gerados resultados insatisfatórios, pois as partes faltantes não detectadas distorciam o aspecto cartum final do vídeo. Outro problema advindo da limitação do efeito do FABL é falta de similaridade com o gênero cartum ao aplicá-lo (ver Figura 72), o que suscitou a alteração do uso do FABL como redutor de cores e seu emprego como módulo de pré-processamento.

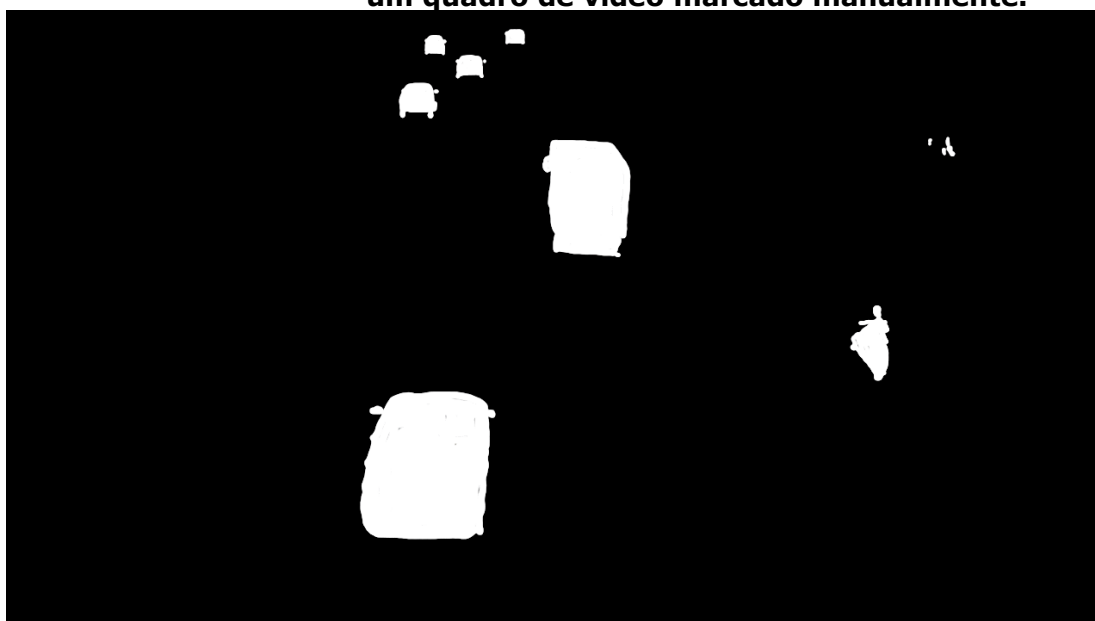
Nas Figuras 74 e 75, são ilustrados, respectivamente, a máscara resultante final do processamento realizado nos módulos removidos e o *ground truth* correspondente ao mesmo quadro. Ao se comparar as duas figuras, é possível constatar a diferença significativa existente entre elas.

Figura 74 - Exemplo de máscara resultante do processamento de um quadro nos módulos removidos.



Fonte: Autoria própria.

Figura 75 - Exemplo de *ground truth* correspondente a um quadro de vídeo marcado manualmente.



Fonte: Autoria própria.

Ao levar em consideração as questões discutidas anteriormente, decidiu-se remover estes módulos problemáticos e investir na análise de outras estratégias que se mostraram mais promissoras.

Apêndice C

Parâmetros Utilizados Pelas Abordagens

As abordagens concorrentes utilizadas nesta pesquisa tiveram seus parâmetros ajustados seguindo as recomendações dos seus autores, enquanto os parâmetros da abordagem desenvolvida nesta dissertação foram definidos experimentalmente. Elas são dispostas a seguir juntamente com os links referentes as implementações disponibilizadas pelos autores.

Quadro 3 - Parâmetros utilizados pelas abordagens.

Parâmetro	Anisotropic kuwahara filter	Coherence-enhancing filtering	XDoG Filter	Abordagem proposta
Σt	2,00	6,00	-	-
Σd	-	1,00	-	-
Σi	-	0,00	-	-
Σg	-	1,50	-	-
Σa	-	1,50	-	-
Σc	-	-	2,28	-
Σdg	-	-	3,00	-
Σdt	-	-	3,00	-
Σrg	-	-	4,25	-
Σrt	-	-	4,25	-

Σe	-	-	1,40	-
Σm	-	-	4,40	-
Σ	-	-	1,00	-
precision Σc	-	-	2,44775	-
precision m	-	-	2,00	-
step m	-	-	1,00	-
α	1,00	-	-	-
bf α	-	-	0,00	-
radius	6,00	-	-	-
smoothing	33,33%	-	-	-
q	8,00	-	-	-
N	-	5	-	-
T r	-	0,002	-	-
T s	-	0,005	-	-
max angle	-	22,50	-	-
r	-	2,00	-	-
st type	-	-	sobel-L	-
etf N	-	-	3	-
bilateral filter: type	-	-	xy	-
n e	-	-	1	-
n a	-	-	3	-
precision g	-	-	2,00	-
precision t	-	-	2,00	-
precision e	-	-	3,00	-

dog: type	-	-	Flow-based	-
dog k	-	-	1,60	-
adj func	-	-	smoothstep	-
ξ	-	-	3,50220	-
T	-	-	0,95595	-
ϕ	-	-	0,38590	-
dog fgauss	-	-	euler	-
quant type	-	-	adaptive	-
nbins	-	-	8	-
ϕq	-	-	2,00	-
$\lambda\Delta$	-	-	0,00	-
$\Omega\Delta$	-	-	2,00	-
$\lambda\phi$	-	-	0,90	-
$\Omega\phi$	-	-	1,60	-
final smooth: type	-	-	flow-nearest	-
Iterações FABL	-	-	-	3
Ajuste de saturação	-	-	-	30%
limiar	-	-	-	70

Fonte: Autoria própria.

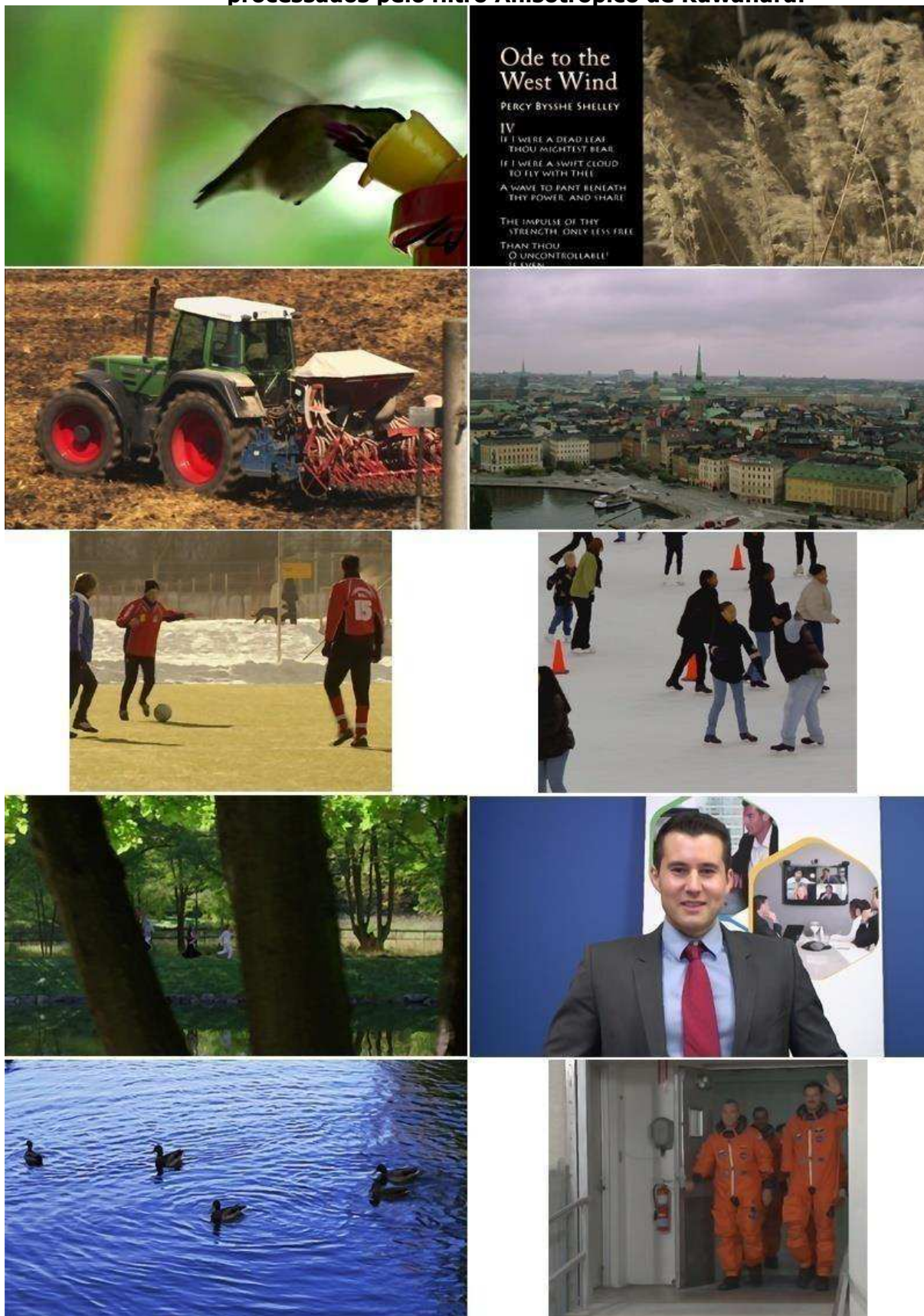
Anisotropic kuwahara filter: <https://code.google.com/archive/p/gpuakf/downloads>

Coherence-Enhancing Filtering: <https://code.google.com/archive/p/cefabs/downloads>

XDoG Filter: <https://code.google.com/archive/p/xdog-demo/downloads>

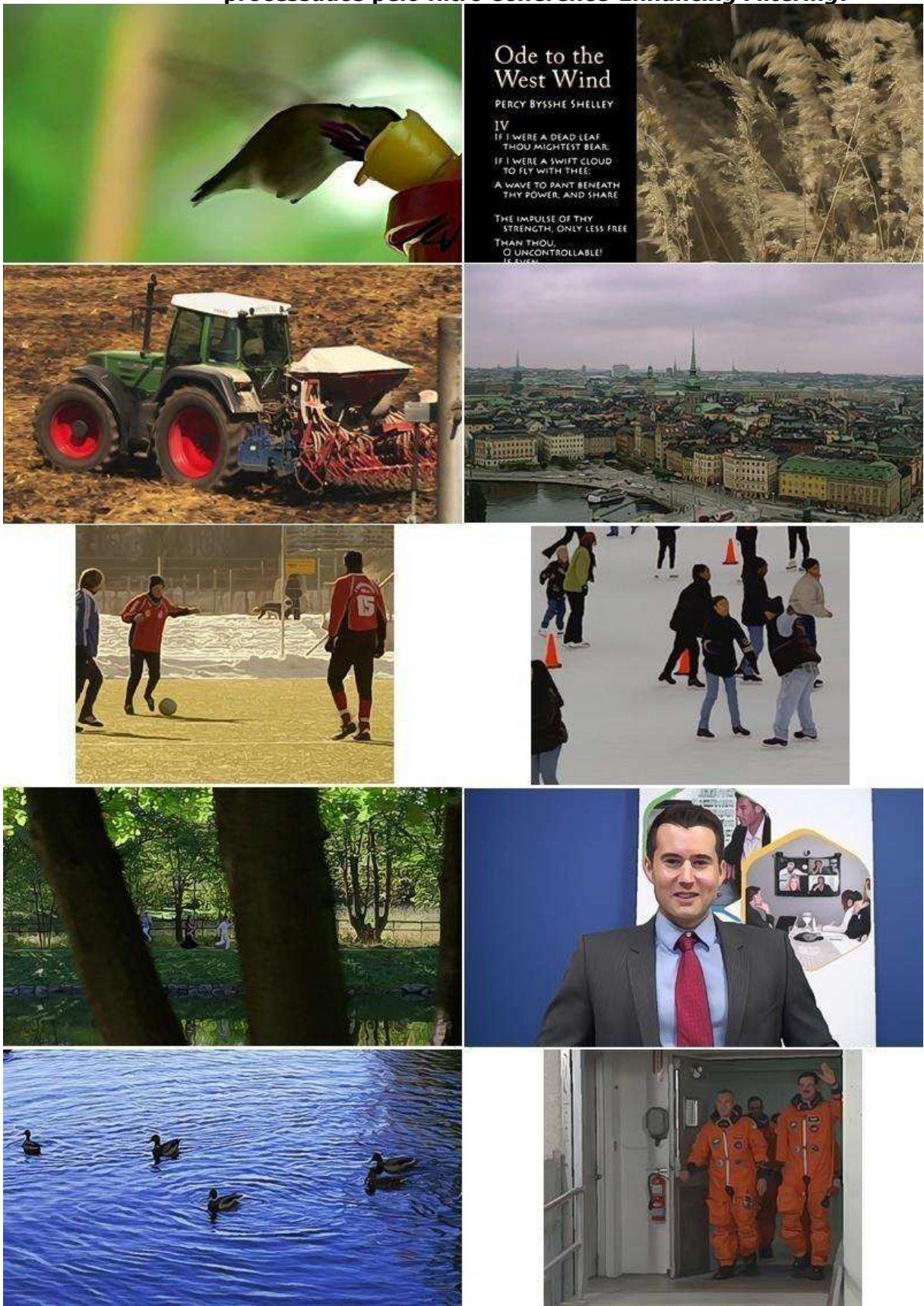
Abordagem Proposta: <https://github.com/albertofagner/Cartoon>

Figura 76 - Quadro inicial de cada um dos 10 vídeos selecionados processados pelo filtro Anisotrópico de Kuwahara.



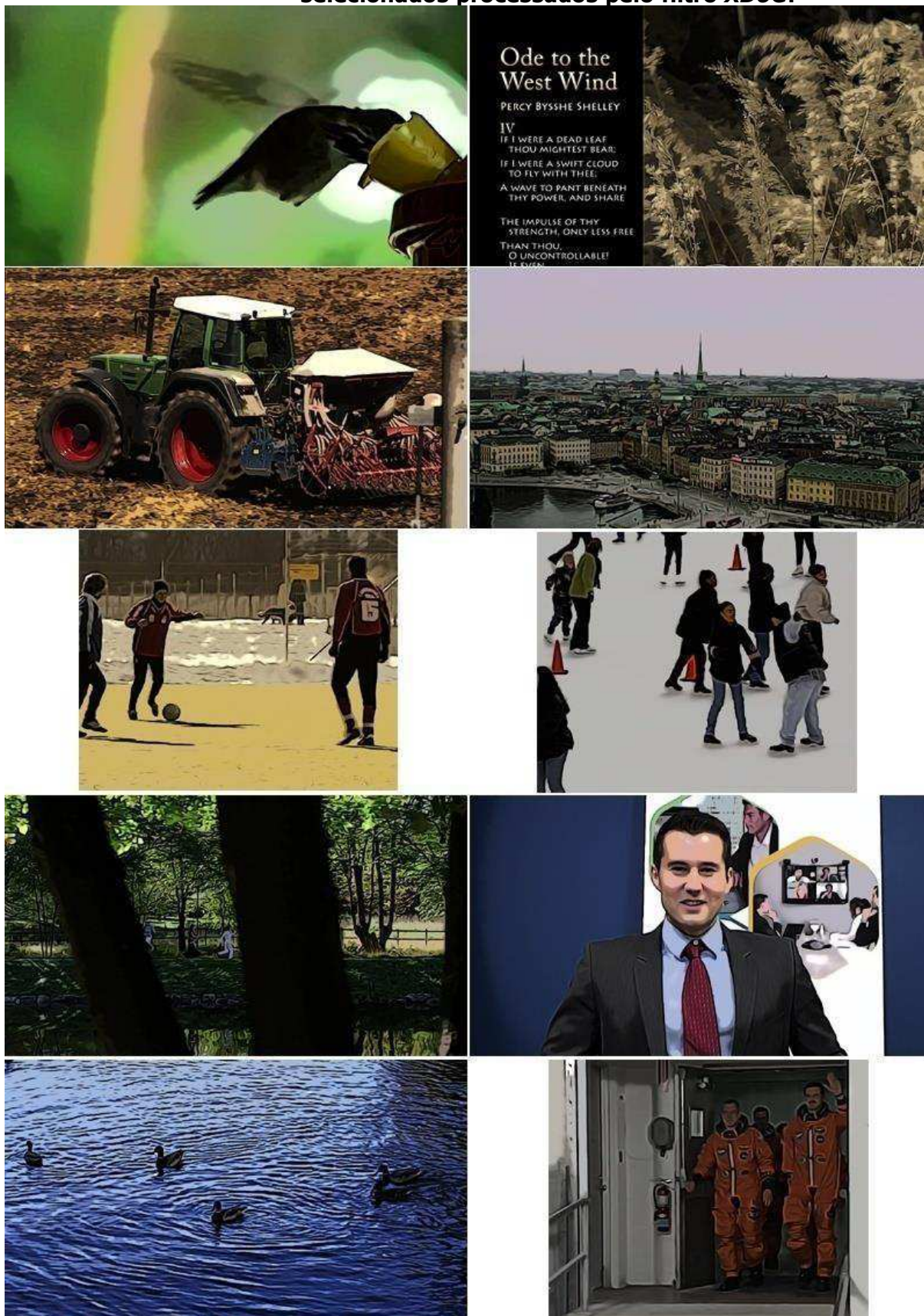
Fonte: Autoria própria.

Figura 77 - Quadro inicial de cada um dos 10 vídeos seleccionados processados pelo filtro *Coherence-Enhancing Filtering*.



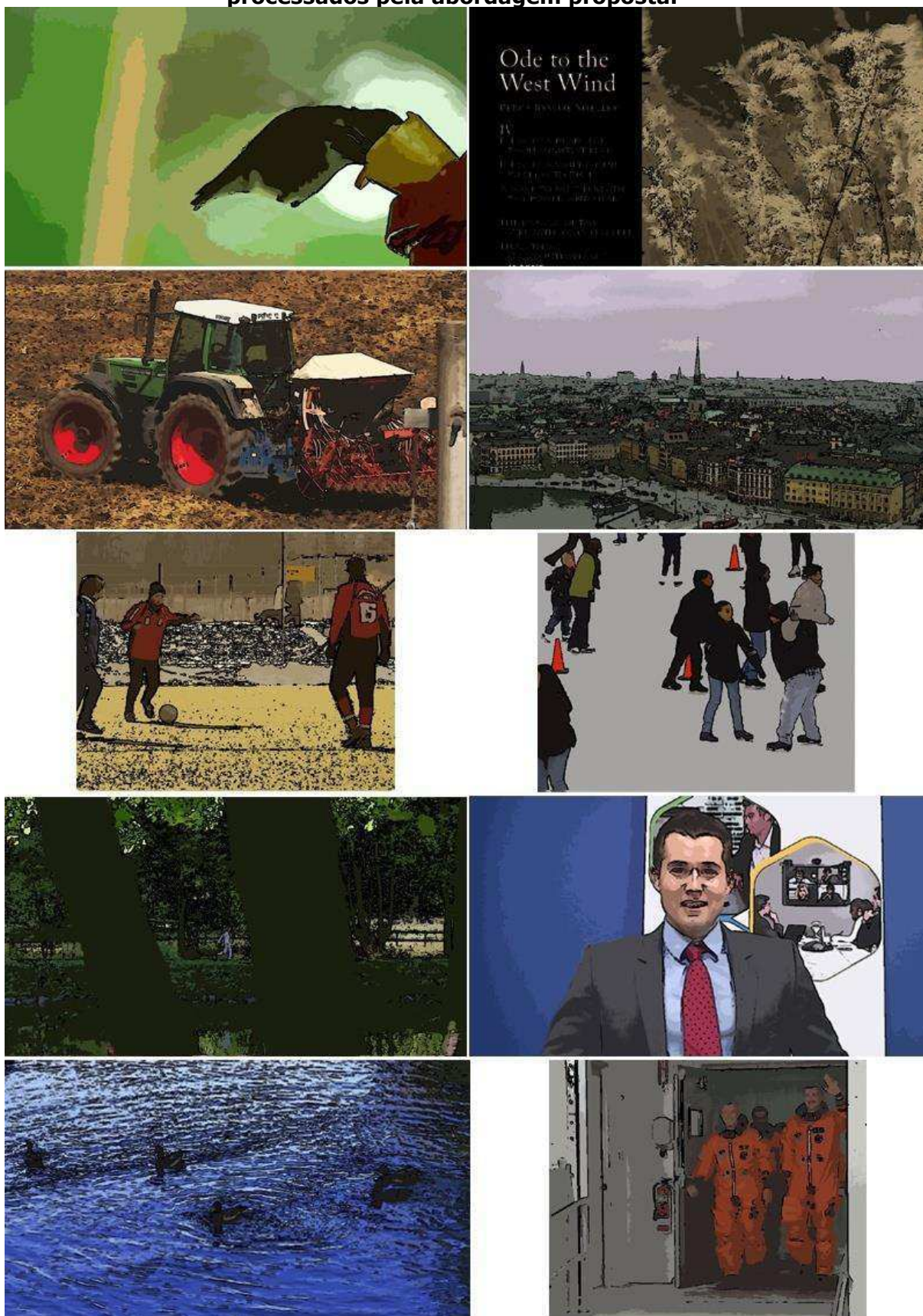
Fonte: Autoria própria.

Figura 78 - Quadro inicial de cada um dos 10 vídeos selecionados processados pelo filtro XDoG.



Fonte: Autoria própria.

Figura 79 - Quadro inicial de cada um dos 10 vídeos seleccionados processados pela abordagem proposta.



Fonte: Autoria própria.