

Métodos de abstração em vídeo e imagem

José Miguel Ferreira Azevedo Matos
Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Matemática
2022

Métodos de Abstração em vídeo e imagem

José Miguel Ferreira Azevedo Matos

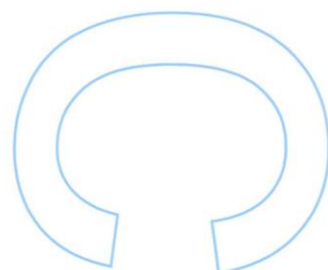
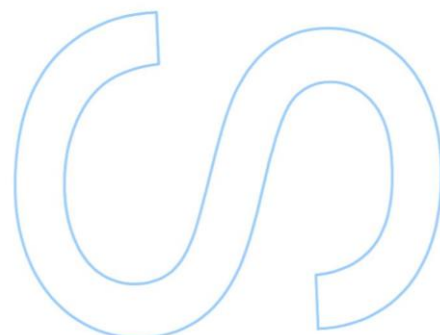
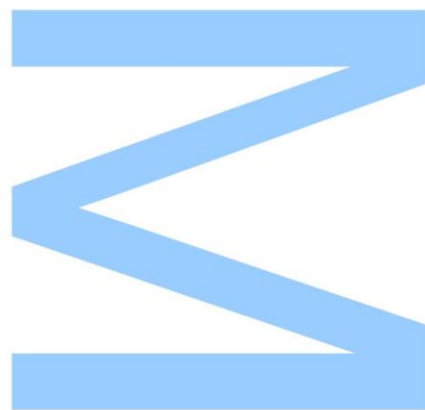
Mestrado em Engenharia Matemática

Departamento de Matemática

2022

Orientador

André Ribeiro da Silva de Almeida Marçal, Professor, Faculdade de Ciências da
Universidade do Porto

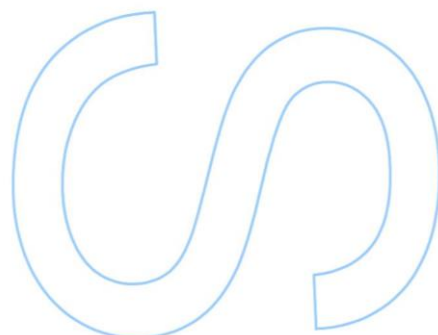
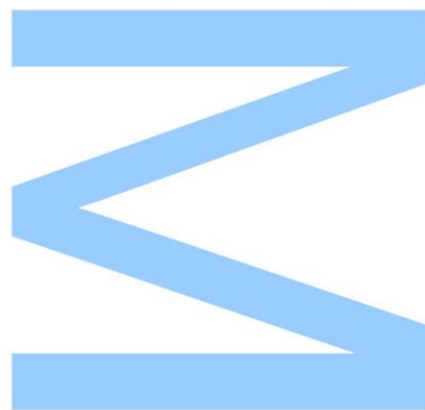




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



Declaração de Honra

Eu, José Miguel Ferreira Azevedo Matos, inscrito no Mestrado em Engenharia Matemática da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação/relatório de estágio/projeto [indicar o aplicável] reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

José Miguel Ferreira Azevedo Matos

Porto, 05/01/2023

Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais, irmãos, avós, tios e primos por me acompanharem ao longo de todo o percurso até aqui ao meu lado. De seguida, agradeço também ao meu orientador, o Professor André Marçal por toda a ajuda que me deu e paciência que teve ao longo deste trabalho. Agradeço ainda a todos os amigos, em especial ao Gringo que me mantiveram motivado e me deram muitas memórias que irei guardar e a todos os professores e colegas que me acompanharam ao longo destes anos.

Abstract

In this work, the topic of non-photorealistic rendering is approached. Designates the area of research and development focused on techniques for creating computer-generated images with a non-photorealistic appearance and arose from the need to create a distinction from more traditional computer graphics whose main objective is to obtain computer-generated images with a photorealistic appearance. There are very different techniques, therefore it is difficult to objectively evaluate each one beyond the execution time of the algorithm used.

The method approached was proposed by Holger Winnemoller et al. for being one of the most accurate and with the particularity of being able to process the images in real time in an attempt to increase the efficiency of the method in view of certain characteristics existing in the input image.

In the first part, there was a brief overview of several relevant image abstraction methods as well as their differences and approaches. Then, the method used was analyzed and the various results obtained were evaluated. A collection of videos and images with different characteristics was made, which will be covered by the method to analyze and compare the various results.

Among the possible existing evaluation methods, the only ones that are objective and factual, such as the rendering execution time are not sufficient to conclude the superior method. Therefore, most of them are supported by the subjective analysis of the final result. To evaluate the results, a subjective analysis was also carried out, taking into account details, contours, color and sharpness. A description was made for each video and image, in terms of characteristics such as color, movement and detail. The set of ideal values of the parameters used was presented, establishing the relationship between them and the characteristics of the videos and images as well as a comparison between the original video or image and the respective output after the abstraction process.

The method's image base warping (IBW) was not included in the method analysis because it wasn't possible to implement. An adaptation to the

method was made using a Sobel filter and a Gaussian blur, both mentioned in the IBW. The results are considered to be more satisfactory in relation to the implementation made with Difference of Gaussians (DoG).

After analyzing each video and image individually, it was possible to find a relationship between the value of each parameter of the method implementation and the characteristics of an image in a table. A low, medium and high value were defined based on the most important criteria to take into account for the abstraction process to each parameter. It was possible to find a coherent connection between criteria and values of each parameter, which facilitates the user of this method to get a more effective and personalized abstraction for each object.

Keywords: Image processing, Video processing, Image abstraction, Bilateral filter

Resumo

Neste trabalho é abordado o tema de renderização não foto-realista. Designa a área de investigação e desenvolvimento focada nas técnicas de criação de imagens geradas por computador com aspeto não-fotorrealista e surgiu pela necessidade de criar uma distinção da computação gráfica mais tradicional que tem como principal objetivo obter imagens geradas por computador com aspeto fotorrealista. Por haver técnicas muito diferentes, torna-se difícil a avaliação objectiva de cada uma além do tempo de execução do algoritmo utilizado.

Foi abordado o método proposto por Holger Winnemoller et al. [1] por ser um dos mais precisos e com a particularidade de conseguir processar as imagens em tempo real na tentativa de aumentar a eficiência do método perante certas características existentes na imagem do *input*.

Numa primeira parte, houve um breve enquadramento de vários métodos de abstração de imagem relevantes bem como as suas diferenças e abordagens. De seguida, foi analisado o método utilizado e avaliados os vários resultados obtidos. Foi feita uma recolha de vídeos e imagens com características diferentes que irão ser alvos do método para analisar e comparar os vários resultados.

Dentro dos possíveis métodos de avaliação existentes, os únicos que são objectivos e factuais como o tempo de execução da renderização não são suficientes para concluir o método superior por isso a maioria são suportados pela análise subjectiva do resultado final. Para a avaliação dos resultados, foi igualmente feita uma análise subjectiva, tendo em atenção os detalhes, contornos, cor e nitidez. Para cada vídeo e imagem foi feita uma descrição do mesmo, em termos de características tais como a cor, movimento e detalhe e foi apresentado o conjunto dos valores ideais dos parâmetros utilizado estabelecendo a relação entre ambos bem como uma comparação entre vídeo ou imagem original e o respetivo output após o processo de abstração.

A deformação de imagem opcional do método não foi incluída na análise do método por não ter sido possível fazer a sua implementação. Foi feita uma adaptação ao método através de um filtro de Sobel e um desfoque Gaussiano,

ambos mencionados na deformação de imagem. Os resultados demonstram ser mais satisfatórios em relação à implementação feita com Diferença de Gaussianos (DoG).

Após a análise a cada vídeo e imagem de forma individual foi possível encontrar uma relação entre o valor de cada parâmetro da implementação do método e as características de uma imagem numa tabela. Foi associado a cada parâmetro um valor baixo, médio, alto e os critérios mais preponderantes que se devem ter em conta para definir o valor de cada variável para o processo de abstração. Foi possível encontrar uma conexão coerente entre critérios e valores de cada parâmetro o que facilita ao utilizador deste método uma abstração personalizada e mais eficaz para cada objeto.

Palavras-chave: Processamento de Imagem; Processamento de Vídeo; Abstração de Imagem; Filtro Bilateral

Índice

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
2 Estado de arte	3
3 Real time video abstraction	7
3.1 Filtro Bilateral	8
3.2 Quantização de luminância	9
3.3 Contornos de diferenças de Gaussianas	10
3.4 Implementação do método	12
4 Avaliação experimental	13
4.1 Filtro bilateral	16
4.1.1 Parâmetro s	16
4.1.2 Parâmetro σ_{color}	18
4.1.3 Parâmetro σ_{space}	19
4.2 Quantização de luminância	20
4.2.1 Parâmetro φ_{eq}	20
4.3 Contornos DoG	22
4.3.1 Parâmetro σ	22
4.3.2 Parâmetro φ_e	24
4.3.3 Parâmetro τ	26
5 Resultados	27
6 Conclusão	43
Bibliografia	45

Lista de Figuras

3.1	Esquema do método <i>Real time video abstraction</i>	7
3.2	Frame do vídeo "Dança" com filtro bilateral	9
3.3	Frame do vídeo "Conversa" com 2 valores de q	10
4.1	Frame do vídeo "Estrada" para vários valores de s	16
4.2	Imagem "Ventoinha" para vários valores de s	17
4.3	Imagem "Ventoinha" para vários valores de σ_{color}	18
4.4	Imagem "Ventoinha" para vários valores de σ_{space}	19
4.5	Imagem "Ventoinha" para vários valores de σ_{color} e σ_{space}	20
4.6	Imagem "Ventoinha" com dois valores para φ_{eq}	21
4.7	Frame do vídeo "Conversa" com dois valores de φ_e	21
4.8	Frame do vídeo "Casal" para vários valores de σ	22
4.9	Imagem "Baseball" para vários valores de σ	23
4.10	Imagem "Conversa" para vários valores de φ_e	24
4.11	Imagem "Ventoinha" para vários valores de φ_e	25
4.12	Frame do vídeo "Casal" para vários valores de τ	26
5.1	Dois frames originais retirados do vídeo "Conversa" e respetivas imagens de abstração	30
5.2	Dois frames originais retirados do vídeo "Dança" e respetivas imagens de abstração	31
5.3	Dois frames originais retirados do vídeo "Trânsito" e respetivas imagens de abstração	32
5.4	Dois frames originais retirados do vídeo "Casal" e respetivas imagens de abstração	33
5.5	Dois frames originais retirados do vídeo "Cavalo" e respetivas imagens de abstração	34
5.6	Dois frames originais retirados do vídeo "Estrada" e respetivas imagens de abstração	35
5.7	Imagem "Ventoinha" original e repetitiva imagem de abstração	36
5.8	Imagem "Baseball" original e repetitiva imagem de abstração	37
5.9	Comparação da abstração de imagem a partir de vários métodos	40

Lista de Tabelas

4.1	Parâmetros utilizados no método de abstração em Python . . .	13
4.2	Vídeos utilizados para avaliação do método	14
4.3	Imagens utilizadas para avaliação do método	15
5.1	Vídeos utilizados e o respetivo conjunto de parâmetros associado	28
5.2	Imagens utilizadas e o respetivo conjunto de parâmetros asso- ciado	29
5.3	Critério-tipo usado para a escolha dos valores dos parâmetros	38

Capítulo 1

Introdução

A estilização de imagem é o processo de modificação de uma imagem para que represente uma certa realidade, e assim dar uma nova forma à imagem. Muitos dos sistemas de estilização de imagens existentes foram projectados para propósitos puramente artísticos, como criar novas formas de arte digital ou, então, ajudar artistas com tarefas laboriosas ou tecnicamente desafiantes. Um desses sistemas, denominado renderização não foto-realista (NPR), incorpora um paradigma, designado por *image abstraction*, que se foca, essencialmente, na comunicação visual e na redução de dados. Este processo consiste na animação e na representação de uma certa imagem através da redução de detalhes irrelevantes de maneira a realçar os dados mais relevantes da mesma. Sendo a cor e a forma as duas características mais importantes a transmitir, os métodos existentes focam-se na resolução de dois problemas: *line drawing* (capturar descontinuidades relevantes) e *region smoothing* (remover descontinuidades insignificantes).

No artigo de Winnemoller et al. [1] é desenvolvido um método automático de abstração de imagens, diferenciando-se dos restantes métodos existentes pela rapidez e eficiência devido à capacidade de processar a imagem em tempo real. Apesar dos resultados serem satisfatórios na maior parte dos casos para os valores estipulados pelos autores, existem imagens que resultam numa abstração com falhas devido às suas características.

Como objetivo principal da tese, é proposta uma abordagem menos automática do método. Consiste em adaptar o valor do conjunto de variáveis pré-definidas no método em função das características da imagem ou vídeo. Desta forma, seriam corrigidas algumas falhas existentes perante um certo tipo de características já identificadas no artigo [1] como a falta de contraste de cores entre elementos. Um vez que a abstração de imagem é muito subjetiva, esta opção também permitiria ao utilizador definir com mais precisão o resultado pretendido.

Foi utilizada uma implementação do método no Python 3.10 e recolhido um conjunto de vídeos e imagens para fazer a sua abstração e testar os vários valores de cada parâmetro. O objetivo é encontrar os valores das variáveis que apresentam os resultados mais satisfatórios e procurar associá-los com as características relevantes presentes para cada imagem usada.

Após serem revistos vários trabalhos anteriores referentes a este tema, a avaliação de cada foi feita através de estudos. Em cada estudo, é pedido a um grupo que faça uma avaliação subjetiva. A avaliação consiste na escolha da abstração mais eficaz entre um conjunto de imagens, cada uma com um método de abstração diferente ou, uma comparação entre uma imagem original e respetiva abstração. Em termos objetivos, o único tipo de avaliação existente é o tempo de execução, que ainda assim, não é suficiente para ter uma avaliação completa ao método.

Para a avaliação dos resultados obtidos, foi implementado um programa que permite ao utilizador seleccionar dois vídeos com parâmetros diferentes e compará-los frame a frame em simultâneo para escolher os valores do conjunto de parâmetros considerados como mais eficaz. Em relação ao processo de recolha de vídeos e imagens, para abranger um maior número de características possível, estes foram escolhidos com características diferentes, tanto em relação à velocidade e frequência de movimentos nos vídeos como ao detalhe e cor de todos os vídeos e imagens. Para assegurar um tamanho de cada vídeo e imagem o mais semelhante possível, em termos de tamanho de cada frame/imagem ou duração de cada vídeo, cada um foi adaptado para terem um tamanho semelhante e para o vídeo ser de curta duração, permitindo também um tempo de execução melhor. Este processo é explicado no capítulo 4.

Será feita numa primeira parte um enquadramento do tema, expondo a evolução dos métodos de abstração ao longo dos anos dando um destaque a alguns trabalhos anteriores relevantes. De seguida, é feita a descrição do método utilizado e da implementação do mesmo em Python. No capítulo seguinte (4), é apresentado o conjunto dos vídeos e imagens recolhidos para uma posterior avaliação dos mesmos bem como as variáveis presentes no algoritmo implementado retirado da plataforma github, onde é explicado em detalhe o efeito que cada uma produz no output final. No capítulo "Resultado" são apresentados os resultados da abstração dos vários vídeos e imagens obtidos através da comparação entre frames originais e respetivos outputs, os valores que apresentam melhores resultados para cada parâmetro e a justificação para essas escolhas.

Finalmente, é apresentada uma comparação entre alguns vídeos após a abstração com os parâmetros originais definidos em [1] e os resultados obtidos no capítulo 5 e concluir qual a abordagem superior em termos de precisão.

Capítulo 2

Estado de arte

Antes de analisar o método abordado neste trabalho, é importante destacar alguns artigos e métodos anteriores que contribuíram para a evolução da estilização de imagens e vídeos. A estilização de imagens começou a ganhar relevância no início dos anos 90 com um sistema semi-automático de pintura [2]. O método consiste no utilizador clicar numa imagem e cada clique gerava uma pincelada de cor de fundo. Este conceito serviu de base para a maioria dos trabalhos de estilização nos anos 90. Este tipo de método é conhecido como *Stroke base rendering* (SBR).

A passagem de SBR para estilização de vídeo requer duas condições: o movimento das pinceladas estão de acordo com o conteúdo do vídeo e a sequência de frames tem de ser fluida, ou seja, não pode dar a sensação de cortes no vídeo.

Em 1997, Litwinowicz [3] contribuiu para a primeira extensão de um algoritmo de estilização para vídeo. O processo consiste em colocar pinceladas com um tamanho fixo no primeiro frame e estimar a localização das pinceladas dos frames seguintes através de um filtro de Sobel para cumprir a coerência temporal. Contudo, este método não é totalmente automático uma vez que requer que o utilizador especifique alguns parâmetros no início do processo.

O artigo de DeCarlo e Santella [4] foi importante para servir de base para vários sistemas de estilização de vídeo através da segmentação de imagens usando uma variante do desvio-médio [5]. A imagem é dividida em várias camadas sob uma hierarquia de espaço de escala. Esta hierarquia permite que a imagem seja renderizada de uma forma abstrata. O processo é obtido através da identificação dos elementos importantes da imagem com a utilização de um modelo de percepção humano e a gravação do movimento do olho humano ao observar a imagem.

Entre as várias técnicas que surgiram relativamente à abstração de imagens, nenhuma era viável para processar vídeos em tempo real. Já técnicas de processamento de imagens que recorrem a operações envolvendo filtros locais constituem uma alternativa válida uma vez que a paralelização e implementação em GPU é direta na maior parte dos casos.

Winnemoller et al. [1] apresenta um método automático e em tempo real de abstração de vídeo e imagem que consideram apresentar resultados mais satisfatórios e mais rápidos em relação a trabalhos anteriores.

Em relação ao método já mencionado de DeCarlo e Santella [4], não inclui a coerência temporal ao contrário do modelo de Winnemoller et al. [1] que, apesar de ser menos elaborado, pode ser executado em tempo real. Este também não requer a gravação do movimento do olho humano.

Raskar et al. [6] apresentou um método baseado numa câmara com múltiplos flashes estrategicamente posicionados de modo a facilitar a deteção de descontinuidades, contornos e, assim, conseguir reduzir a textura de forma mais eficiente. Por contrapartida, exige um hardware especializado e enfrenta dificuldades técnicas em relação ao multi-flash em vídeo.

Na abordagem de Fisher et al. [7] onde pretende fornecer ao utilizador uma imagem enriquecida através de um objeto virtual, apresentam 2 métodos diferentes. No primeiro método, utilizam o filtro bilateral com o auxílio do filtro Gaussiano para a segmentação de cores com o objetivo de otimizar o tempo de execução do seu método, reduzindo o número de píxeis da imagem antes de aplicar o filtro bilateral para depois voltar a aumentar para o seu tamanho original. Para os contornos, aplicam o *Canny edge detector* na imagem original. Para o método *cartoon-like* tem como duas principais componentes uma técnica para desenhar contornos grossos nos objetos e um método de *shading* personalizado que tem como objetivo reduzir o número de cores na imagem gerada. Esta sua implementação acaba por ser limitada em relação ao detalhe e os contornos tendem a ter muito ruído.

Gooch et al. [8] utiliza a Diferença de Gaussianos e um modelo simples de percepção de brilho no seu método automático de caricaturas faciais monocromáticas. O artigo de Winnemoller et al. [1] [2006] adota um modelo semelhante para os contornos além de ter em conta a coerência temporal, cor e *performance* em tempo real.

Kang et al. [9] propôs um método baseado no filtro bilateral e Diferença de Gaussianos tal como Winnemoller et al. [1], em que os filtros seguem um campo de fluxos, o *Edge tangent flow*. O método além de não ser em tempo real, é menos satisfatório em imagens com maior contraste em relação a Winnemoller et al. [1].

Redmond et al. [10] apresenta um método de abstração de fotografias 3D através do filtro de Kuwahara, Diferença de Gaussianas e uma técnica de

sombreamento plano. O filtro de Kuwahara simplifica a forma, arestas e cor simultaneamente enquanto preserva recursos importantes. A diferença de Gaussianos preserva os contornos da imagem. O método falha uma vez que o filtro de Kuwahara perde recursos importantes da imagem, quando existe pouca luminosidade.

Capítulo 3

Real time video abstraction

Neste capítulo será explicado o método de Winnemoller et al. [1]. O capítulo está dividido de acordo com a figura 3.1 que mostra as várias etapas existentes: Filtro bilateral, quantização da luminância, contornos de Diferença de Gaussianos e a etapa opcional Deformação de Imagem (IBW).

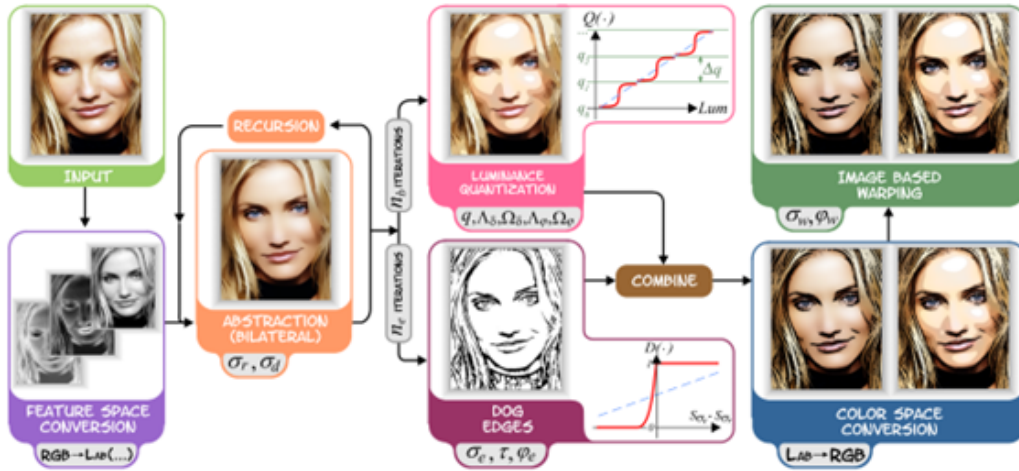


Figure 3.1: Esquema com os vários parâmetros e outputs associados a cada parte do método de abstração

3.1 Filtro Bilateral

Inicialmente, a imagem é convertida do sistema de cores RGB para o espaço de cores CIELAB, que expressa a cor em três valores: L para luminosidade perceptual e a e b para as quatro cores únicas da visão humana: vermelho, verde, azul e amarelo. Daqui em diante, este espaço de cores será designado por $L^*a^*b^*$. Em seguida, é aplicado o filtro bilateral a cada componente $L^*a^*b^*$ para suavizar a imagem onde o valor de cada pixel é substituído pela média dos valores dos pixels vizinhos. Neste filtro, $\hat{x}$ é a localização do pixel e x são os pixels vizinhos, σ_d é referente ao raio do *blur* onde aumento do seu valor se traduz em maior *blurring*. Dado o input da imagem $f(\cdot)$, o filtro definido é dado por $H(\cdot)$:

$$H(\hat{x}, \sigma_d, \sigma_r) = \frac{\int_e^{-\frac{1}{2}(\frac{\|\hat{x}-x\|}{\sigma_d})^2} w(x, \hat{x}) f(x) dx}{\int_e^{-\frac{1}{2}(\frac{\|\hat{x}-x\|}{\sigma_d})^2} w(x, \hat{x}) dx} \quad (3.1)$$

A função $w(\cdot)$ determina em que partes da imagem os contrastes são realçados ou suavizados através de aplicações de iterações de $H(\cdot)$:

$$w(x, \hat{x}, \sigma_r) = (1 - m(\hat{x})) \cdot w'(x, \hat{x}, \sigma_r) + m(\hat{x}) + u(\hat{x}) \quad (3.2)$$

$$w'(x, \hat{x}, \sigma_r) = e^{-\frac{1}{2} \left(\frac{\|f(\hat{x}) - f(x)\|}{\sigma_r} \right)^2} \quad (3.3)$$

Para a aplicação em tempo real, $m(\cdot) = 0$ tal que $w(\cdot) = w'(\cdot)$ (equações 3.2 e 3.3), a equação 3.1 transforma-se no conhecido filtro bilateral que se vê aplicado na figura 3.2 onde σ_r determina como o contraste vai ser preservado ou *blurred*. Valores baixos de σ_r preservam quase todos os contrastes e os filtros têm pouco impacto na imagem, enquanto que para valores elevados, $w'(\cdot) \xrightarrow{\sigma_r \rightarrow \infty} 1$, tornando $H(\cdot)$ num *standard Gaussian blur* linear. Os autores usam $\sigma_d = 3$ em todos os casos e $\sigma_r = 4.25$ para a maior parte dos vídeos e imagens. Na figura 3.2 é possível ver uma comparação entre uma imagem original e a mesma após ser aplicado o filtro bilateral.



Figure 3.2: Frame do vídeo “Dança” [11] original (esq.) e com filtro bilateral aplicado (dir.)

3.2 Quantização de luminância

A quantização de luminância consiste em atribuir uma escala de luminância com menor diversidade do que a escala original, que é feito através da equação 3.4.

$$Q(\hat{x}, q, \varphi_q) = q_{nearest} + \frac{\Delta q}{2} \tanh(\varphi_q \cdot (f(\hat{x}) - q_{nearest})) \quad (3.4)$$

Onde $Q(.)$ representa a imagem pseudo-quantizada, Δq a largura dos intervalos, $q_{nearest}$ o valor da fronteira do intervalo mais próximo a $f(\hat{x})$ e φ_q é um parâmetro que controla a suavidade entre intervalos. O número de intervalos usado costuma situar-se em $q \in [8, 10]$ igualmente espaçados. A função descrita pela equação 3.4 é formalmente descontínua, mas para valores suficientemente grandes de φ_q as discontinuidades não são visíveis. Para minimizar as transições dissonantes, os autores admitem que φ_q é função do gradiente de luminosidade da imagem abstraída. Para gerar valores de φ_q , primeiramente, restringiram os gradientes calculados ao intervalo $[\Lambda\delta, \Omega\delta]$ e, de seguida, estabeleceram uma correspondência linear com o intervalo $[\Lambda\varphi, \Omega\varphi]$. Tipicamente, usa-se $[\Lambda\delta = 0, \Omega\delta = 2]$ e $[\Lambda\varphi = 3, \Omega\varphi = 14]$. Na figura 3.3 encontra-se uma comparação lado a lado entre uma imagem quantizada dentro do intervalo estipulado para q ($q=8$) e uma quantizada a 4 níveis que causa manchas demasiado irrealistas por usar uma escala demasiado curta para a luminância.



Figure 3.3: Frame do vídeo "Conversa" [12] com 2 valores de q (número de níveis)

3.3 Contornos de diferenças de Gaussianas

Para os contornos da imagem foi utilizada a Diferença de Gaussianas pela sua simplicidade e eficiência computacional quando comparado com outros filtros. Neste método é utilizado o filtro de Gauss por duas vezes, com um raio de desfocagem diferente em cada execução na imagem original e o resultado final é dado pela subtração de uma imagem pela outra. Para este filtro, é aplicado um degrau na função $D(\cdot)$ para aumentar a coerência temporal nas animações. Para este filtro, o parâmetro τ controla a quantidade da diferença *center-surround* requerida para a ativação da célula, para valores baixos de τ é detetado menos ruído mas também o nível de contorno real é reduzido. À medida que $\tau \rightarrow 1$, o filtro fica cada vez mais instável. Os autores utilizam $\tau = 0.98$. O σ_e determina a escala espacial para detecção de contornos e à medida que o seu valor aumenta, maior é a espessura dos contornos. O parâmetro φ_e é responsável pela nitidez das arestas e o seu valor encontra-se no intervalo $[0.75, 5]$. Para n_b iterações do filtro bilateral, os contornos são obtidos após $n_e < n_b$ iterações para reduzir o ruído. Tipicamente, $n_e \in \{1, 2\}$ e $n_b \in \{3, 4\}$.

$$D(\hat{x}, \sigma_e, \tau, \varphi_e) = \begin{cases} 1 & \text{se } S_{\sigma_e} - \tau \cdot S_{\sigma_r} > 0 \\ 1 + \tanh(\varphi_e \cdot (S_{\sigma_e} - \tau \cdot S_{\sigma_r})) & \text{caso contrário} \end{cases}$$

Na equação 3.5 está definida a função de ruído, onde S_{σ_e} e S_{σ_r} estão definidos nas equações 3.6 e 3.7 respectivamente.

$$S(\hat{x}, \sigma_e) = \frac{1}{2\pi\sigma_e^2} \int f(x) e^{-\frac{1}{2}(\frac{\|\hat{x}-x\|}{\sigma_e})^2} dx \quad (3.5)$$

$$S_{\sigma_e} \equiv S(\hat{x}, \sigma_e) \quad (3.6)$$

$$S_{\sigma_r} \equiv S(\hat{x}, \sqrt{1.6}\sigma_e) \quad (3.7)$$

Deformação de imagem (IBW)

Para corrigir alguns erros nos contornos DoG e para melhorar a aparência final do resultado os autores fazem um procedimento opcional que consiste na deformação da imagem (IBW). IBW é uma técnica criada por Arad e Gotsman [13] em que se movem píxeis ao longo da imagem em direção aos contornos. Loviscach [14] propôs um método de IBW mais simples, em que o campo deformado é o resultado escalado e desfocado do filtro de Sobel da imagem. É usado este método com $\sigma_w = 1.5$ e $\varphi_w = 2.7$.

3.4 Implementação do método

Para a avaliação do método de abstração de imagem, foi efetuada uma implementação em Python 3.10. O programa consiste na abstração de uma imagem ou vídeo, onde o resultado final resulta da concatenação de cada frame após a abstração.

O método está dividido em 3 partes: Filtro bilateral, definido pelas variáveis s , σ_{color} e σ_{space} ; Quantização da cor, através da variável φ_{eq} e Diferença de Gaussianos com as variáveis σ , φ_e e τ . (variáveis definidas na tabela).

Em primeiro lugar, é feita a conversão do formato RGB para L*a*b. É aplicado o filtro bilateral aos canais L, a, b com os parâmetros escolhidos pelo utilizador s , σ_{color} e σ_{space} através da função do filtro bilateral pré-implementada do módulo cv2.

A quantização da cor é obtida através da função 3.8, onde a função $qnearest(f)$ aproxima o valor da luminância (L) em cada pixel dependendo do número do número de níveis utilizado. Nesta implementação foi escolhida a quantização em 8 níveis por se encontrar no intervalo normalmente utilizado [8,10]. Assim, o valor de luminância de cada pixel é aproximado para o valor associado do nível inserido onde cada pixel está inserido.

$$Q(\cdot) = qnearest(L) + \frac{5tanh(L - qnearest(L))}{\varphi_{eq}} \quad (3.8)$$

Para os contornos da imagem é utilizado o método da diferença de Gaussianos. É implementada a função $Gauss2D(shape, \sigma)$ que permite obter o filtro de Gauss de dimensões N por M (shape). É aplicado o filtro ao canal de luminância, com $shape=(5,5)$ para dois valores diferentes: σ e $1.5*\sigma$, definidos pelo utilizador e obtém-se o contorno final através da função $D(\cdot)$, onde S_{σ_e} e S_{σ_r} representam o filtro para σ e $1.5*\sigma$ respetivamente e τ e φ_e são ambos parâmetros definidos pelo utilizador. Após juntar as diferentes partes, é feita a conversão de volta do formato L*a*b para RGB.

Capítulo 4

Avaliação experimental

O método implementado em Python tem 7 parâmetros, sendo que estes podem ser repartidos pelas 3 partes já abordadas no capítulo 3.4. Os parâmetros s , σ_{color} e σ_{space} são os responsáveis pelo filtro bilateral. Para a quantização da luminância o único responsável é o φ_{eq} enquanto que para os contornos DoG são associados os parâmetros σ , φ_e e τ . Na tabela 4.1 é descrito cada parâmetro, valores possíveis e o nível de impacto na imagem.

Parâmetro	Descrição	Valores possíveis	Impacto	Observações
s	Diâmetro da vizinhança do pixel usado para o filtro bilateral	Valor mínimo: 0	Relevante	Para valores elevados deste parâmetro o tempo de computação é maior (uma imagem de 640x640 demora 8 minutos para $s=400$)
σ_{color}	As cores são agrupadas cada vez mais à medida que o valor aumenta	Valor mínimo: 0	Relevante	
σ_{space}	Quanto maior é o valor, mais os píxeis se agrupam tal como as cores no sigma-color	Valor mínimo: 0	Relevante	
φ_{eq}	Controla a quantização da cor	$]0; \text{inf}]$	Pouco relevante	Apenas relevante para vídeos com alvos de tamanho elevado
σ	Sigma gaussiano	$]0; \text{inf}]$	Pouco Relevante	Valor ideal em $\text{sigma}=1$
φ_e	controls the sharpness of the activation falloff	$[0.75, 5]$	Pouco Relevante	Apenas obtém resultados satisfatórios para valores elevados
τ	control the center-surround difference required for cell activation	$[0, 1]$	Muito relevante	Impacto quase nulo para valores inferiores a 0.97

Table 4.1: Parâmetros utilizados no método de abstração em Python e a descrição, valores possíveis, impacto e observações de cada um

Para a avaliação do método foram recolhidos vários vídeos de teste da web, apresentados na tabela 4.2 juntamente com uma breve descrição e características, tais como o tempo, tamanho, número de frames e tempo de execução médio por frame.

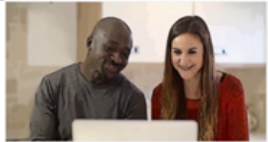
Nome	Tempo original	Descrição	Nº Frames	Tempo Execução	Tamanho	Frames
Conversa	20 s.	Homem e mulher a falar em frente ao computador	622	25s.	480x910	
Dança	17 s.	Grupo de pessoas a dançar em público	415	23s.	480x854	
Trânsito	16 s.	Vídeo acelerado de uma rua movimentada de noite	501	22s.	480x854	
Casal	10 s.	Casal a ver fotografias na máquina	300	30s.	480x854	
Cavalo	11 s.	Cavalo a andar no hipódromo	344	34s.	480x854	
Estrada	22 s.	Vídeo do percurso de uma estrada	530	13s.	480x270	

Table 4.2: Descrição de todos os vídeos utilizados para avaliação do método bem como o número de frames, tamanho, tempo de cada um e o respetivo tempo de execução médio por frame

Apesar do método ser em tempo real, a implementação em Python é demasiado lenta. Para reduzir o tempo de execução de forma considerável, o tamanho de todos os vídeos e imagens utilizados foi ajustado para que não ultrapasse os 480 píxeis de altura, respeitando a razão entre altura e largura original. Após esta alteração, é possível ver na tabela 4.2 o tempo de execução por frame de cada vídeo. O número de frames de cada vídeo é também ajustado para obter um tempo total de execução razoável. O vídeo é limitado a um máximo de 20 segundos e apenas são utilizados frames intervalados de 4 em 4, eliminando do vídeo os restantes. Após estas alterações, a coerência temporal mantém-se e permite que todos os vídeos utilizados

nunca ultrapassem o tempo de execução de 70 minutos.

Na tabela 4.3 encontram-se as imagens recolhidas, bem como uma descrição e o tamanho de cada uma. Ao contrário dos vídeos, as imagens não sofreram qualquer alteração por se tratar de apenas um frame. Como tal, o seu tempo de execução é muito baixo.

Nome	Descrição	Tamanho	Imagem
Ventoinha	Paisagem com Ventoinha eólica	420x631	
Baseball	Treinador e jogador num campo de baseball	184x185	

Table 4.3: Descrição das imagens utilizadas para avaliação do método bem como o tamanho de cada uma

Para a seleção dos vídeos e imagens, a prioridade foi encontrar características diferentes para conseguir comparar as diferenças nos valores escolhidos para cada parâmetro. Vídeos como "Dança" e "Estrada" têm muito movimento e desafiam a coerência temporal da abstração, sendo que o "Estrada" é responsável por uma velocidade maior além do tamanho reduzido dos carros. Os vídeos "Conversa" e "Casal" são os únicos dos seleccionados com a abstração do rosto de pessoas, onde no caso do "Casal" acrescenta ao pouco contraste entre as pessoas e o fundo do vídeo uma maior claridade no vídeo. O pouco contraste também está presente no vídeo "Estrada". O vídeo "Cavalo" não é tão detalhado mas requer atenção à variação da claridade ao longo do vídeo bem como aos contornos no solo e árvores. Para as imagens, foram seleccionadas apenas duas com elementos simples e claramente identificados uma vez que a abstração de vídeos é a mais desafiante.

4.1 Filtro bilateral

Os parâmetros s , σ_{color} e σ_{color} são associados ao filtro bilateral e têm o seu valor definido pelo utilizador para serem usados na função `BilateralFilter` já implementada em Python. Nesta secção será explicado em maior detalhe o que faz cada parâmetro bem como quais são os valores indicados para cada um tendo em conta as características de cada imagem.

4.1.1 Parâmetro s

O parâmetro s define o diâmetro da vizinhança do pixel usado. Para esta variável qualquer valor igual ou superior a 0 é aceite, sendo que quanto maior for, menos detalhes terá a imagem. Tal é visível no exemplo apresentado na figura 4.1 - para um valor mais baixo (4.1a) é possível distinguir as árvores na zona da floresta apesar do pouco contraste existente enquanto que para valores elevados (4.1c) apenas se vê uma mancha monótona na mesma parte da imagem. Podemos retirar deste exemplo que um valor relativamente baixo deverá ser o mais adequado para imagens com este tipo de característica.

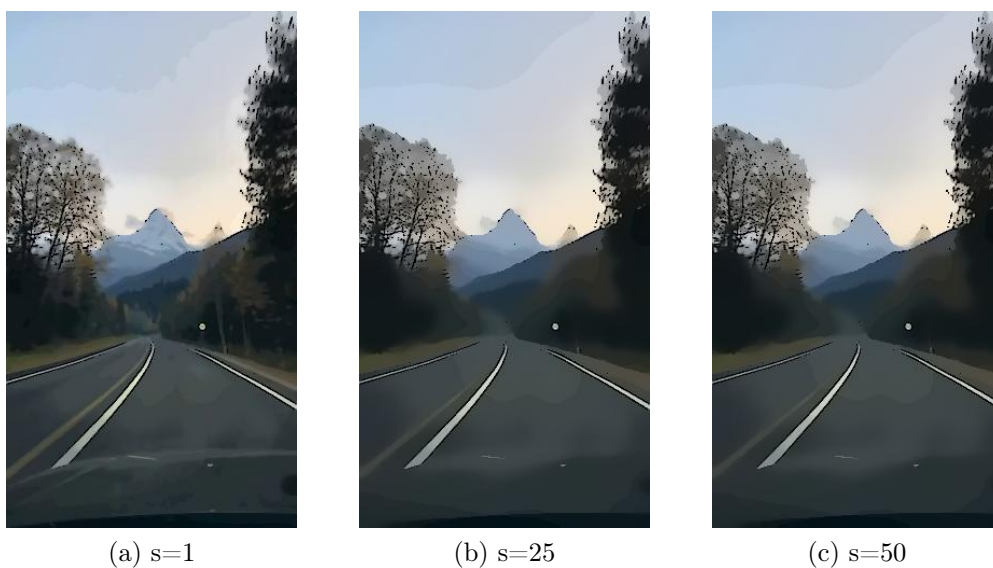


Figure 4.1: Frame do vídeo “Estrada” [15] para vários valores do parâmetro s

Na figura 4.2 encontra-se um exemplo de um tipo de imagem diferente, com mais contrastes e menos detalhe. Neste caso, o valor mais adequado do parâmetro é mais elevado devido à facilidade de distinguir os vários elementos da imagem e à perda de detalhe. Foi escolhido o valor de $s=30$ uma vez que para valores mais elevados, a imagem não sofre qualquer alteração (Figuras 5.7c e 5.7d). Para valores mais baixos, nota-se o pouco impacto do filtro na imagem com destaque para o relevo da zona verde. Aconselha-se então um valor mais elevado neste parâmetro para imagens com uma menor necessidade de detalhe e um contraste evidente entre as várias partes da imagem. Um valor mais baixo é mais indicado para casos como o da figura 4.1 com um contraste menos evidente que pode ocultar partes importantes da imagem-alvo.



(a) $s=1$



(b) $s=10$



(c) $s=30$



(d) $s=50$

Figure 4.2: Imagem "Ventoinha" [16] para vários valores do parâmetro s

4.1.2 Parâmetro σ_{color}

O parâmetro σ_{color} é responsável pelo agrupamento das cores da imagem, ou seja, à medida que o valor do parâmetro diminui, torna a imagem mais monótona em relação ao esquema de cores. A figura 4.3 mostra o exemplo de uma imagem "Ventoinha" para 4 valores de σ_{color} . Para o valor mais elevado do parâmetro σ_{color} apresentado (figura 4.3d), nota-se uma maior monotonia de cores numa imagem pouco nítida. À medida que o valor diminui é possível identificar um maior detalhe e uma maior diferença entre os vários elementos existentes na imagem, sobretudo em casos com maior variabilidade e contraste de cores como o exemplo da figura 4.3. Para o caso da figura 4.3a ($\sigma_{color} = 1$), as alterações presentes na imagem em relação à original são quase nulas. Foi então considerado o valor $\sigma_{color} = 5$ (figura 4.3b) como o mais adequado por ser facilmente identificável os elementos da imagem e consegue ocultar o relevo irrelevante para a abstração presente nas montanhas e na relva.



Figure 4.3: Imagem "Ventoinha" [16] para vários valores do parâmetro σ_{color}

4.1.3 Parâmetro σ_{space}

Tal como o σ_{color} , o parâmetro σ_{space} está associado ao agrupamento de cores na imagem. Este parâmetro controla o nível de agrupamento de píxeis. Na figura 4.4 podemos ver o resultado na imagem "Ventoinha" para 2 valores de σ_{space} ao aumentar o valor da variável, onde quanto maior for o seu valor, maior a mistura dos píxeis da imagem. Para esta variável, a partir de valores elevados como em 4.4b, o seu efeito deixa de ser perceptível.



(a) $\sigma_{space} = 5$



(b) $\sigma_{space} = 25$

Figure 4.4: Imagem "Ventoinha" [16] para vários valores do parâmetro σ_{space}

O parâmetro σ_{space} e o σ_{color} estão associados entre si. Assim, o efeito que provocam na imagem é mais notório quando são alterados os valores de ambos em simultâneo. Um valor elevado dos 2 parâmetros vai agrupar as cores de mais píxeis em simultâneo e torna-se ainda mais notória a diferença entre valores extremos. Na figura 4.5 encontram-se 2 imagens com valores distintos para ambos os parâmetros σ_{color} e σ_{space} . Para valores mais altos dos parâmetros (figuras 4.5c e 4.5d) destaca-se a monotonia de cores, falta de contornos e pouco contraste entre as várias partes da imagem quando comparado com as figuras 4.5a e 4.5b onde os valores dos parâmetros são mais baixos.



(a) $\sigma_{color}=1$ e $\sigma_{space}=1$



(b) $\sigma_{color}=5$ e $\sigma_{space}=5$



(c) $\sigma_{color}=15$ e $\sigma_{space}=15$



(d) $\sigma_{color}=25$ e $\sigma_{color}=25$

Figure 4.5: Imagem "Ventoinha" [16] para vários valores dos parâmetros σ_{color} e σ_{space}

4.2 Quantização de luminância

4.2.1 Parâmetro φ_{eq}

O parâmetro φ_{eq} é o único responsável pela quantização da luminância do canal L que controla a nitidez da transição de intensidade entre píxeis vizinhos. Através da figura 4.6, podemos identificar as diferenças presentes entre as imagens resultantes da aplicação do processo de abstração de imagem com 2 valores diferentes de φ_{eq} (5 e 25). Ao escolher um valor mais elevado existe uma diferença maior nos vários tons de cor. Este efeito é indesejado pois provoca um contraste exagerado e desnecessário em certas partes da imagem como se observa na figura 4.6b.



(a) $\varphi_{eq} = 5$



(b) $\varphi_{eq} = 25$

Figure 4.6: Imagem "Ventoinha" [16] com dois valores para φ_{eq}

Quando a imagem-alvo envolve rostos de pessoas, trata-se de um caso onde se evidencia mais o impacto do parâmetro φ_{eq} . Na figura 4.7, encontra-se um exemplo claro onde um valor mais elevado (figura 4.7b) provoca uma maior diferença na cor, destacando-se o tom de pele das pessoas que prejudica o realismo da abstração num elemento chave da imagem. É mais adequado utilizar um valor reduzido para conseguir manter a coerência da imagem.



(a) $\varphi_{eq} = 2$



(b) $\varphi_{eq} = 30$

Figure 4.7: Frame do vídeo "Conversa" [12] com dois valores do parâmetro φ_e

4.3 Contornos DoG

4.3.1 Parâmetro σ

O parâmetro σ é um dos 3 responsáveis pela detecção de contornos através do filtro Gaussiano, podendo tomar qualquer valor positivo. Apesar disso, para valores de σ entre 0 e 0.2, a imagem perde todo o contorno (figura 4.8a). O mesmo acontece quando o σ toma valores maiores que 2.2 (figura 4.8d) em que o nível de contorno é praticamente nulo. Em relação aos restantes valores permitidos pelo parâmetro, o número de contornos da imagem aumenta à medida que σ diminui. Na imagem "Casal", o cabelo e a roupa são os exemplos mais evidentes da variação causada pelo parâmetro. Devido ao ruído provocado na camisa do homem e a alguns contornos indesejados na face da mulher, o resultado para $\sigma = 1$ (figura 4.8c) é o mais eficaz.

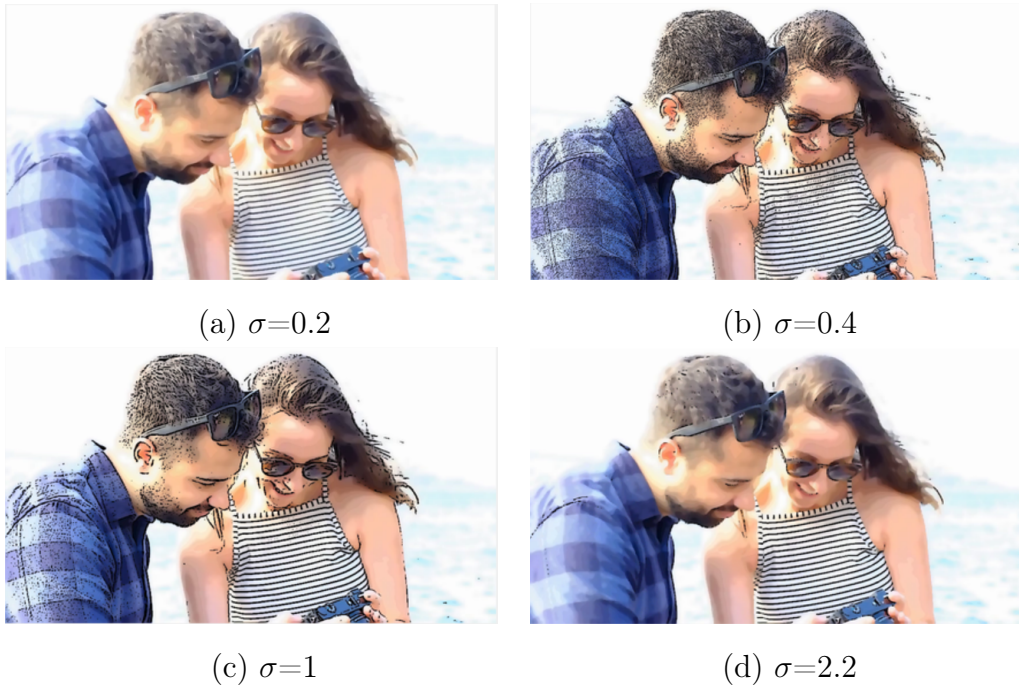


Figure 4.8: Frame do vídeo “Casal” [17] para vários valores do parâmetro σ

A figura 4.9 mostra a imagem "Baseball" para 4 valores diferentes de σ . Neste caso, um valor mais baixo do σ é mais eficaz por ser necessário um maior detalhe na imagem como se comprova na diferença de contorno entre $\sigma = 0.6$ (figura 4.9b) e $\sigma = 1.2$ (figura 4.9c). Para valores superiores ($\sigma = 1.2$) apesar de causarem menos ruído no solo, também detetam menos contornos relevantes nos elementos importantes da imagem. Tal como no

exemplo anterior, para valores muito baixos até 0.2 (figura 4.9a) e superiores a 2.2 (figura 4.9d) a imagem fica com um nível nulo de contorno.



(a) $\sigma=0.2$



(b) $\sigma=0.6$



(c) $\sigma=1.2$



(d) $\sigma=2.2$

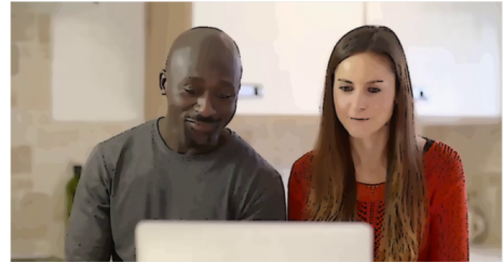
Figure 4.9: Imagem “Baseball”[18] para vários valores do parâmetro σ .

4.3.2 Parâmetro φ_e

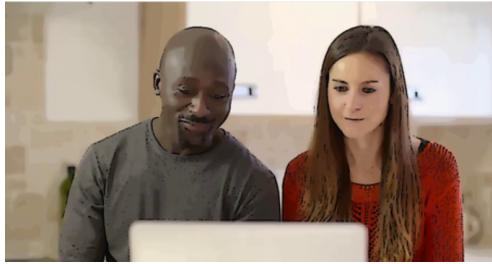
O segundo parâmetro responsável pela detecção de contornos é o parâmetro φ_e , que controla a nitidez dos contornos. Os valores variam entre 0.75 e 5 sendo que quanto maior for, mais contornos a imagem irá mostrar. Na figura 4.11 temos um exemplo de uma imagem para um valor baixo de φ_e (figura 4.10a) e alto (figura 4.10d). Este exemplo requer um maior detalhe em relação às expressões faciais das pessoas, portanto é necessário um valor de φ_e mais elevado. No caso da figura 4.10d o único contorno excessivo na imagem situa-se nas camisolas e à medida que o valor do φ_e diminui, também os contornos correspondentes à fronteira entre as pessoas e o fundo da imagem vão desaparecendo até chegar ao valor mínimo na figura 4.10a que elimina todos os contornos. Assim, o valor determinado como mais favorável para esta imagem foi de $\varphi_e=5$, para assegurar que todos os contornos necessários estejam presentes na imagem.



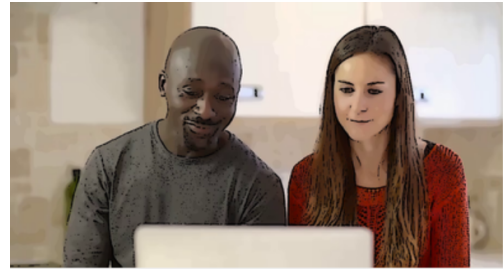
(a) $\varphi_e=0.75$



(b) $\varphi_e=2$



(c) $\varphi_e=3$



(d) $\varphi_e=5$

Figure 4.10: Imagem “Conversa”[12] para vários valores do parâmetro φ_e .

Na figura 4.11 temos um exemplo oposto, tratando-se de uma imagem bastante simples em termos de detalhe. Começando pelo valor máximo desta variável (figura 4.11d) é possível observar um excesso de contorno ao longo da imagem, tanto no ruído desnecessário no solo e montanhas como junto das nuvens. O valor do parâmetro é então ajustado até obter uma imagem satisfatória que não contenha o ruído indesejado. Uma vez que a imagem não necessita de um nível elevado de contorno, os valores intermédios (figuras 4.11b e 4.11c) também são excessivos. Apesar de no caso para $\varphi_e = 3$ os contornos necessários estarem presentes em relação a valores mais baixos como $\varphi_e = 0.75$, o ruído presente provoca demasiado impacto negativo na imagem. O valor escolhido para este exemplo é o valor mínimo $\varphi_e = 0.75$ (Figura 4.11a). Este valor permite eliminar quase todo o ruído existente com exceção de uma parte da zona verde do solo. Por se tratar de uma imagem com elementos facilmente distinguíveis, o facto de faltarem alguns contornos não compromete a abstração.



(a) $\varphi_e = 0.75$



(b) $\varphi_e = 2$



(c) $\varphi_e = 3$



(d) $\varphi_e = 5$

Figure 4.11: Imagem “Ventoinha”[16] para vários valores do parâmetro φ_e .

4.3.3 Parâmetro τ

O parâmetro τ controla o tamanho da região envolvente (*center-surround difference*) requerida para a ativação da célula. O valor de τ varia entre 0 e 1. A figura 4.12 mostra um exemplo de abstração da imagem "Casal" para 4 valores de τ (0.98, 0.987, 0.994 e 1). Observamos que para o valor máximo permitido obtemos um nível de contorno muito elevado. Por um lado consegue captar todos os detalhes da imagem e dá destaque aos intervenientes/objetos na imagem em detrimento do fundo. Por outro lado, acaba por ter algum ruído indesejado. Para valores definidos abaixo dos 0.97 para o parâmetro τ qualquer imagem fica sem contorno. Desta forma, conclui-se que os valores ideais se situam entre 0.98 e 1 uma vez que mesmo para valores como 0.98 (figura 4.12a) o nível de contorno é muito reduzido. No caso específico do vídeo "Casal" foi escolhido um valor elevado devido à clareza presente na imagem que não permite detetar os contornos das pessoas em relação ao fundo e à necessidade de destacar alguns detalhes tanto no cabelo como no rosto de cada um.

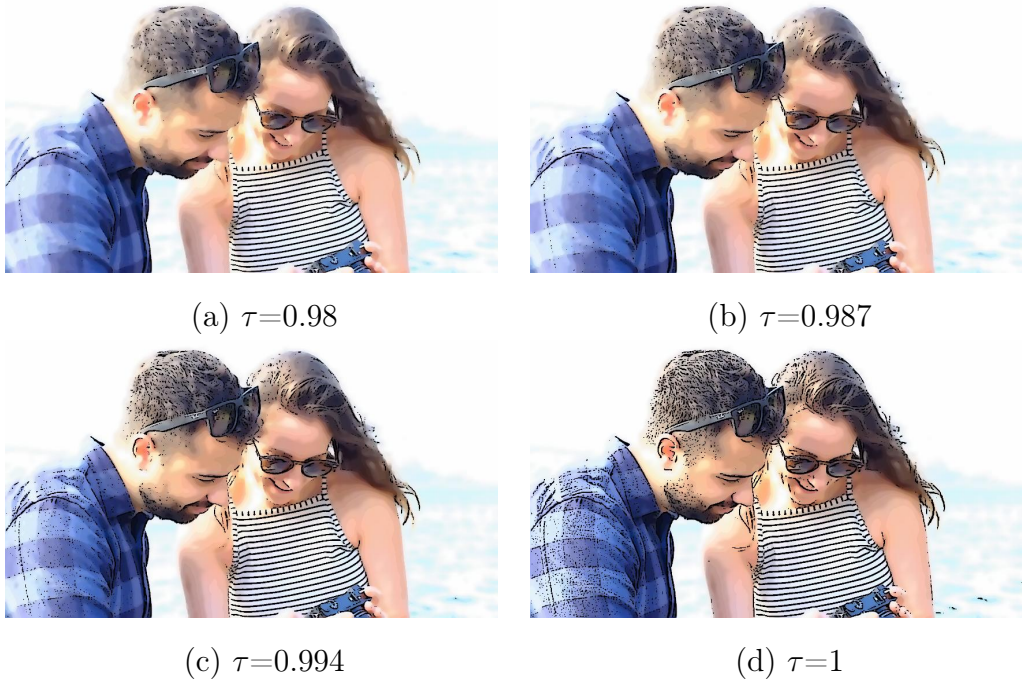


Figure 4.12: Frame do vídeo "Casal"[17] para vários valores do parâmetro τ .

Capítulo 5

Resultados

Foram efetuados testes experimentais com 6 vídeos e diversas configurações de parâmetros do processo de abstração de imagem. Na tabela 5.1 encontra-se uma breve descrição de cada vídeo utilizado neste artigo para análise do método bem como os parâmetros escolhidos, um frame retirado do output de cada abstração bem como as dificuldades mais relevantes para cada vídeo. De seguida, cada um irá ser analisado em maior detalhe, através da explicação dos valores obtidos e como estes se associam às características presentes no vídeo, como o movimento, contraste de cores ou o nível de detalhe dos vários elementos.

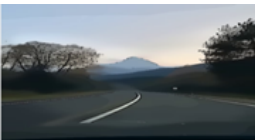
Nome	Tempo	Descrição	Frame	Parâmetros	Dificuldades
Conversa	7 s.	Homem e mulher a falar em frente ao computador		$s=7$ $\sigma_{color} = 4$ $\sigma_{space} = 4$ $\phi_{eq} = 2$ $\sigma = 1$ $\phi_e = 4.9$ $\tau = 0.993$	Fundo monótono dificulta a abstração. Ruído na roupa.
Dança	5 s.	Grupo de pessoas a dançar em público		$s=5$ $\sigma_{color} = 4$ $\sigma_{space} = 4$ $\phi_{eq} = 2$ $\sigma = 1$ $\phi_e = 5$ $\tau = 0.985$	Quando passam pessoas a uma distância muito próxima não difere muito do original.
Trânsito	6 s.	Vídeo acelerado de uma rua movimentada de noite		$s=2$ $\sigma_{color} = 3$ $\sigma_{space} = 3$ $\phi_{eq} = 2$ $\sigma = 1$ $\phi_e = 5$ $\tau = 0.992$	Vídeo a uma velocidade muito rápida, a qualidade da abstração diminui cada vez que o alvo se afasta da câmara.
Casal	3 s.	Casal a ver fotografias na máquina		$s=7$ $\sigma_{color} = 6$ $\sigma_{space} = 6$ $\phi_{eq} = 2$ $\sigma = 1$ $\phi_e = 2.5$ $\tau = 1$	Falhas nos contornos nas zonas com pouco contraste ou muita claridade, ruído na camisa
Cavalo	4 s.	Cavalo a andar no hipódromo		$s=30$ $\sigma_{color} = 5$ $\sigma_{space} = 5$ $\phi_{eq} = 2$ $\sigma = 0.6$ $\phi_e = 0.75$ $\tau = 1$	Quando existe maior claridade na imagem os contornos são menos eficazes. Solo e árvores com algum ruído.
Estrada	6 s.	Vídeo do percurso de uma estrada		$s=5$ $\sigma_{color} = 5$ $\sigma_{space} = 5$ $\phi_{eq} = 2$ $\sigma = 0.6$ $\phi_e = 0.75$ $\tau = 1$	Pequenas falhas nos contornos das montanhas.

Table 5.1: Descrição dos vídeos utilizados e o respetivo conjunto de parâmetros associado

Foram recolhidas também algumas imagens para testar o método que, sem o factor do movimento a ter em conta e com um tempo de execução menor em relação aos vídeos, facilita a abstração e permite encontrar o conjunto de parâmetros favorável mais facilmente. A tabela 5.2 mostra algumas das imagens testadas bem como uma breve descrição, o conjunto de parâmetros escolhido, dificuldades encontradas e a própria imagem.

Nome	Tempo	Descrição	Frame	Parâmetros	Dificuldades
Ventoinha	Imagem	Imagem de paisagem com ventoinha eólica		$s=30$ $\sigma_{color} = 5$ $\sigma_{space} = 5$ $\phi_{eq} = 2$ $\sigma = 0.6$ $\phi_e = 0.75$ $\tau = 1$	Pequenas falhas nos contornos das montanhas mais distantes.
Baseball	Imagem	Treinador e jogador num campo de baseball		$s=30$ $\sigma_{color} = 5$ $\sigma_{space} = 5$ $\phi_{eq} = 2$ $\sigma = 0.6$ $\phi_e = 0.75$ $\tau = 1$	Contorno excessivo no chão.

Table 5.2: Descrição das imagens utilizadas e o respetivo conjunto de parâmetros associado

Vídeo “Conversa”

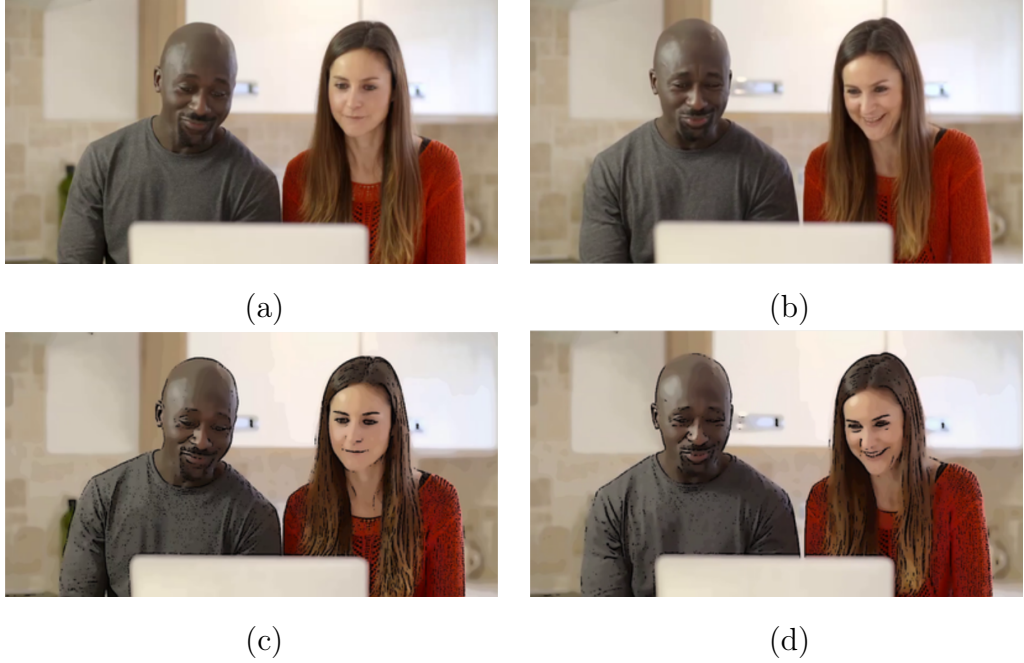


Figure 5.1: Dois frames originais retirados do vídeo "Conversa"[12] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=7$; $\sigma_{color}=4$; $\sigma_{space}=4$; $\phi_{eq}=2$; $\sigma=1$; $\phi_e=4.9$; $\tau=0.993$

Neste exemplo, temos um vídeo de 2 pessoas a falar em frente a um computador. Destaca-se em cada frame as expressões faciais das pessoas, que requerem maior atenção ao detalhe na sua abstração. Como tal, o resto da imagem acaba por ter algumas falhas por terem características diferentes: camisolas quase lisas acabam por detetar algum contorno indesejado e o fundo tem manchas desnecessárias devido à sua monotonia.

Em relação à escolha dos parâmetros foi dada maior prioridade à abstração das caras das pessoas por se tratar do foco principal da imagem. Os valores do σ_{color} e σ_{space} são reduzidos uma vez que para valores elevados, a imagem fica com pouca nitidez de cores na face das pessoas e os limites dos contornos ficam mal definidos. O parâmetro τ controla os contornos na imagem. Apesar de haver falhas nesta componente por não contornar completamente cada uma das pessoas, o valor definido é o mais adequado para que a face não tenha excesso de detalhes e mantenha a expressão de cada um. Para valores superiores a 0.993, seria necessário reduzir o valor de ϕ_e para compensar o

excesso de contorno. Esta alternativa não tem o resultado desejado uma vez que ao reduzir os excessos de contorno, retira também partes importantes. Para valores do parâmetro ϕ_{eq} elevados torna a imagem demasiado irrealista em relação à luminosidade e sombras da imagem portanto adotou-se um valor baixo. Em relação ao parâmetro σ os resultados mais satisfatórios foram obtidos com $\sigma = 1$, um valor baixo resulta numa imagem com mais ruído indesejado e para valores mais altos o edifício e as pessoas perdem contornos importantes apesar da redução de ruído no chão.

Vídeo “Dança”



Figure 5.2: Dois frames originais retirados do vídeo "Dança"[11] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=5$; $\sigma_{color}=4$; $\sigma_{space}=4$; $\phi_{eq}=2$; $\sigma=1$; $\phi_e=5$; $\tau=0.985$

Neste vídeo de um grupo de pessoas a dançar ao ar livre, existe um maior detalhe ao longo de toda a imagem, tanto no edifício, nas árvores e nas pessoas que se movimentam ao longo do vídeo. Como tal, é necessário ter isto em atenção na escolha dos parâmetros, privilegiando o detalhe, coerência temporal das cores e dos contornos. Os valores responsáveis pelo filtro bilateral são baixos para que a imagem consiga manter os efeitos de sombra e o contraste entre todos os elementos da imagem. O movimento existente na foto não é

tão próximo da câmera como o exemplo da “Conversa” por isso $\tau=1$ acaba por ser excessivo e não permite destacar os detalhes importantes, criando vários pontos de incoerência no vídeo. O valor escolhido consegue salientiar as sombras e contornos tanto nas pessoas em movimento como nos objetos imóveis. Para valores elevados de ϕ_{eq} nota-se alguns efeitos de sombras irrealistas no céu, chão e em partes do corpo das pessoas portanto $\phi_{eq} = 2$ é o valor mais adequado pois mantém o tom de cor nos vários elementos da imagem sem criar efeitos de sombra desnecessários. Como ponto negativo as faces das pessoas têm algumas falhas devido à sua distância da câmera, ao movimento rápido e ao pouco detalhe que permite realçar com os contornos sem prejudicar a imagem.

Vídeo “Trânsito”



Figure 5.3: Dois frames originais retirados do vídeo "Trânsito"[19] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=2$; $\sigma_{color}=3$; $\sigma_{space}=3$; $\phi_{eq}=2$; $\sigma=1$; $\phi_e=5$; $\tau=0.992$

Este vídeo é o mais desafiante para conseguir obter uma boa abstração uma vez que existe muito movimento a uma velocidade demasiado rápida para ter coerência ao longo de todo o vídeo. Os parâmetros escolhidos são os que apresentam um melhor resultado quando os veículos estão parados e

tanto os contornos como as cores estão de acordo com a imagem original. Os parâmetros s , σ_{color} e σ_{space} são mais reduzidos do que nos outros vídeos para que a imagem não fique pouco nítida e os objetos imperceptíveis quando estão muito longe enquanto os valores de $\phi_{eq} = 2$, $\sigma = 1$ e $\phi_e = 5$ são os que alteram menos a imagem original. Com isso, apesar de não haver tantos detalhes nos carros em movimento não existe uma incoerência flagrante que aconteceria com um conjunto de parâmetros diferente.

Vídeo “Casal”

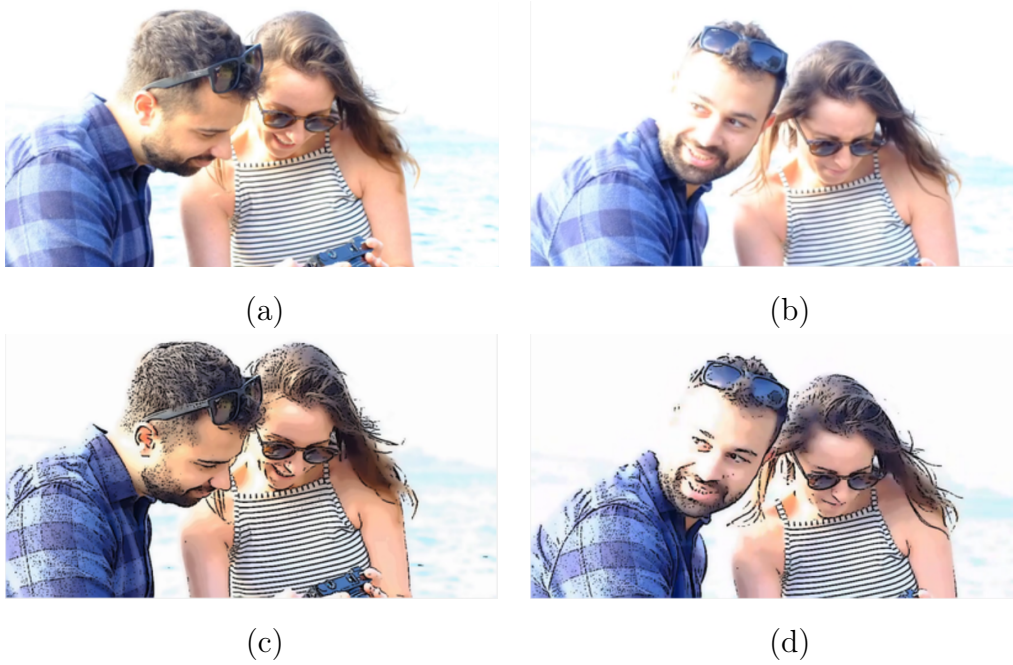


Figure 5.4: Dois frames originais retirados do vídeo "Casal"[17] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=7$; $\sigma_{color}=6$; $\sigma_{space}=6$; $\phi_{eq}=2\sigma=1$; $\phi_e=2.5$; $\tau=1$

Para este vídeo o maior obstáculo no processo de abstração foi o pouco contraste existente na imagem entre as pessoas e o fundo devido à elevada luminosidade. Os parâmetros responsáveis pelos contornos ϕ_e e τ foram determinantes para haver uma boa distinção entre os elementos da imagem. Desta forma, o valor que apresenta melhores resultados foi para $\tau = 1$ uma vez que é o valor máximo deste parâmetro. Para retirar os excessos indesejados, o valor do $\phi_e = 2.5$ é o mais adequado apesar de algumas falhas no reconhecimento das faces das pessoas nos locais com maior claridade como comprova

a figura 5.4d. Os parâmetros s , σ_{color} e σ_{space} têm valores que permitem ao vídeo ter variabilidade no tom de pele de cada pessoa (valores elevados originam mais monotonia de cores) mas não ao ponto de ter manchas irrealistas na face ou nos braços (caso os valores fossem muito baixos).

Vídeo “Cavalo”

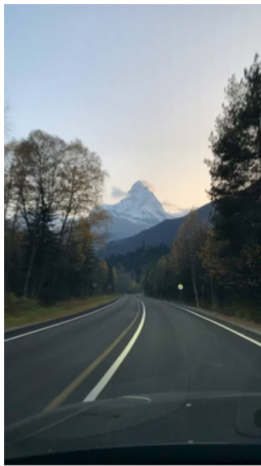


Figure 5.5: Dois frames originais retirados do vídeo "Cavalo" [20] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=30$; $\sigma_{color}=5$; $\sigma_{space}=5$; $\phi_{eq}=2$; $\sigma=0.6$; $\phi_e=0.75$; $\tau=1$

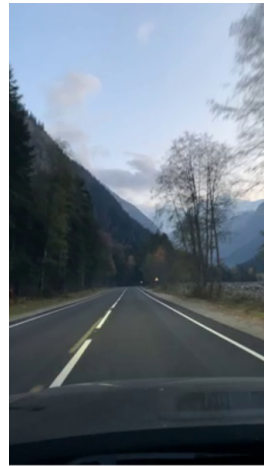
Para este vídeo destaca-se características como o relevo do solo que pode provocar ruído, as árvores no fundo que necessitam de um bom nível de contorno e a claridade que incide no cavalo em certas partes do vídeo. A maior diferença nos valores dos parâmetros comparado com outros exemplos é o valor de s muito elevado. O facto da imagem não ter pormenores relevantes permite aumentar o diâmetro utilizado para o filtro bilateral que permite reduzir algum ruído no solo e aumenta o nível de abstração no cavalo uma vez que um valor reduzido provocada um efeito reduzido. Já σ_{color} e σ_{space} têm valores baixos para não prejudicar a nitidez da imagem. Mais uma vez, o valor de τ é o máximo permitido, ajustando os excessos de contornos através do $\sigma = 0.6$ e, principalmente, o ϕ_e que toma o mínimo possível. O relevo

do solo acaba por ser o maior entrave à abstração por ter vários contornos irrelevantes uma vez que os restantes elementos são facilmente identificados e por cada um ter poder ser representado por poucos tons de cores facilita o processo.

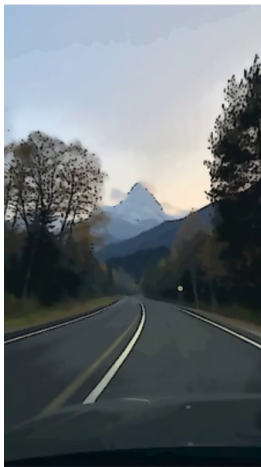
Vídeo “Estrada”



(a)



(b)



(c)



(d)

Figure 5.6: Dois frames originais retirados do vídeo "Estrada" [15] (a,b) e respectivas imagens de abstração (c,d), obtidas com os seguintes parâmetros: $s=5$; $\sigma_{color}=5$; $\sigma_{space}=5$; $\phi_{eq}=2$ $\sigma=0.6$; $\phi_e=0.75$; $\tau=1$

Este vídeo tem a particularidade de ter pouco detalhe ao longo de toda a imagem e ter uma maior monotonia de cores. Assim, os parâmetros s , σ_{color} e σ_{space} têm valores baixos para que seja perceptível a distinção das várias árvores apesar da semelhança de cores. Os parâmetros relativos aos contornos são semelhantes ao vídeo "Cavalo" por ambos terem pouco detalhe. O maior entrave neste exemplo são algumas árvores isoladas e com poucas folhas que aliadas ao movimento existente causa pequenas falhas nos contornos dos ramos.

Imagem "Ventoinha"

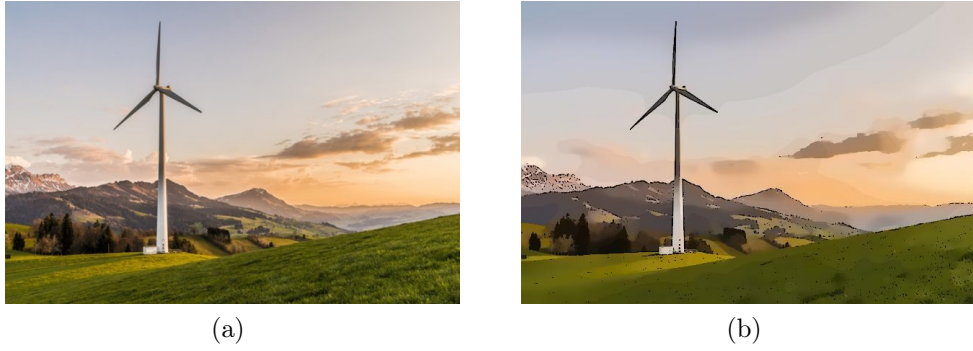


Figure 5.7: Imagem "Ventoinha" [16] original (a) e repetitiva imagem de abstração (b) obtida com os seguintes parâmetros:

$s=30$; $\sigma_{color}=5$; $\sigma_{space}=5$; $\phi_{eq}=2$; $\sigma=0.6$; $\phi_e=0.75$; $\tau=1$

Nesta imagem existe pouco detalhe por se tratar de uma paisagem. Para a abstração o valor do conjunto de parâmetros foi escolhido de modo a obter uma imagem simplificada em relação à cor e detalhe. Deste modo, s toma um valor elevado para ter uma imagem com maior monotonia de cores em cada elemento. Os parâmetros ϕ_e e σ tomam valores muito baixos, seguindo o exemplo acima do vídeo "Cavalo" para evitar ruído indesejado causado por $\tau=1$. Isto resulta num nível de contorno mais baixo mas sem prejudicar a imagem uma vez que existe uma distinção clara entre todas as partes da imagem.

Imagem "Baseball"



Figure 5.8: Imagem "Baseball" [18] original (a) e repetitiva imagem de abstração (b) obtida com os seguintes parâmetros:

$s=30$; $\sigma_{color}=5$; $\sigma_{space}=5$; $\phi_{eq}=2$; $\sigma=0.6$; $\phi_e=0.75$; $\tau=1$

A figura 5.8 mostra outro exemplo onde existe um contraste entre elementos evidente. Assim, aliado ao facto de não haver pormenores difíceis de distinguir, podemos atribuir um valor elevado ao parâmetro s tal como aconteceu no exemplo da figura 5.6. Igualmente, temos o defeito do excesso de contorno no solo em comum devido ao parâmetro τ se encontrar no máximo permitido para que seja possível detetar os contornos importantes nas pessoas da imagem, destacando-se as dobras das roupas.

Critério para escolha dos valores dos parâmetros

Após serem analisadas várias imagens e vídeos para a abstração, foi definido um critério para se proceder à escolha dos valores dos vários parâmetros utilizados no método de abstração de uma forma mais simples e automática. Assim, recorrendo à tabela 5.3 facilmente é possível encontrar um conjunto de parâmetros considerado como mais indicado para a abstração de qualquer imagem, dependendo das suas características.

Parâmetro	Valor			V.Baixo	Critério	V.Alto
	Baixo	tipo	alto			
s	3	7	30	+	Detalhe	-
σ_{color}	2	5	7	+	Contraste de cores	-
σ_{space}	2	5	7	-	Monotonia de cores	+
φ_{eq}	2	2	20	-	Nitidez na transição de píxeis	+
σ	0.5	1	1	+	Detalhe nos contornos	-
φ_e	0.75	5	5	-	Nitidez dos contornos	+
τ	0.98	0.994	1	-	Nível de contorno	+

Table 5.3: Critério-tipo usado para a escolha dos valores dos parâmetros

Para cada parâmetro foi definido um valor baixo, médio e alto, com base na característica da imagem da coluna "Critério" da tabela 5.3. Seguindo o exemplo do parâmetro s , foi utilizado o valor baixo para a maioria das imagens analisadas neste documento. Tratando-se de imagens com vários detalhes importantes, é escolhido o valor baixo do parâmetro que define o diâmetro da vizinhança de cada pixel usado para o filtro bilateral para não perder partes relevantes de elementos da imagem. O vídeo "Cavalo" e as imagens "Baseball" e "Ventoinha", por se tratarem de um vídeo e imagens mais simples, onde os elementos são claramente distinguíveis entre si, adequa-se um valor alto. Destaca-se também o parâmetro φ_{eq} que tem um valor tipo e baixo igual. O efeito que um valor mais alto desta variável produz é demasiado irrealista perante a maioria das imagens, que resultam em manchas indesejadas devido à diferença dos níveis de luminosidade em várias zonas da imagem. O mesmo acontece nos parâmetros σ e φ_e que diminuem substancialmente a quantidade de contorno na imagem para valores abaixo do valor tipo.

Adaptação do método de abstração

As imagens anteriores foram obtidas através do método de Winnemoller et al. [1] à exceção do IBW opcional por não ter sido possível fazer a sua implementação. Com isto, alguns contornos têm falhas nas imagens, através de descontinuidades ou omissão de partes de contornos que seriam corrigidos através do IBW. Como tal, sendo um processo que utiliza um filtro de Sobel e um desfoque Gaussiano, foi implementado um filtro de sobel uma vez que os resultados demonstram ser mais satisfatórios em relação aos contornos do que a anterior implementação da Diferença de Gaussianos. É então aplicado um desfoque Gaussiano na imagem original e de seguida, o filtro de Sobel. O filtro de Sobel é uma operação que utiliza 2 matrizes G_x e G_y (1) que são convoluídas com a imagem original para calcular as aproximações das derivadas. As matrizes G_x e G_y são utilizadas para calcular as variações horizontais e verticais da imagem. Após obter ambos os resultados, a magnitude e a direção do gradiente são dados por (2).

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (5.1)$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (5.2)$$

Para adicionar os contornos à imagem, é definido o valor a partir do qual os píxeis da matriz G são considerados como contornos. É definido então o parâmetro *perc* que permite ao utilizador escolher o nível de contorno. Este valor pode variar entre 0 e 1, sendo a percentagem utilizada na escala [0,255] presente nos píxeis de G (*perc*=0.1 daria o valor máximo de 25.5 para o pixel ser considerado como contorno). Além deste parâmetro, é necessário escolher os valores de *sigmaX*, *sigmaY* e o tamanho do kernel (*ksize*), variáveis utilizadas na função pré-implementada `cv2.filter2D`

Na figura 5.9 encontra-se a abstração da imagem original (a) ,utilizada em [1], feita com 3 abordagens diferentes dentro do método usado por Winnemoller et al. [1]: imagem retirada do artigo (b), a implementação do método sem IBW (c) e com contornos obtidos através do filtro de Sobel (d).

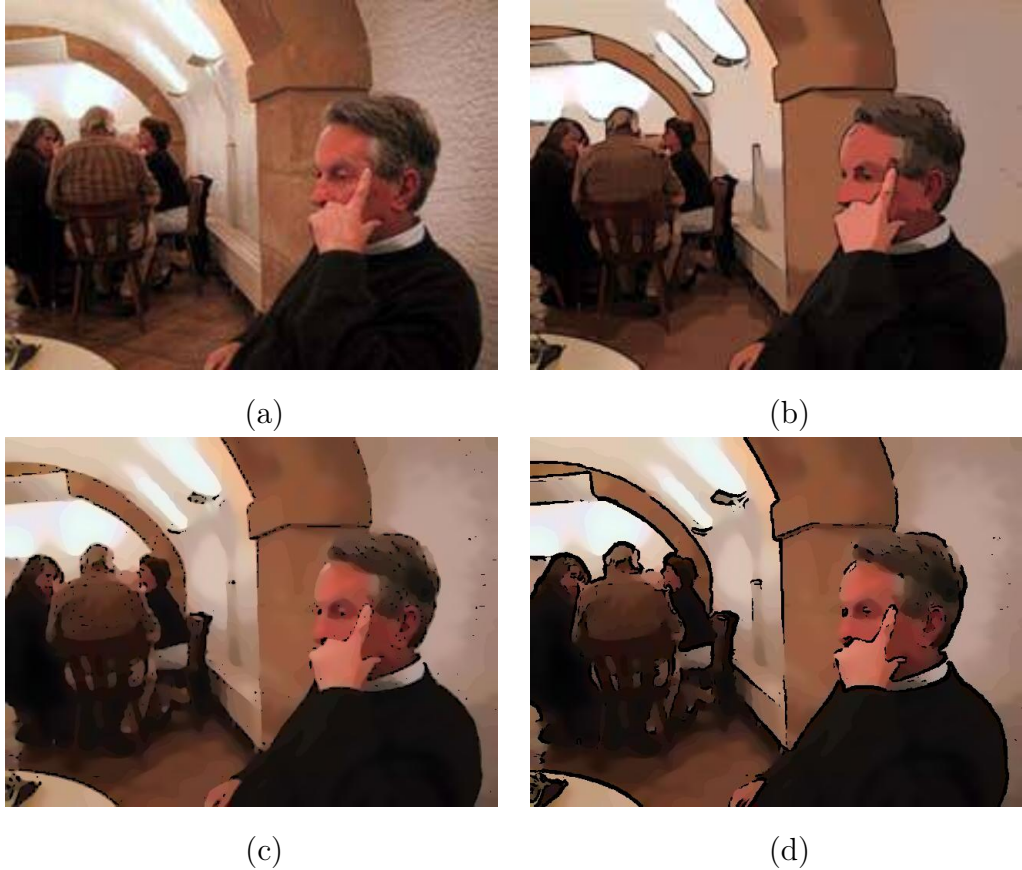


Figure 5.9: Imagem original (a); Imagem retirada de [1] com abstração automática do método usado (b); Resultado da imagem com implementação do método [1] sem IBW (c) e com contornos recorrendo a filtro de Sobel (d)

Nesta imagem, os valores dos parâmetros s , σ_{color} , σ_{space} e φ_{eq} é comum entre as imagens 5.9(c) e (d). O valor de s é 10 para que não haja uma variação da cor nos vários elementos da imagem, seguindo o exemplo em 5.9(b). Os valores das restantes variáveis referentes ao filtro bilateral e quantização de luminância são muito próximas do valor tipo indicado na tabela 5.3 uma vez que valores elevados tornam a imagem pouco nítida ou com manchas demasiado irrealistas. Em relação aos contornos, a imagem 5.9(c) apresenta claramente o pior resultado por omitir várias partes de contornos

importantes além de permitir algum ruído na parede. A abordagem de 5.9(d) corrige grande parte dos erros permitidos aproximando-se muito do método original com IBW. Mesmo assim, existem descontinuidades mínimas nos contornos em relação a 5.9(b).

Capítulo 6

Conclusão

Este trabalho teve como objetivo analisar o método proposto de Winnemoller et al. [1] e se existe alguma relação entre os valores do conjunto de parâmetros escolhidos do resultado mais satisfatório e as características da imagem para que seja possível aplicar o método de forma personalizada de acordo com a preferência do utilizador.

Ao longo da tese, foram analisados vários vídeos e imagens diferentes, de forma a abrangerem uma maior variabilidade de características presentes. Foram escolhidos os valores de cada parâmetro de forma a obter o resultado considerado como mais eficaz para cada caso de forma subjetiva. Foi possível associar os valores obtidos para cada imagem e vídeo com as características presentes em cada uma. Esta abordagem pode ser uma vantagem em relação ao método automático devido ao facto de não ser possível avaliar objectivamente este tipo de resultados, uma vez que o tema de abstração de imagens tem um número elevado de técnicas diferentes, cada uma com critérios e abordagens diferentes. Ao dar a escolha ao utilizador do nível de abstração pretendido para cada imagem, permite a que este proceda à sua avaliação subjetiva e esteja mais perto do resultado pretendido.

O facto de não ter sido possível replicar exatamente o método de Winnemoller et al. [1] acaba por prejudicar vários resultados, em relação aos contornos. O tempo de execução da abordagem sugerida não foi conseguida em tempo real, retirando uma das vantagens deste método em relação a outras técnicas existentes. Apesar disso, a abordagem pretende focar-se na qualidade do resultado da abstração preterindo a velocidade do método uma vez que é adicionada uma vertente manual a um método automático. Com a adaptação final do método, após a escolha do valor dos parâmetros tendo em conta as características da imagem, foi possível obter uma abstração que corrige grande parte das falhas identificadas à implementação anterior e que apresenta poucas diferenças para o método original, sendo discutível qual é

a abordagem mais adequada, estando sempre dependente de uma avaliação subjetiva.

Como trabalho futuro, é proposta uma melhoria em relação ao tempo de execução da implementação do método usado bem como aos contornos da imagem. Para isso, é sugerido replicar a deformação de imagem (IBW) para poder comparar de forma mais precisa ambas as abordagens referidas ao longo do trabalho.

Bibliografia

- [1] Holger Winnemöller, Sven C Olsen, and Bruce Gooch. Real-time video abstraction. *ACM Transactions On Graphics (TOG)*, 25(3):1221–1226, 2006.
- [2] Paul Haeberli. Paint by numbers: Abstract image representations. In *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 207–214, 1990.
- [3] Peter Litwinowicz. Processing images and video for an impressionist effect. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 407–414, 1997.
- [4] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. *ACM transactions on graphics (TOG)*, 21(3):769–776, 2002.
- [5] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.
- [6] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM transactions on graphics (TOG)*, 23(3):679–688, 2004.
- [7] Jan Fischer, Dirk Bartz, and Wolfgang Straber. Stylized augmented reality for improved immersion. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pages 195–202. IEEE, 2005.
- [8] Bruce Gooch, Erik Reinhard, and Amy Gooch. Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics (TOG)*, 23(1):27–44, 2004.
- [9] Henry Kang, Seungyong Lee, and Charles K Chui. Flow-based image abstraction. *IEEE transactions on visualization and computer graphics*, 15(1):62–76, 2008.

- [10] Niall Redmond and John Dingliana. Adaptive abstraction of 3d scenes in realtime. 2007.
- [11] LeonidRaikov. Dance street style street style people dancing. <https://pixabay.com/videos/dance-street-style-street-style-17352/>. [Online; acedido a 22 Março, 2022].
- [12] Videvo. Couple making a video call. <https://www.videvo.net/video/couple-making-a-video-call/7761/>. [Online; acedido a 22 Março, 2022].
- [13] Nur Arad and Craig Gotsman. Enhancement by image-dependent warping. *IEEE Transactions on Image Processing*, 8(8):1063–1074, 1999.
- [14] Joern Loviscach. Scharfzeichner: Klare bilddetails durch verformung. *Computer Technik*, 22(236-237):74, 1999.
- [15] Yaroslav Shuraev. Aventura façanha carro veículo. <https://www.pexels.com/pt-br/video/aventura-facanha-carro-veiculo-4434242/>. [Online; acedido a 22 Março, 2022].
- [16] Pixabay. Moinho de vento branco. <https://pixabay.com/videos/photographer-beach-photography-91/>. [Online; acedido a 22 Março, 2022].
- [17] Hans. Pair trust conversation talk discussion few. <https://pixabay.com/videos/pair-trust-conversation-talk-19706/>. [Online; acedido a 22 Março, 2022].
- [18] Pixabay. Homem ajoelhado no campo de beisebol ao lado do homem. <https://www.pexels.com/pt-br/foto/homem-ajoelhado-no-campo-de-beisebol-ao-lado-do-homem-264337/>. [Online; acedido a 22 Março, 2022].
- [19] MagicTv. Traffic city cityscape urban night illuminated. <https://pixabay.com/videos/traffic-city-cityscape-urban-night-88921/>. [Online; acedido a 22 Março, 2022].
- [20] Can Köksal. Cavalo galopando em um rancho. <https://www.pexels.com/pt-br/video/cavalo-galopando-em-um-rancho-2122952/>. [Online; acedido a 22 Março, 2022].