

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

RAFAEL PACHECO KOVALESKI

**Tone Mapping Reverso de Alta Qualidade
para uma Ampla Gama de Exposições**

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Prof. Dr. Manuel Menezes de Oliveira Neto
Orientador

Porto Alegre, setembro de 2013

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Kovaleski, Rafael Pacheco

Tone Mapping Reverso de Alta Qualidade para uma Ampla Gama de Exposições / Rafael Pacheco Kovaleski. – Porto Alegre: PPGC da UFRGS, 2013.

69 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2013. Orientador: Manuel Menezes de Oliveira Neto.

I. Oliveira Neto, Manuel Menezes de. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Pró-Reitor de Coordenação Acadêmica: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	5
LISTA DE FIGURAS	6
RESUMO	7
ABSTRACT	8
1 INTRODUÇÃO	9
2 CONCEITOS	11
2.1 Faixa Dinâmica	11
2.1.1 Faixa Dinâmica em Câmeras Digitais	11
2.1.2 Captura de Imagens HDR	12
2.1.3 Sistema Visual Humano	13
2.1.4 Visualização de Imagens HDR	13
2.1.5 Tone Mapping	14
2.2 Filtro Bilateral	17
2.2.1 Fast Bilateral Filter	20
2.2.2 Bilateral Grid	20
2.2.3 Real Time Bilateral Filter	21
2.2.4 Adaptive Manifolds	22
3 TRABALHOS RELACIONADOS	24
3.1 Modelos de Descontorno	24
3.1.1 Amplitude Dithering for High Contrast Displays	25
3.1.2 Countouring Removal	25
3.2 Modelos Globais	25
3.2.1 A Power Function Model for Range Expansion	26
3.2.2 Linear Scaling for HDR Monitors	26
3.2.3 Expanding Over-exposed Content	27
3.3 Modelos de Classificação	30
3.3.1 Highlight Generation for HDR Monitors	30
3.3.2 Enhancement of Bright Video Features for HDR Display	31
3.4 Modelos de Mapas de Expansão	32
3.4.1 Non-Linear Expansion using Expand Maps	32
3.4.2 LDR2HDR	34
3.4.3 High-quality Brightness Enhancement Functions	35
3.5 Modelos Dependentes do Usuário	36

4	AVALIANDO OPERADORES DE TONE MAPPING REVERSO	38
4.1	Visual Difference Predictor	38
4.2	HDR Visual Difference Predictor	39
4.3	HDR Visual Difference Predictor 2	40
4.4	Dynamic Range Independent Image Quality Metric	41
5	TONE MAPPING REVERSO PARA UMA AMPLA GAMA DE EXPO- SIÇÕES	44
5.1	Generalização	44
5.2	Simplificação	48
6	RESULTADOS	53
7	CONCLUSÃO	63
	REFERÊNCIAS	64

LISTA DE ABREVIATURAS E SIGLAS

BEF	Brightness Enhancement Function
CRT	Cathodic Ray Tube
CSF	Contrast Sensitivity Function
DRIM	Dynamic Range Independent Image Quality Metric
HDR	Hygh Dynamic Range
JND	Just Noticeable Difference
LCD	Liquid Crystal Display
LDR	Low Dynamic Range
VDP	Visual Differences Predictor
cd/m^2	Candelas por metro quadrado, unidade de medida de luminância

LISTA DE FIGURAS

2.1	Captura de imagens HDR	12
2.2	Monitor Dolby DR37-P	14
2.3	Técnicas de tone mapping global	15
2.4	Técnicas de tone mapping local	16
2.5	Técnicas de tone mapping baseadas no domínio frequência	16
2.6	Técnicas de tone mapping baseadas no domínio gradiente	17
2.7	Exemplo de filtragem bilateral	18
3.1	Exemplos de artefatos de contorno	25
3.2	Classificação de áreas difusas e especulares	30
3.3	Interface da aplicação de DIDYK et al. [2008]	32
3.4	Mapa de expansão	33
3.5	BEFs geradas por REMPEL et al. [2007]	34
3.6	BEFs de alta qualidade	35
3.7	Exemplo de alucinação	37
4.1	Detecção de diferenças pelo VDP	39
4.2	Detecção de diferenças pelo HDR-VDP	40
4.3	Detecção de diferenças pelo HDR-VDP2	42
4.4	Exemplo de distorções detectadas pelo DRIM	43
5.1	Comparação da qualidade das BEFs	45
5.2	Relação entre tons de cinza e máximo entre canais RGB	48
5.3	Vantagens da utilização de tons de cinza	49
5.4	Comparação entre espaços de cor RGB e CIELAB	50
5.5	Demonstração de BEFs com diferentes filtros (1)	51
5.6	Demonstração de BEFs com diferentes filtros (2)	52
6.1	Comparação entre técnicas: série <i>building</i>	56
6.2	Comparação entre técnicas: série <i>graffiti</i>	57
6.3	Comparação entre técnicas: série <i>sunset</i>	58
6.4	Aplicação em frames de vídeo: <i>fireball</i>	59
6.5	Aplicação em frames de vídeo: <i>fireball-smoke</i>	59
6.6	Comparação entre técnicas: série <i>car</i>	60
6.7	Comparação entre técnicas: série <i>pencils</i>	61
6.8	Comparação entre técnicas: série <i>flowers</i>	62

RESUMO

Operadores de tone mapping reverso (RTMOs) realçam imagens e vídeos de baixa faixa dinâmica para visualização em monitores de alta faixa dinâmica. Um problema comum encontrado por operadores anteriores é a maneira com que tratam conteúdo sub ou superexposto. Sob tais condições, eles podem não ser eficientes, e até mesmo causar perda e reversão de contraste visível. Apresentamos uma classe de operadores de tone mapping reverso, baseados no filtro bilateral cruzado (*cross bilateral filter*), capazes de gerar imagens HDR de alta qualidade. Experimentos utilizando uma métrica objetiva de avaliação de imagens demonstram que nosso método é a única técnica capaz de realçar detalhes perceptíveis ao longo de uma grande gama de exposições de imagem, a qual inclui desde imagens subexpostas até imagens superexpostas.

Palavras-chave: Tone mapping reverso, alta faixa dinâmica, filtro bilateral cruzado.

High-Quality Reverse Tone Mapping for a Wide Range of Exposures

ABSTRACT

Reverse tone mapping operators (rTMOs) enhance low-dynamic-range images and videos for display on high dynamic range monitors. A common problem faced by previous rTMOs is the handling of under or overexposed content. Under such conditions, they may not be effective, and even cause loss and reversal of visible contrast. We present a class of local rTMOs based on cross bilateral filtering that is capable of generating high-quality HDR images and videos for a wide range of exposure conditions. Experiments performed using an objective image quality metric show that our approach is the only single technique available that can gracefully enhance perceived details across a large range of image exposures.

Keywords: reverse tone mapping, high dynamic range, cross bilateral filter.

1 INTRODUÇÃO

A tecnologia para visualização de imagens em alta faixa dinâmica (HDR, do inglês *high dynamic range*) tem recebido grande atenção por parte da indústria e do meio acadêmico. Algoritmos para comprimir a faixa dinâmica de imagens para monitores convencionais estão cada vez mais rápidos, e protótipos para dispositivos de visualização de alta faixa dinâmica são cada vez mais comuns. A possibilidade de utilizar imagens HDR para iluminar uma cena em computação gráfica sinteticamente atrai estúdios de cinema, que utilizam esta tecnologia para a criação de cenários realistas em seus filmes. No entanto, dispositivos de visualização convencionais apresentam um contraste muito menor que aquele encontrado em uma cena real. Por esse motivo, operadores de tone mapping (como [PATTANAİK et al., 1998; REINHARD et al., 2002; DURAND; DORSEY, 2002; FATTAL et al., 2002]) são utilizados para comprimir a faixa dinâmica presente em imagens HDR para que possam ser visualizadas em monitores do tipo CRT ou LCD, por exemplo, preservando detalhes que poderiam ser perdidos.

Recentes desenvolvimentos em hardware HDR [HOEFFLINGER, 2007] nos permitem antecipar que monitores com alta faixa dinâmica serão comuns no futuro. Porém, a substituição de monitores de baixa faixa dinâmica (LDR, do inglês *low dynamic range*) por monitores com uma faixa dinâmica estendida introduz um novo problema: adaptar as imagens e vídeos que possuímos atualmente para estes novos dispositivos, tanto para uso profissional quanto pessoal. Diversos pesquisadores propuseram soluções para aumentar a faixa dinâmica deste legado de mídia. Estas técnicas são chamadas de operadores de **tone mapping reverso** (*reverse tone mapping*), e buscam aumentar o contraste e brilho das imagens e vídeos originais, de forma a aprimorar a visualização deste conteúdo. A eficiência destes algoritmos é suportada por estudos recentes [YOSHIDA et al., 2006; SEETZEN et al., 2006; AKYÜZ et al., 2007; BANTERLE et al., 2009], que demonstram que usuários preferem imagens realçadas com estes operadores.

Operadores de tone mapping reverso apresentam uma limitação em comum: nenhum deles é capaz de tratar tanto imagens e vídeos subexpostos quanto superexpostos. Em condições não ótimas, estes operadores podem não aprimorar o contraste percebido, ou até mesmo acarretar uma perda de detalhes. Neste trabalho apresentamos uma nova classe de operadores de tone mapping reverso, baseada no filtro bilateral cruzado (*cross bilateral filter*), capaz de gerar imagens e vídeos HDR de alta qualidade para uma alta gama de condições de exposição.

Um componente importante de operadores de tone mapping reverso recentes é uma função de aumento de brilho (BEF, do inglês *Brightness Enhancement Function*) [REMPPEL et al., 2007; KOVALESKI; OLIVEIRA, 2009], também conhecida como um mapa de expansão [BANTERLE et al., 2006, 2008]. BEFs são geradas através da definição de áreas onde houve perda de informação, principalmente devido à saturação dos sensores

da câmera utilizada para captura da imagem, e um posterior preenchimento destas áreas com uma função de suavização. A qualidade das BEFs influencia significativamente o resultado da operação de tone mapping reverso [KOVALESKI; OLIVEIRA, 2009]. Nossa técnica é capaz de gerar estas funções automaticamente.

As **contribuições** deste trabalho incluem:

- Uma nova classe de operadores de tone mapping reverso para imagens e vídeos que suporta uma alta gama de condições de exposição;
- Uma técnica eficiente para a geração de funções de aumento de brilho de alta qualidade, para operadores de tone mapping reverso, baseada no filtro bilateral cruzado. Nosso método é capaz de utilizar qualquer técnica de aceleração do filtro bilateral, liberando o usuário de limitações de uma implementações específicas.

O restante desta dissertação apresenta-se organizado do seguinte modo:

- O Capítulo 2 apresenta informações gerais sobre faixa dinâmica (seção 2.1) e o filtro bilateral (seção 2.2);
- O Capítulo 3 apresenta uma revisão sobre trabalhos de tone mapping reverso, além de mostrar como são classificados de acordo com suas características;
- O Capítulo 4 apresenta a evolução de técnicas capazes de detectar diferenças em imagens. Estas técnicas são utilizadas por operadores de tone mapping e tone mapping reverso para avaliar a qualidade de seus resultados;
- O Capítulo 5 apresenta o método desenvolvido para generalizar e simplificar a técnica de KOVALESKI; OLIVEIRA [2009], mostrando comparações entre a técnica desenvolvida e resultados anteriores;
- O Capítulo 6 discute os resultados da técnica desenvolvida, comparando-os com métodos recentes;
- Finalmente, o Capítulo 7 sumariza os principais pontos apresentados nesta dissertação.

2 CONCEITOS

Neste capítulo serão apresentados pontos importantes para um bom entendimento do trabalho aqui apresentado. A seção 2.1 apresenta a idéia de faixas dinâmicas, um importante conceito em trabalhos envolvendo tone mapping e tone mapping reverso, enquanto a seção 2.2 introduz o filtro bilateral, componente essencial da técnica exibida nesta dissertação.

2.1 Faixa Dinâmica

O termo faixa dinâmica (em inglês, *dynamic range*) é utilizado para descrever a razão entre o maior e o menor valor possível de uma determinada quantidade. No escopo deste trabalho, faixa dinâmica é a razão entre a maior e a menor luminância presentes em uma imagem.

A luminância de uma cena pode ser descrita através de valores de candelas por metro quadrado (cd/m^2). Cenas reais geralmente apresentam uma grande faixa de luminância, cobrindo aproximadamente 14 ordens de magnitude, variando entre a luz do sol direta (de $10^5 cd/m^2$ a $10^8 cd/m^2$) e a luz de estrelas (de $10^{-3} cd/m^2$ a $10^{-6} cd/m^2$). Em um ambiente digital, a razão N entre a maior e menor luminância é chamada de **contraste**, e é expressa pela notação $N : 1$.

Imagens digitais tradicionais não armazenam valores de luminância reais, armazenando em seu lugar a informação de cor da cena em três canais de 8 bits. Estas imagens, incapazes de guardar toda a informação de faixa dinâmica da cena capturada, são chamadas de **imagens de baixa faixa dinâmica** ou **imagens LDR** (do inglês *Low Dynamic Range*). Por outro lado, imagens capazes de armazenar toda ou quase toda a informação de luminância de uma cena real, através de valores de ponto flutuante, são chamadas de **imagens de alta faixa dinâmica** ou **imagens HDR**.

2.1.1 Faixa Dinâmica em Câmeras Digitais

A luminância de uma cena, em dispositivos de captura digitais, é calculada através de células fotoelétricas que medem a quantidade de fótons que as atingem. Podemos pensar nessas células como recipientes que armazenam fótons como se fossem gotas de água. Assim, se a célula fotoelétrica enche, não terá mais espaço para novos fótons. Este estado define o ponto branco da câmera. Em uma câmera ideal, o contraste seria simplesmente o número máximo de fótons em uma célula fotoelétrica dividido pelo mínimo valor possível de medir, chamado de ponto preto. Ou seja: se cada célula fotoelétrica pudesse conter 1000 fótons, o contraste seria de 1000 : 1.

O problema é que câmeras convencionais não conseguem contar apenas um fóton.

Cada célula fotoelétrica está sujeita a diversos tipos de ruído, como o aquecimento dos circuitos (chamado ruído térmico) ou o vazamento de fótons de células vizinhas (chamado *blooming*), que alteram o valor mínimo possível do ponto preto. Por esse motivo, a faixa dinâmica real de uma câmera digital é descrita como a razão entre a máxima intensidade de luz (ou seja, o ponto de saturação da célula fotoelétrica) e a menor intensidade de luz (acima do ruído dos circuitos) possíveis de se medir.

A unidade mais comumente utilizada para medir a faixa dinâmica em câmeras digitais é o *f-stop*, que descreve o alcance total da luz em potências de dois. Assim, um contraste de 1024 : 1, por exemplo, pode também ser representado como possuindo uma faixa dinâmica de 10 *f-stops* (pois $2^{10} = 1024$). Dependendo da aplicação, um *f-stop* também pode ser chamado de **zona** ou *eV* (*exposure value*).

2.1.2 Captura de Imagens HDR

Para capturar imagens HDR são necessárias técnicas de composição de imagens (Figura 2.1), pois até mesmo câmeras fotográficas comerciais mais recentes não são capazes de capturar toda a informação de faixa dinâmica de uma cena real, sendo limitadas a imagens de 8 bits ou imagens de 16 bits em formato RAW. A primeira menção de técnicas para gerar imagens HDR foi feita por Charles Wyckoff, que criou filmes fotográficos de três camadas para capturar diferentes exposições de uma mesma fotografia [WYCKOFF; FEIGENBAUM, 1962].

Um dos maiores problemas encontrados na composição das imagens é que câmeras digitais e filme fotográfico não têm uma resposta linear à luminância capturada. Por este motivo, técnicas de composição precisam estimar a **função de resposta da câmera** (chamada **CRF**, *Camera Response Function*). Esta função é, então, utilizada para converter o valor dos pixels para seus valores de luminância correspondentes.

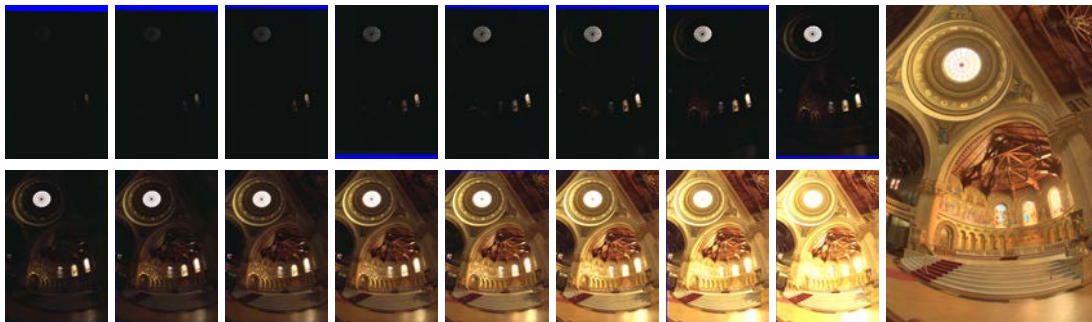


Figura 2.1: Diferentes exposições utilizadas para gerar uma imagem HDR. A imagem à direita corresponde ao resultado do operador de tone mapping de REINHARD et al. [2002]. Observe os detalhes mantidos nas janelas ao fundo e na clarabóia no topo da imagem.

MANN [1993] descreve um método simples para estimar a CRF utilizando técnicas de *bracketing* (fotos da mesma cena capturadas com configurações diferentes da câmera fotográfica), ajustando os valores de pixels de diferentes exposições a uma CRF fixa, $CRF(x) = ax^\gamma + b$. No entanto, esta CRF é limitada e não corresponde à maioria das CRFs reais.

Em 1997, Paul Debevec e Jitendra Malik desenvolveram uma nova maneira de gerar imagens HDR a partir de imagens de baixa faixa dinâmica [DEBEVEC; MALIK, 1997]. A técnica utilizada é similar à de Steve Mann, porém o método utilizado para capturar a

curva de resposta da câmera é mais robusto. MITSUNAGA; NAYAR [1999] aperfeiçoaram este método gerando uma representação polinomial da CRF, enquanto ROBERTSON et al. [1999, 2003] propõem uma alternativa capaz de calcular a CRF utilizando um método iterativo, além de utilizar uma ponderação probabilística para estimar os pesos de cada exposição em uma seqüência de imagens.

KANG et al. [2003] exploram métodos para captura de vídeos em HDR utilizando múltiplas exposições. Uma câmera programada para alterar a velocidade de captura a cada frame é utilizada, alinhando e deformando frames correspondentes a diferentes velocidades de captura e compondo-os para recuperar um frame HDR. No entanto, este método não é capaz de mostrar objetos movendo-se em alta velocidade sem introduzir artefatos de visualização. Além disso, a técnica é capaz de capturar apenas cerca de 15 frames por segundo. NAYAR; BRANZOI [2003] desenvolveram uma câmera com faixa dinâmica adaptável, onde um modulador de luz controlável de cristal líquido é colocado em frente à câmera. Este modulador adapta a exposição de cada pixel, permitindo a captura de cenas com altas faixas dinâmicas.

Algumas empresas dispõem de câmeras baseadas no método de múltiplas exposições para captura de imagens HDR, como a SpheroCam HDR [SPHEROCAM], da Spheron-VR, e Panoscan MK-3 [PANOSCAN], da Panoscan Ltd. Estas câmeras capturam imagens esféricas ($360^\circ \times 180^\circ$) e são capazes de alcançar uma faixa dinâmica muito alta. A SpheroCam HDR, por exemplo, é capaz de capturar até 26 *f-stops* de faixa dinâmica em uma única varredura, com 50 *megapixels*.

A alternativa para câmeras baseadas em múltiplas exposições é o uso de sensores CCD (*Charge-Coupled Device*, dispositivo de carga acoplada) capazes de capturar valores HDR. O problema destes sensores é que atualmente ainda têm baixa resolução ($\sim 640 \times 480$) e apresentam imagens com bastante ruído. Por este motivo, suas aplicações são voltadas ao uso industrial, como segurança e automação.

2.1.3 Sistema Visual Humano

O sistema visual humano, embora capaz de se adaptar a uma grande faixa dinâmica em uma cena, não consegue visualizar mais que uma pequena região desta faixa em um determinado instante. Pesquisas diferem no valor exato, mas a maioria concorda em um limite de contraste em torno de 150 : 1 [SEETZEN et al., 2004]. Acima deste valor, arestas com uma diferença mínima de contraste parecem borradas e indistintas, e o olho não consegue julgar as magnitudes das regiões adjacentes.

O processo de adaptação do olho humano move a faixa dinâmica perceptível de acordo com a luminância da cena observada. Porém, este processo não é instantâneo: sair de um ambiente claro para um ambiente muito escuro pode levar alguns minutos de adaptação, enquanto a situação inversa é mais rápida, levando apenas alguns segundos para a visão se adaptar totalmente [LEDDA et al., 2004].

2.1.4 Visualização de Imagens HDR

Dispositivos de visualização, assim como a visão humana, não conseguem mostrar toda a faixa dinâmica de uma cena real. Porém, ao contrário da nossa visão, eles não conseguem se adaptar a diferentes condições de iluminação. Monitores CRT, LCD e projetores apresentam um contraste de cerca de 300 : 1 [SEETZEN et al., 2004]. Alguns monitores, porém, são capazes de trabalhar com uma faixa dinâmica muito maior. Por exemplo, o modelo DR37-P (Figura 2.2), da Brightside Technologies, empresa posteriormente adquirida pela Dolby [DOLBY], é capaz de operar com contraste de até 200.000 : 1

[BITTECH].



Figura 2.2: Representação do monitor Dolby DR37-P, desenvolvido pela Brightside Technologies. As imagens inferiores demonstram as capacidades de contraste de um monitor de alta definição convencional (esquerda) e de um monitor HDR (direita). Fonte: [BITTECH].

A alternativa mais simples para comprimir a faixa dinâmica de uma imagem é mapear linearmente os valores da imagem para o monitor, utilizando um passo fixo para a quantização de toda a faixa de valores. Porém, devido à resposta não-linear à intensidade luminosa de uma cena por nossos olhos, é possível perceber artefatos devido a esse mapeamento, principalmente nas partes escuras da imagem. Uma alternativa ao mapeamento linear é utilizar a correção gamma [CGSD], que mapeia os valores não-linearmente, de acordo com o valor de luminância no pixel da imagem. No entanto, para imagens HDR, a correção gamma não irá produzir uma boa distribuição de valores [REINHARD et al., 2005]. Por esse motivo, operadores de tone mapping são utilizados para melhorar o resultado da compressão de faixa dinâmica [TUMBLIN; RUSHMEIER, 1993].

2.1.5 Tone Mapping

Operadores de tone mapping buscam gerar uma imagem o mais fiel possível à percepção da cena real utilizando a faixa dinâmica disponível no dispositivo de visualização desejado. De acordo com REINHARD et al. [2005], algoritmos de tone mapping podem se enquadrar em quatro tipos principais: (i) operadores globais, (ii) operadores locais, (iii) operadores de domínio frequência e (iv) operadores de domínio gradiente. Os dois primeiros mapeiam tons de acordo com a percepção de luminância do olho humano, geralmente baseados no *framework* proposto por TUMBLIN; RUSHMEIER [1993], enquanto os dois últimos baseiam-se em modelos matemáticos para melhorar o nível de detalhes em áreas comprimidas. Discutiremos brevemente estes operadores a seguir.

2.1.5.1 Operadores Globais

Nesta categoria se enquadram o mapeamento linear dos valores da imagem, além do mapeamento com correção gamma. Os algoritmos de tone mapping globais tendem à simplicidade e velocidade, e muitos são rápidos o suficiente para permitir aplicações em tempo real, ao custo da perda de detalhes nas áreas comprimidas. Alguns exemplos se encontram na Figura 2.3.

DRAGO et al. [2003], por exemplo, estendem as curvas de resposta logarítmicas para utilizar uma faixa dinâmica maior. A base do logaritmo varia entre 2 e 10, baseado em regiões com menor ou maior brilho na imagem. Esta variação preserva o contraste em regiões mais escuras e comprime regiões mais claras.

Já REINHARD et al. [2002] basearam seu operador de tone mapping global em técnicas utilizadas em fotografia, especificamente o sistema de zonas, formulado por Ansel Adams e Fred Archer por volta de 1940 e popularizado mais tarde por Minor White [WHITE; ZAKIA; LORENZ, 1976]. Sua técnica mapeia a luminância média da imagem para a média intensidade do monitor a ser utilizado.



Figura 2.3: Tone mapping global por DRAGO et al. [2003] (esquerda) e REINHARD et al. [2002] (direita).

2.1.5.2 Operadores Locais

Operadores de tone mapping locais ajustam a luminância de cada pixel de acordo com sua vizinhança. Desta maneira, são capazes de comprimir faixas dinâmicas maiores e manter mais detalhes, em comparação com operadores globais (Figura 2.4). Porém, devido à complexidade dos processos envolvidos, muitos operadores locais de tone mapping não podem ser executados em tempo real, seja devido ao custo computacional ou pela necessidade de configurar manualmente parâmetros para a geração da imagem final.

PATTANAIAK et al. [1998] baseiam seu operador de tone mapping em um modelo do sistema visual humano. Seu modelo incorpora uma representação de diferentes escalas de luminância, padrões e cores que causam mudanças na iluminação de uma cena, utilizando os dados processados por esse modelo para gerar uma imagem com uma faixa dinâmica menor.

REINHARD et al. [2002] apresentam um operador de tone mapping local baseado em técnicas utilizadas em fotografia. Para o caso local, a técnica simulada foi dodging-and-burning. Um filtro que varia espacialmente, de acordo com a intensidade do pixel na imagem, melhora o contraste de detalhes, utilizando um modelo do sistema visual humano. Ao contrário de outros operadores de tone mapping locais, a solução encontrada

neste trabalho não produz halos luminosos perto de regiões de alta frequência iluminadas. Esta afirmação foi provada com inúmeros testes apresentados em seu trabalho, utilizando outros operadores e diversas imagens HDR.



Figura 2.4: A mesma imagem HDR com sua faixa dinâmica comprimida por dois operadores de tone mapping local diferentes: PATTANAIAK et al. [1998] (esquerda) e REINHARD et al. [2002] (direita).

2.1.5.3 Operadores em Domínio Frequência

O conceito de que uma imagem pode ser considerada como uma composição de uma componente LDR com altas frequências espaciais e uma componente HDR de baixa frequência foi estudado por DURAND; DORSEY [2002] (Figura 2.5). Se estas componentes podem ser separadas, apenas os elementos de baixa frequência precisam ser comprimidos. Seu operador de tone mapping utiliza o filtro bilateral [TOMASI; MANDUCHI, 1998], que é uma operação que suaviza a imagem (*blur*), mantendo as arestas intactas. Mais informações sobre este filtro serão apresentadas na seção 2.2.

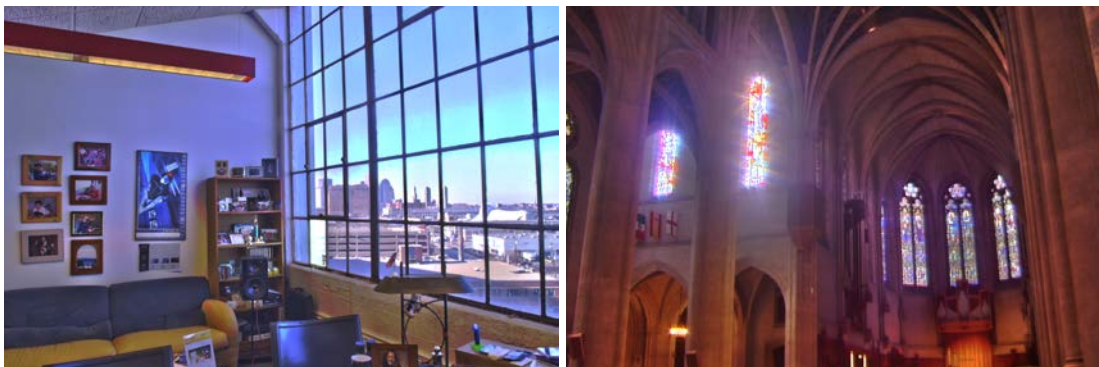


Figura 2.5: Imagens LDR geradas pelo algoritmo de DURAND; DORSEY [2002]. A percepção da cena não é a mesma por um observador humano: o método aplicado realça os detalhes, sem modelar o sistema visual de uma pessoa.

2.1.5.4 Operadores em Domínio Gradiente

Altas frequências em uma imagem introduzem fortes mudanças no gradiente da luminância, enquanto baixas frequências tendem a ser mais suaves. Com base nessa observação, FATTAL et al. [2002] criaram seu operador de tone mapping, identificando gradientes altos em diversas escalas e atenuando sua magnitude, porém mantendo sua direção

(Figura 2.6). A atenuação é progressiva: quanto maior a magnitude do gradiente, maior a compressão. A partir deste mapa de gradientes atenuado a imagem é reconstruída.



Figura 2.6: Tone mapping por FATTAL et al. [2002]. Alterando a magnitude dos gradientes altos de uma imagem, é possível reduzir a faixa dinâmica de uma imagem. Porém, a percepção da cena é diferente daquela de um observador humano.

2.2 Filtro Bilateral

Um filtro bilateral é um filtro não-linear capaz de borrar uma imagem, respeitando arestas proeminentes, através da combinação de filtragem espacial e filtragem pela intensidade dos pixels: a saída de uma filtragem simples no domínio espacial é regulada por uma segunda filtragem, que é calculada sobre a intensidade dos pixels de sua vizinhança. Assim, pixels que apresentam uma grande diferença de intensidade e/ou distância terão pouca ou nenhuma influência sobre a filtragem do pixel atual. A Figura 2.7 demonstra um exemplo de filtragem utilizando o filtro bilateral.

Este filtro surgiu em 1995 sob o nome de *nonlinear Gaussian filter* (filtro Gaussiano não-linear) [AURICH; WEULE, 1995]. Mais tarde, TOMASI; MANDUCHI [1998] chamaram este filtro de *bilateral filter* (filtro bilateral), nome pelo qual é mais conhecido atualmente.

Embora a função mais utilizada para a filtragem bilateral seja a Gaussiana, qualquer função de decaimento pode ser aplicada à equação. No entanto, o uso de outras funções introduz poucas mudanças nos resultados [DURAND; DORSEY, 2002]. Para uma função Gaussiana G como função de decaimento, domínio espacial S e domínio intensidade R (*Range*), o resultado I^b da passagem do filtro bilateral em uma imagem I é definido pela Equação 2.1:

$$I_p^b = \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q$$

$$W_p^b = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|)$$
(2.1)

onde p é o pixel sendo filtrado, q é um pixel da vizinhança de p , $\|p - q\|$ é a distância Euclidiana entre p e q , I_p e I_q são os valores dos pixels p e q na imagem I , σ_s e σ_r são os desvios padrões das funções Gaussianas para o domínio espacial e de intensidade, respectivamente, e W_p^b é um fator de normalização.

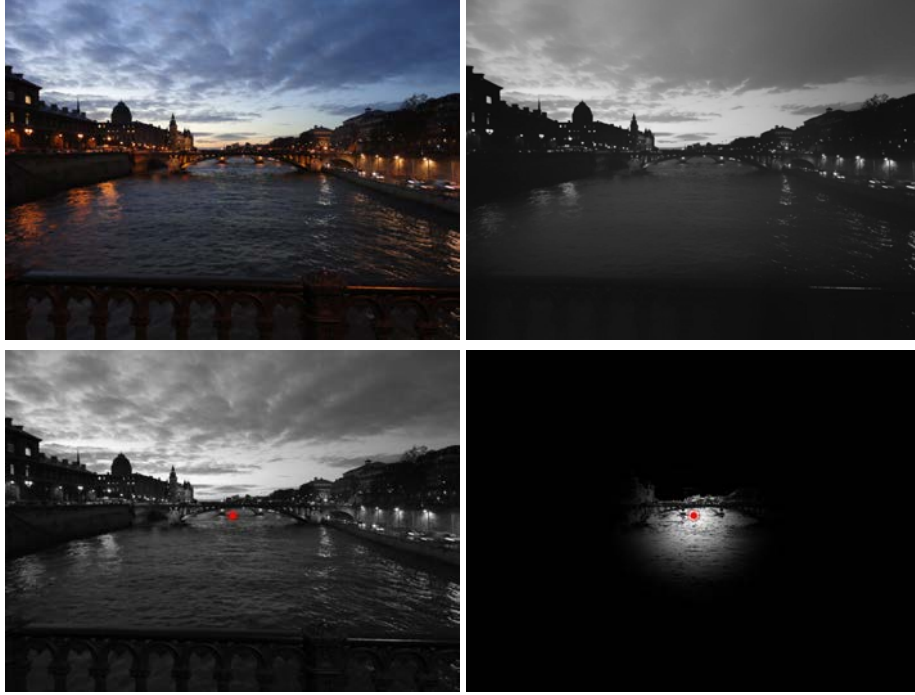


Figura 2.7: Exemplo de aplicação do filtro bilateral, com $\sigma_s = 50$ e $\sigma_r = 0.1$. **Topo à esquerda.** Imagem a ser filtrada em tons de cinza. **Abaixo à esquerda.** Imagem de entrada em tons de cinza. O ponto destacado corresponde ao mesmo pixel na imagem à direita. **Abaixo à direita.** Exemplo da influência dos vizinhos do pixel em destaque no valor final da filtragem. Observe que tons similares e pixels mais próximos têm maior influência no resultado. **Topo à direita.** Imagem em tons de cinza filtrada. Observe como as arestas são mantidas, embora o valor de σ_s seja relativamente grande.

Em algumas aplicações de fotografia computacional, às vezes é necessário separar a informação da imagem I , a ser filtrada, da imagem E , que define as arestas que devem ser preservadas. Por exemplo, em um cenário com/sem flash [EISEMANN; DURAND, 2004; PETSCHNIGG et al., 2004], uma imagem P_{sf} é tirada de um ambiente sem flash e outra imagem P_f é tirada com flash. Filtrar P_{sf} é difícil devido à quantidade de ruído que uma imagem com pouca luz possui. Para resolver este problema, EISEMANN; DURAND [2004] e PETSCHNIGG et al. [2004] introduziram o filtro bilateral cruzado (*cross* ou *joint bilateral filter*) como uma variação do filtro bilateral comum. Esta técnica é capaz de filtrar a imagem sem flash $P_{sf} = I$, enquanto se utiliza da versão com flash $P_f = E$ para localizar as arestas que devem ser preservadas. A definição do filtro bilateral cruzado é similar à Equação 2.1, exceto que E substitui I na Gaussiana de intensidades (G_{σ_r}):

$$\begin{aligned}
 I_p^c &= \frac{1}{W_p^c} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) I_q \\
 W_p^c &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|)
 \end{aligned} \tag{2.2}$$

O filtro bilateral tem sido usado extensamente em diversas áreas como redução de ruído [BUADES; COLL; MOREL, 2005], edição de texturas e reiluminação [OH et al., 2001], tone mapping [DURAND; DORSEY, 2002; PETSCHNIGG et al., 2004] e estilização [WINNEMÖLLER; OLSEN; GOOCH, 2006]. No entanto, devido à natureza

não-linear deste filtro, seu uso em aplicações em tempo real é limitado. Algumas técnicas permitem que a saída do filtro seja calculada em velocidades interativas através de uma aproximação do resultado real. Uma breve revisão destes métodos será apresentada nesta seção, e alguns algoritmos serão expostos mais extensamente devido a suas contribuições.

DURAND; DORSEY [2002] linearizaram o filtro bilateral, o que fez possível utilizar a transformada de Fourier para aplicar o filtro no domínio frequência. Eles também diminuem a resolução dos dados de entrada (*downsampling*) para acelerar a computação da filtragem.

PHAM; VLIET [2005] descrevem um filtro bilateral com kernel separável. Aplicando o filtro primeiramente na dimensão X e, em seguida, aplicando o mesmo filtro na dimensão Y (ou vice-versa) no resultado do passo anterior, foi possível acelerar sua execução em até 3.5 vezes para kernels 2D. Porém, esta aproximação depende de kernels pequenos para obter um bom resultado. Para kernels maiores, artefatos devido à complexidade das cenas podem ser notados, principalmente devido ao fato do filtro considerar linhas e colunas separadamente [PARIS; DURAND, 2006].

WEISS [2006] descreve um método que calcula o resultado exato do filtro bilateral utilizando uma função box para definir a influência espacial dos pixels. Essa implementação é baseada em seu algoritmo para o *fast median filter*. Embora este método tenha um bom desempenho, sua transformada de Fourier e consequente inversa introduzem efeito *Mach band* devido ao uso de uma função box. O autor utiliza um método iterativo para corrigir este problema, utilizando sempre a imagem original na filtragem de intensidades, o que, segundo ele, converge para uma função Gaussiana sem induzir uma aparência cartunesca.

PARIS; DURAND [2006] criaram o algoritmo *fast bilateral filter*, baseado no processamento de sinais em espaços de mais alta dimensionalidade. Uma imagem I é armazenada em uma estrutura de dados em três dimensões, onde as dimensões X e Y correspondem às posições espaciais dos pixels e a dimensão Z corresponde à intensidade do pixel na posição (x, y) de I . Estudaremos este método mais profundamente na seção 2.2.1.

O *bilateral grid* [CHEN; PARIS; DURAND, 2007] é uma generalização da estrutura apresentada no trabalho de PARIS; DURAND [2006]. Esta generalização permite que operações sensíveis a arestas sejam aplicadas facilmente à imagem de entrada. A seção 2.2.2 descreve este método com mais detalhes.

PORIKLI [2008] demonstrou que o filtro bilateral pode ser computado com complexidade $O(1)$ em relação ao tamanho do kernel para três tipos de filtro bilateral: (1) Kernel box espacial e kernel de intensidade arbitrário, utilizando um histograma integral para evitar operações redundantes. (2) Kernel arbitrário espacial e kernel de intensidade polinomial, que pode ser interpretado como uma soma ponderada da filtragem espacial da imagem de entrada e suas potências. (3) Kernel espacial e de intensidades arbitrários, em que séries de Taylor são utilizadas para aproximar a função Gaussiana de intensidades até a quarta derivada. Para variâncias Gaussianas pequenas, no entanto, este método apresenta uma péssima aproximação do resultado, em relação à saída exata da filtragem.

Baseando-se no trabalho de DURAND; DORSEY [2002], um filtro bilateral de complexidade $O(1)$ para kernels arbitrários foi demonstrado por YANG; TAN; AHUJA [2009]. As intensidades contidas na imagem são discretizadas em um número de valores e um filtro linear é computado para cada valor. O valor final é calculado através de uma interpolação entre os dois filtros mais próximos. A seção 2.2.3 apresenta este método mais extensamente.

Para acelerar a interpolação utilizada por técnicas como o *fast bilateral filter* [PARIS;

DURAND, 2006] e o *bilateral filter* [CHEN; PARIS; DURAND, 2007], ADAMS; BAEK; DAVIS [2010] propõem a amostragem dos espaços multidimensionais criados por estas técnicas nos vértices de uma grade permutohédrica (*permutohedral lattice*). Esta grade é composta de *simplices* idênticos, e o *simplex* que contorna um ponto pode ser achado por um algoritmo simples de arredondamento. A distribuição dos pontos no espaço multidimensional e sua posterior amostragem podem ser, então, aplicados através de interpolação baricêntrica, que é exponencialmente mais rápida que a utilizada em trabalhos anteriores.

Finalmente, GASTAL; OLIVEIRA [2012] utilizam variedades (*manifolds*) que se adaptam ao conteúdo da imagem. As amostras do sinal (imagem) são então projetadas nestes manifolds, filtradas, e em seguida reamostradas para obter a imagem final. Esta técnica será apresentada com maiores detalhes na seção 2.2.4.

2.2.1 Fast Bilateral Filter

PARIS; DURAND [2006] reinterpreta o filtro bilateral como uma convolução em um espaço de maior dimensionalidade, considerando os valores de cada pixel como uma nova dimensão da imagem (ao lado de sua posição). Uma posterior amostragem (operação chamada *slicing*) extrai os dados da imagem final deste espaço.

Esta técnica introduz a noção de intensidade homogênea. Valores de intensidade homogênea são definidos como um par (wV, w) , onde w é o peso relacionado a um valor V . No caso do filtro bilateral, w corresponde ao fator de normalização (W_p^b na Equação 2.1). Assim, é possível restaurar o valor V pela divisão wV/w . O nome "intensidade homogênea" foi inspirado nas coordenadas homogêneas da geometria projetiva, que utiliza o mesmo conceito. [PARIS; DURAND, 2006] mostraram que a Equação 2.1 pode ser expressa como uma convolução por um kernel 3D, seguida de duas operações não-lineares. As operações não-lineares são uma divisão pelo fator de normalização contido no vetor de intensidade homogênea e uma operação chamada de *slicing*, que corresponde à amostragem do espaço tridimensional nas coordenadas $(x, y, I(x, y))$, onde I é a intensidade da imagem de entrada do filtro na posição (x, y) .

$$\text{linear: convolução 3D} \quad (w^b i^b, w^b) = g_{\sigma_s, \sigma_r} \otimes (wi, w) \quad (2.3)$$

$$\text{não linear: slicing+divisão} \quad I_p^b = \frac{(w^b(p, I_p) i^b(p, I_p))}{w^b(p, I_p)} \quad (2.4)$$

Além da possibilidade de utilizar a convolução no domínio frequência, também é possível diminuir a resolução do espaço 3D onde a imagem é armazenada e filtrada. Durante a operação de *slicing*, a resolução pode ser restaurada utilizando interpolação trilinear.

2.2.2 Bilateral Grid

O *bilateral grid* [CHEN; PARIS; DURAND, 2007] surgiu a partir da estrutura de dados proposta no *fast bilateral filter* [PARIS; DURAND, 2006]. Esta estrutura permite operações sensíveis a arestas devido à maneira com que a imagem de entrada é tratada.

O uso do *bilateral grid* consiste de três passos: (i) criação do *bilateral grid*, (ii) processamento dos dados e (iii) *slicing*. Os parâmetros de entrada para a criação do *bilateral grid* são um fator de amostragem para o domínio espacial (s_s) e um fator de amostragem para o domínio intensidade (s_r). Assim, para uma imagem de entrada I , com tamanho (x_{max}, y_{max}) e com intensidade máxima v_{max} , o *bilateral grid* resultante terá um tamanho $(x_{max}/s_s, y_{max}/s_s, v_{max}/s_r)$.

Antes de utilizarmos o *bilateral grid*, inicializamos todas as suas células com $(0, 0)$. No *bilateral grid*, é importante guardar o número de pixels em uma célula. Para isso,

utilizam-se quantidades homogêneas no formato (wV, w) , onde w é o número de pixels ou peso, e V um valor escalar. Caso deseja-se usar cores, o vetor passa a ser (wR, wG, wB, w) , onde R , G e B correspondem aos canais RGB de uma imagem. Com esta estrutura, podemos facilmente calcular médias ponderadas: $(w_1V_1, w_1) + (w_2V_2, w_2) = (w_1V_1 + w_2V_2, w_1 + w_2)$; após normalizarmos o resultado por $(w_1 + w_2)$ temos o valor da média ponderada entre os valores V_1 e V_2 , com pesos w_1 e w_2 .

Para preencher o *bilateral grid* precisamos inserir o valor do pixel na célula correta. Assim, para cada pixel (x, y) de uma imagem I , um *bilateral grid* Γ , fator de amostragem espacial s_s e fator de amostragem de intensidade s_r :

$$\Gamma([x/s_s], [y/s_s], [I(x, y)/s_r]) + = (I(x, y), 1)$$

onde $[.]$ é a operação de inteiro mais próximo e a função $I(x, y)$ corresponde ao valor do pixel da imagem I na posição (x, y) . Notem que acumulamos a intensidade do pixel e o número de pixels de cada célula, respeitando o conceito de intensidades homogêneas.

Para o processamento de dados, qualquer função que recebe como entrada uma função 3D pode ser aplicada ao *bilateral grid*. No caso específico do filtro bilateral, a função utilizada é uma convolução por um kernel Gaussiano, onde os desvios padrões são σ_s para o componente espacial e σ_r para o componente de intensidade.

O processo de *slicing* recupera o pixels da posição (x, y) no *bilateral grid* após o processamento. Dado um *bilateral grid* Γ e uma imagem de referência R (geralmente a mesma utilizada no processo de criação do *grid*), extrai-se o mapa de valores final M amostrando o *bilateral grid* na posição $(x/s_s, y/s_s, R(x, y)/s_r)$ utilizando interpolação trilinear, recuperando um vetor homogêneo, e dividindo este vetor pelo fator de normalização, recuperando o valor final da filtragem.

Utilizando a notação descrita neste artigo, é possível descrever o algoritmo do Fast Bilateral Filter (descrito na seção 2.2.1) através da equação:

$$bf(I) = s_I(G_{\sigma_s, \sigma_r} \otimes g(I)) \quad (2.5)$$

onde um grid g é inicializado com a imagem I , convoluído por uma Gaussiana 3D com parâmetro espacial σ_s e parâmetro de intensidade σ_r , e é finalmente retirado do grid g através da operação de *slicing* com a imagem I (s_I). No caso do filtro bilateral cruzado, simplesmente substitui-se s_I , na equação anterior, por s_E , onde E é uma imagem que contém as arestas que se deseja preservar, e I os dados que se deseja filtrar.

2.2.3 Real Time Bilateral Filter

O trabalho de Yang et al. [YANG; TAN; AHUJA, 2009] apresenta um algoritmo de filtragem bilateral com complexidade computacional $O(1)$. Isto é possível através da decomposição do filtro bilateral em diversos filtros espaciais constantes, que são utilizados para calcular a saída do filtro.

Um filtro bilateral contém um kernel espacial e um kernel para intensidades. Dado p como um pixel da imagem e q como um pixel na vizinhança $N(p)$ de p , $I(p)$ e $I(q)$ como os valores do pixel p e q , o valor filtrado de p é

$$I^B(p) = \frac{\sum_{q \in N(p)} (f_S(p, q) \cdot f_R(I(p), I(q)) \cdot I(q))}{\sum_{q \in N(p)} (f_S(p, q) \cdot f_R(I(p), I(q)))} \quad (2.6)$$

onde f_S e f_R são os kernels espaciais e de intensidade, respectivamente.

Na prática, a intensidade dos pixels de uma imagem I é discreta, com $I(p) \in \{0, \dots, N-1\}$, onde N é o número total de valores em tons de cinza. Dado $I(p) = k$, a Equação 2.6 pode ser expressa como

$$I^B(p) = \frac{\sum_{q \in N(p)} (f_S(p,q) \cdot f_R(k, I(q)) \cdot I(q))}{\sum_{q \in N(p)} (f_S(p,q) \cdot f_R(k, I(q)))}. \quad (2.7)$$

Para cada pixel q e cada valor de intensidade $k \in \{0, \dots, N-1\}$, define-se

$$\begin{aligned} W_k(q) &= f_R(k, I(q)) \text{ e} \\ J_k(q) &= W_k(q) \cdot I(q). \end{aligned} \quad (2.8)$$

A filtragem bilateral pode então ser decomposta em N conjuntos de respostas lineares

$$\begin{aligned} J_k^B(p) &= \frac{\sum_{q \in N(p)} f_S(p,q) J_k(q)}{\sum_{q \in N(p)} f_S(p,q) W_k(q)} \\ \text{tal que } I^B(p) &= J_{I(p)}^B(p) \end{aligned} \quad (2.9)$$

onde J_k^B é um componente principal da imagem filtrada bilateralmente (*Principle Bilateral Filtered Image Component, PBFIC*).

Na prática, pode-se assumir que apenas $\hat{N} \leq N$ PBFICs serão utilizadas (portanto, $k \in \{L_0, \dots, L_{\hat{N}-1}\}$, onde L é um subconjunto dos valores de intensidade presentes na imagem). Como a intensidade do pixel p é $I(p) \in [L_k, L_{k+1}]$, o valor filtrado $I^B(p)$ pode ser interpolado linearmente de $J_k^B(p)$ e $J_{k+1}^B(p)$ através da seguinte equação:

$$I^B(p) = (L_{k+1} - I(p)) J_k^B(p) + (I(p) - L_k) J_{k+1}^B. \quad (2.10)$$

2.2.4 Adaptive Manifolds

O trabalho de GASTAL; OLIVEIRA [2012] utiliza uma abordagem inovadora para acelerar o resultado do filtro bilateral, através da utilização de um conjunto de *manifolds* que se adaptam ao sinal que se deseja filtrar, chamados de *adaptive manifolds*. Esta técnica apresenta melhor performance e utiliza menos memória que abordagens anteriores.

Este filtro possui três estágios: *splatting*, *blurring*, e *slicing*. O primeiro estágio, *splatting*, faz uma projeção Gaussiana das cores f_i de cada um dos pixels p_i em todos os k *manifolds* gerados. Os valores projetados são armazenados nos pontos de amostragem $\hat{\eta}_{ki}$:

$$\Psi_{splat}(\hat{\eta}_{ki}) = \phi \frac{\Sigma_{\mathcal{R}}}{2} (\eta_{ki} - f_i) f_i.$$

$\Sigma_{\mathcal{R}}$ é uma matriz de covariância $d_{\mathcal{R}} \times d_{\mathcal{R}}$ que controla o decaimento do kernel Gaussiano ϕ na dimensão \mathcal{R} .

A etapa de *blurring* filtra cada *manifold* k separadamente, utilizando os valores Ψ_{splat} de todos os pontos de amostragem $\hat{\eta}_{ki}$. Isto resulta em um novo valor $\Psi_{blur}(\hat{\eta}_{ki})$ armazenado em cada ponto de amostragem. As distâncias para esta etapa são calculadas considerando a curvatura do *manifold* em uma versão escalada do espaço $\mathcal{S} \times \mathcal{R}$.

Finalmente, a operação de *slicing* calcula a resposta final g_i da filtragem para cada pixel p_i através da interpolação dos valores Ψ_{blur} agregados de todos os *manifolds*, utilizando os mesmos pontos de amostragem $\hat{\eta}_{ki}$ utilizados na operação de *splatting*. A interpolação é feita utilizando convoluções normalizadas, e o valor final g_i é calculado utilizando valores nas posições $\hat{\eta}_{ki}$:

$$g_i = \frac{\sum_{k=1}^K w_{ki} \Psi_{blur}(\hat{\eta}_{ki})}{\sum_{k=1}^K w_{ki} \Psi_{blur}^0(\hat{\eta}_{ki})}, \quad w_{ki} = \phi \frac{\Sigma_{\mathcal{R}}}{2} (\eta_{ki} - f_i).$$

O valor Ψ_{blur}^0 é a versão borrada de Ψ_{splat}^0 :

$$\Psi_{splat}^0(\hat{\eta}_{ki}) = \phi_{\frac{\Sigma_{\mathcal{R}}}{2}}(\eta_{ki} - f_i),$$

que, ao contrário de Ψ_{splat} , não apresenta o valor f_i multiplicando à direita da equação.

A geração dos *adaptive manifolds* é feita através do algoritmo descrito a seguir:

Passo 1: Gera-se o primeiro *manifold*, η_1 , filtrando o sinal com um filtro passa-baixas: $\eta_1(p_i) = (h_{\Sigma_{\mathcal{S}}} \otimes f)(p_i)$, onde \otimes simboliza uma convolução, e $h_{\Sigma_{\mathcal{S}}}$ é um filtro passa-baixas na dimensão \mathcal{S} com matriz de covariância $\Sigma_{\mathcal{S}}$.

Passo 2: Calcula-se a direção $v_1 \in \mathbb{R}^{d_{\mathcal{R}}}$, que sumariza as variações de cores de pixel f_i em torno do *manifold* η_1 . A direção v_1 corresponde ao autovetor associado ao maior autovalor da matriz de covariância $(\mathbf{f}_1 - \boldsymbol{\eta}_1)(\mathbf{f}_1 - \boldsymbol{\eta}_1)^T$, onde $\mathbf{f}_1 = (f_1 \dots f_N)$ é uma matriz de dimensão $d_{\mathcal{R}} \times N$ contendo todas as cores de pixels f_i , e $\boldsymbol{\eta}_1 = (\eta_{11} \dots \eta_{1N})$ é uma matriz de dimensões $d_{\mathcal{R}} \times N$ contendo todos os valores de manifold η_{1i} associados com o pixel p_i .

Passo 3: Segmenta-se os pixels em dois conjuntos \mathcal{C}_- e \mathcal{C}_+ utilizando o sinal do produto escalar $dot = v_1^T(f_i - \eta_{1i})$:

$$\begin{cases} p_i \in \mathcal{C}_- & \text{se } dot < 0, \\ p_i \in \mathcal{C}_+ & \text{senão.} \end{cases}$$

Isto pode ser entendido como a segmentação dos pixels do sinal de entrada em dois subconjuntos: um que está localmente acima do *manifold* e um que está localmente abaixo, definindo duas populações distintas.

Passo 4: Computa-se um novo manifold η_- através de uma filtragem passa-baixas ponderada do sinal de entrada, dando peso zero para pixels que **não** estão em \mathcal{C}_- :

$$\eta_-(p_i) = \sum_{p_j \in \mathcal{C}_-}^N W_-(p_j) f_j \Big/ \sum_{p_j \in \mathcal{C}_-}^N W_-(p_j),$$

$$W_- = \theta(\eta_{1j} - f_j) h_{\Sigma_{\mathcal{S}}}(p_i - p_j).$$

$h_{\Sigma_{\mathcal{S}}}$ é o filtro passa baixas utilizado para gerar η_1 , e θ é uma função que dá mais peso a pixels p_j que não foram bem representados pelo manifold η_1 :

$$\theta(\eta_{1j} - f_j) = 1 - w_{1j}.$$

A construção do *manifold* η_+ associado com \mathcal{C}_+ é feito de maneira similar.

Passo 5: Baseado em um critério de parada, decide-se se são necessários mais *manifolds*. Caso haja necessidade, repete-se recursivamente o Passo 2, substituindo ocorrências de η_1 com η_- , e somente utilizando pixels $p_i \in \mathcal{C}_-$ para construir as matrizes \mathbf{f}_- e $\boldsymbol{\eta}_-$, e conjuntos \mathcal{C}_{--} e \mathcal{C}_{-+} . Faz-se o mesmo para η_+ utilizando pixels $p_i \in \mathcal{C}_+$.

3 TRABALHOS RELACIONADOS

O problema da compressão de faixa dinâmica tem sido foco de diversos estudos, motivados pela incapacidade de monitores convencionais de exibir simultaneamente toda a faixa dinâmica que pode ser encontrada em uma cena real. Porém, com recentes pesquisas em hardware capaz de aumentar a faixa dinâmica de dispositivos de exibição [SEETZEN et al., 2004; DOLBY], um problema análogo ao tone mapping foi criado, mas com um foco diferente: como melhorar o contraste de imagens e vídeos LDR de modo a tirar o melhor proveito da capacidade disponibilizada por monitores com faixa dinâmica estendida, provendo assim ao usuário uma experiência superior? Este processo é chamado tone mapping reverso (*reverse* (ou *inverse*) *tone-mapping*).

Mapear linearmente todos os valores da imagem LDR para a máxima capacidade do monitor HDR é a opção mais simples para a resolução deste problema. Porém, imagens mais escuras tendem a ficar muito claras devido à escala nos pixels da imagem. Algoritmos para melhorar a visualização de imagens sem que percam sua naturalidade em monitores HDR começaram a ser desenvolvidos para evitar este problema. A eficiência de algoritmos de tone mapping reverso é mostrada em estudos recentes [YOSHIDA et al., 2006; SEETZEN et al., 2006; AKYÜZ et al., 2007; BANTERLE et al., 2009]. Nestes estudos, diferentes configurações de imagens são apresentadas a diversos observadores humanos, a fim de encontrar quais imagens têm melhor qualidade ou são mais parecidas com cenas reais. Estes estudos mostram que observadores preferem observar uma versão da imagem LDR com contraste escalado para monitor HDR a observar a imagem LDR original em monitor LDR.

Algoritmos de tone mapping reverso, de acordo com [BANTERLE et al., 2009], podem ser classificados pela maneira com que interagem com o conteúdo da imagem, como veremos nas próximas seções.

3.1 Modelos de Descontorno

DALY; FENG [2003, 2004] propuseram alguns métodos para aumentar a profundidade de bits de imagens e vídeos de 8 bits para monitores mais novos, capazes de representar 10 bits. Nestes novos monitores LCD, que apresentam um maior contraste (por volta de 1000 : 1) e têm um ponto de luminância máximo maior (tipicamente $400\text{cd}/\text{m}^2$), dados de 8 bits teriam seu conteúdo expandido linearmente, o que resultaria em artefatos de contorno (*contouring artifacts*, Figura 3.1). O objetivo destes métodos é criar uma imagem de média faixa dinâmica, que remove artefatos de contorno em áreas de transição, sem ênfase em áreas super ou subexpostas.

3.1.1 Amplitude Dithering for High Contrast Displays

Este algoritmo [DALY; FENG, 2003] é baseado em *amplitude dithering* [ROBERTS, 1962]. *Amplitude dithering* é uma técnica que adiciona um padrão de ruído à imagem antes de sua quantização. Desta maneira, a profundidade de bits percebida é maior que a real, devido a uma constante adaptação acontecendo no monitor e no sistema visual humano.

A técnica de ROBERTS [1962] foi modificada por Daly e Feng para ser aplicada em dispositivos de visualização capazes de exibir contrastes maiores. Os autores modularam o ruído utilizado combinando o efeito do ruído de padrão fixo do monitor e o ruído percebido pelo sistema visual humano, fazendo com que o ruído aplicado pareça invisível. Os autores utilizaram a função de sensibilidade a contraste (*Contrast Sensitivity Function, CSF*), uma função 2D anisotrópica derivada a partir de experimentos psicofísicos [DALY, 1993], que se estende na dimensão temporal durante imagens em movimento, fazendo com que o ruído aplicado à imagem tenha uma variância maior. O uso desta função permite que os valores da imagem final sejam expandidos em mais um bit.

3.1.2 Countouring Removal

Em outro algoritmo, DALY; FENG [2004] apresentam um método capaz de remover contornos em vez de mascará-los com ruído invisível. O primeiro passo é filtrar a imagem inicial no bit p utilizando um filtro passa-baixas. O filtro precisa ser grande o suficiente para atravessar os falsos contornos. Esta operação aumenta a profundidade de bits para $n > p$, pois durante a média uma maior precisão é necessária do que para os valores iniciais. Esta imagem é, em seguida, quantizada no bit p , e quaisquer contornos que aparecem são falsos, pois a imagem não contém altas frequências. Os falsos contornos são subtraídos da imagem original, e a imagem filtrada no bit p é adicionada para restaurar seus componentes de baixa frequência. A principal limitação deste método é a impossibilidade de remover contornos em áreas de alta frequência na imagem, mas estes são difíceis de detectar pelo sistema visual humano devido ao mascaramento de frequências (*frequency masking*) [FERWERDA et al., 1997].



Figura 3.1: Exemplos de artefatos de contorno [DALY; FENG, 2004].

3.2 Modelos Globais

Modelos globais são métodos que aplicam a mesma função de expansão a todos os pixels da imagem de entrada.

3.2.1 A Power Function Model for Range Expansion

Um dos primeiros métodos de expansão de faixa dinâmica foi proposto por LANDIS [2002]. Este método de expansão, usado primariamente para iluminação baseada em imagens, utiliza funções de potência. A expansão de luminância é definida como

$$L_w(x) = \begin{cases} (1 - k) \cdot L_d(x) + k \cdot L_{w,max} \cdot L_d(x) & \text{se } L_d(x) \geq R \\ L_d(x) & \text{caso contrário} \end{cases}$$

onde R é o limite de expansão, que é igual a 0.5 no trabalho original, $L_{w,max}$ é a luminância máxima que o usuário necessita para a imagem expandida, e α é o expoente de queda que controla a curva de expansão.

Embora esta técnica produza *light-probes* em alta faixa dinâmica de boa qualidade para iluminação baseada em imagens, ela não produz imagens capazes de serem visualizadas em monitores HDR. Isto acontece devido a este método não eliminar artefatos de compressão ou quantização, que são intensificados devido à potenciação.

3.2.2 Linear Scaling for HDR Monitors

Para investigar conteúdo LDR em monitores HDR, AKYÜZ et al. [2007] realizaram uma série de experimentos psicofísicos. Os experimentos executados foram avaliações de imagens transformadas por tone mapping, imagens de exposição simples e imagens HDR, utilizando o monitor Dolby DR-37P. O experimento envolveu 22 participantes com idades entre 20 e 40 anos, e em todos os experimentos 10 imagens HDR foram usadas, contendo luminâncias correspondentes a ambientes internos e externos, em condições de luz variando de escuras a muito claras.

O primeiro experimento foi uma comparação entre imagens HDR e LDR produzidas utilizando diversos operadores de tone mapping, uma imagem capturada automaticamente (o que minimiza o número de pixels super/subexpostos), e uma imagem escolhida por participantes de um estudo piloto. As imagens foram exibidas no monitor DR-37P, utilizando imagens HDR próprias para este monitor e imagens LDR calibradas para simular sua aparência em um monitor LCD Dell UltraSharp 2007FP, de 20.1". Os participantes precisaram classificar as imagens que lhes pareciam melhores. Para cada imagem no teste o participante precisava observar uma diferente versão por 2 segundos, que era aleatoriamente escolhida entre os diferentes tipos disponíveis. Os resultados experimentais mostraram que participantes preferiam imagens HDR. Além disso, os autores não encontraram uma diferença muito grande na preferência dos participantes entre imagens onde tone mapping foi aplicado e as demais imagens LDR (tanto automáticas quanto escolhidas pelo estudo piloto).

No segundo experimento os autores comparam imagens HDR com imagens LDR, tanto originais quanto expandidas através do operador descrito na Equação 3.1:

$$Y'(x) = k \left(\frac{Y(x) - Y_{min}}{Y_{max} - Y_{min}} \right)^\gamma \quad (3.1)$$

onde $Y(x)$ corresponde ao pixel x do canal Y do espaço de cores Yxy [CIEXYZ], Y_{min} e Y_{max} correspondem ao valor mínimo e máximo deste canal, respectivamente, k é a intensidade máxima de luminância do monitor HDR, e γ é o fator não-linear de escala. Para este experimento, imagens com diferentes valores γ iguais a 1, 2.2 e 0.45 foram geradas. A organização do experimento e as tarefas de classificação foram as mesmas do primeiro experimento. Os resultados mostraram que imagens expandidas a partir de exposições

mais iluminadas tiveram uma classificação melhor do que imagens HDR, enquanto as imagens HDR obtiveram uma melhor classificação quando a luminância média entre as imagens era similar. Os autores sugeriram, então, que a luminância média é mais importante que o contraste. Finalmente, outro resultado importante foi que a utilização de um fator de escala linear ($\gamma = 1$) foi a expansão mais favorecida, sugerindo que a escala linear pode ser suficiente para expandir imagens LDR.

Os autores trabalharam somente com imagens HDR de alta resolução, sem artefatos de compressão, e capturadas artisticamente. Enquanto a escala linear funciona bem sob estas condições ideais, em cenários mais realistas, como programas de televisão ou DVDs, onde se emprega a compressão dos dados, esta técnica pode não ser sempre a melhor escolha. Nestes casos, uma expansão mais cuidadosa deve ser empregada para evitar a amplificação de artefatos de compressão e contornos.

3.2.3 Expanding Over-exposed Content

[MASIA et al., 2009a] realizaram uma série de experimentos psicofísicos com o objetivo de mostrar como operadores de tone mapping reverso se comportam com imagens sob condições de exposição não ideais. Os algoritmos utilizados neste estudo são: LDR2HDR [REMPEL et al., 2007], descrito na seção 3.4.2, mapas de expansão [BANTERLE et al., 2006], descrito na seção 3.4.1, e escala de contraste linear [AKYÜZ et al., 2007], descrito na seção 3.2.2.

O primeiro experimento avalia os resultados de três operadores de tone mapping reverso, citados anteriormente, além da imagem em LDR. Fotografias de nove cenas com diferentes condições de iluminação foram utilizadas, e cada uma das cenas foi capturada com quatro tempos diferentes de exposição. Cinco cenas eram compostas de imagens claras (corretamente expostas a superexpostas), e as outras quatro consistiam de imagens mais escuras (corretamente expostas a subexpostas). Cada uma destas imagens foi expandida utilizando os algoritmos citados anteriormente e exibida em um monitor Dolby DR-37P simultaneamente. A posição de cada algoritmo foi randomizada para cada teste, e a ordem destes testes também foi randomizada, com a restrição de que testes consecutivos não mostravam a mesma cena. A tarefa apresentada foi avaliar a qualidade de cada uma das imagens exibidas em uma escala de 1 a 7, de acordo com o quanto as imagens seriam similares a uma cena real. Para auxiliar nesta tarefa, o grupo foi treinado com alguns testes antes do início do experimento.

De acordo com este estudo, os usuários preferem imagens consideradas escuras expandidas utilizando um operador de tone mapping reverso do que imagens consideradas claras, expandidas da mesma maneira. Além disso, os algoritmos LDR2HDR [REMPEL et al., 2007] e de escala linear [AKYÜZ et al., 2007] obtiveram o melhor desempenho, em média, neste teste, seguidos das imagens em LDR e o algoritmo de mapas de expansão [BANTERLE et al., 2006], nesta ordem. O fraco desempenho deste último algoritmo é devido à maneira com que ele exagera os erros de imagens não expostas corretamente, resultando em artefatos visíveis.

O segundo experimento foi motivado pela observação de que artefatos introduzidos pelos algoritmos LDR2HDR e de mapas de expansão são tipicamente visíveis quando se aplica um operador de tone mapping na imagem. Isto acontece por apresentarem um componente espacial, como arestas ou ruído amplificado, e não devido a níveis inapropriados de intensidades para certas características da imagem. Neste novo experimento, em que as imagens foram exibidas em um monitor LDR através da aplicação de um operador de tone mapping (neste caso, ajuste de histograma [WARD et al., 1997]), observou-se que

as notas dadas a cada algoritmo tinham forte correlação com as do primeiro experimento ($R^2 = 0.9018$). Este resultado não indica que as imagens pareçam iguais tanto em LDR quanto em HDR, mas demonstra que o padrão de preferências é bem conservado. A forte correlação também sugere que uma avaliação razoável da qualidade de algoritmos de tone mapping reverso pode ser feita sem a utilização de um monitor HDR, e que as avaliações subjetivas de imagens LDR expandidas dependem muito mais da presença ou não de artefatos de visualização do que de intensidades exatas das diferentes estruturas presentes nas imagens.

Com base nestas informações, um operador de tone mapping reverso foi proposto, buscando apresentar a imagem de maneira que detalhes se mostrem mais proeminentes, ao invés de simplesmente expandir áreas saturadas, como os algoritmos de tone mapping reverso geralmente fazem. A expansão sugerida para valores de luminância linearizados segue uma curva γ , que tem o efeito de escurecer a aparência geral da imagem (o que, de acordo com [MEYLAN; DALY; SÜSTRUNK, 2006], produz melhores resultados para imagens claras) ao mesmo tempo que amplifica o contraste (o que melhora a qualidade percebida da imagem, de acordo com [REMPEL et al., 2007]). A linearização dos valores de luminância antes da expansão é feita utilizando uma curva gamma ($\gamma = 2.2$), que pode ser utilizada no lugar da inversa da função de resposta da câmera sem produzir artefatos visíveis [REMPEL et al., 2007].

Para obter um valor γ adequado para cada imagem, primeiramente deve-se obter uma medida da intensidade da imagem. Isso é feito calculando-se um valor chave (*key*) da imagem, que indica se a cena é subjetivamente clara ou escura. Como a intensidade média da imagem pode ser aproximada pelo logaritmo da luminância média da imagem [TUMBLIN; RUSHMEIER, 1993; REINHARD et al., 2002], o valor chave k é calculado como em [AKYÜZ; REINHARD, 2006]:

$$k = \frac{\log L_{avg} - \log L_m}{\log L_M - \log L_m}$$

onde $\log L_{avg} = (\sum_{x,y} \log L(x,y) + \delta)/n$. L_m e L_M são as luminâncias mínima e máxima da imagem, respectivamente, n é o número de pixels, $L(x,y)$ é a luminância do pixel em (x,y) e δ é um pequeno valor para evitar $\log 0$. Em seguida, calcula-se o expoente $\gamma = 10.44 \cdot k - 6.282$, que é utilizado para expandir a imagem de entrada.

Para avaliar o resultado deste novo operador, os autores utilizaram a métrica de avaliação de imagens DRIM [AYDIN et al., 2008], que é capaz de detectar três tipos de distorção diferentes: perda de contraste visível, amplificação de contraste invisível e reversão de contraste (mais detalhes sobre esta técnica na seção 4.4). Os resultados apresentados demonstram que o algoritmo proposto produz resultados de maior qualidade em imagens superexpostas, revelando mais detalhes e minimizando a perda e reversão de contraste, quando comparado com os algoritmos utilizados nos experimentos anteriores. Para imagens subexpostas, no entanto, obteve resultados similares aos outros operadores testados.

É possível perceber, contudo, que este algoritmo falha em imagens onde $k < 0.6$, pois o expoente γ se torna negativo e distorce a imagem final. Além disso, a regressão apresentada para o cálculo de γ não corresponde com os valores apresentados em uma tabela do próprio artigo, levando os autores a publicarem posteriormente um relatório técnico [MASIA; GUTIERREZ, 2011] onde discutem as limitações de seu trabalho anterior e propõem novas e mais robustas maneiras de calcular o valor γ utilizado por sua técnica.

Neste relatório técnico, os autores utilizam diferentes estatísticas da imagem, além de um valor γ ajustado manualmente a fim de obter sua melhor representação subjetiva,

para construir diferentes regressões multilíneas. Estas regressões utilizam o valor γ ajustado manualmente como objetivo, e avaliam a correlação de cada uma das diferentes estatísticas calculadas para definir se devem ser incluídas ou não. Três equações, obtidas utilizando diferentes restrições, são apresentadas:

$$\gamma = 3.8872 + 0.3752 \log L_H - 2.9941k_1 + 0.016P_{se} \quad (3.2)$$

$$\gamma = 0.9855 + 2.8972L_H - 0.8232L_{med} + 0.2734skew_L - 0.0898kurt_L \quad (3.3)$$

$$\gamma = 2.4379 + 0.2319 \log L_H - 1.1228k_1 + 0.0085P_{se} \quad (3.4)$$

onde L corresponde à imagem de luminância da imagem de entrada, L_H corresponde à média geométrica dos valores em L , k_1 corresponde ao valor chave (*key*) de L , P_{se} é a percentagem de pixels superexpostos, L_{med} é a média aritmética dos valores de L , $skew_L$ é a medida de assimetria da distribuição dos valores de L (*skewness*), e a curtose $kurt_L$ descreve a forma da distribuição (*kurtosis*).

A média geométrica dos valores de luminância da imagem (L_H) é obtida através da equação:

$$L_H = \exp\left(\frac{1}{N} \sum_{i=1}^N \log(L(i) + \epsilon)\right) \quad (3.5)$$

onde N corresponde ao número de pixels em L e ϵ é um valor muito pequeno, para evitar singularidades quando $L(i) = 0$.

Obtém-se o valor chave k_1 da imagem ignorando-se 1% dos maiores e menores valores de L :

$$k_1 = \frac{\log L_H - \log L_{min}}{\log L_{max} - \log L_{min}} \quad (3.6)$$

onde L_{min} e L_{max} correspondem ao menor e maior valores de luminância, respectivamente, e L_H é obtido através da Equação 3.5.

A média aritmética da luminância L_{med} é dada por $1/N \sum_{i=1}^N L(i)$, com N sendo o número de pixels da imagem. A percentagem de pixels superexpostos p_{se} é calculada definindo pixels superexpostos como aqueles em que sua luminância $l \cdot 255 \geq 254$. Já o valor de assimetria $skew_L$ e a curtose $kurt_L$ são calculados utilizando o terceiro e quarto momentos centrais, respectivamente:

$$\begin{aligned} skew_L &= \mu_3/\sigma_3 \\ kurt_L &= \mu_4/\sigma_4 \\ \mu_k &= \int_{-\infty}^{+\infty} (x - \mu)^k f(x) dx \end{aligned} \quad (3.7)$$

onde $f(x)$ é a função da distribuição de frequência dos valores dos pixels da imagem, e σ_n corresponde ao desvio padrão da distribuição no momento n .

As equações obtidas foram, então, avaliadas através do método de mínimos quadrados, comparando o valor γ calculado pela equação com o valor γ ajustado manualmente. Destas, a Equação 3.4, descrita como mais robusta por seus autores (MASIA; GUTIERREZ), obteve melhor desempenho.

Esta técnica é considerada o estado da arte em operadores de tone mapping reverso. No Capítulo 6 demonstramos que, embora esta técnica produza resultados com boa qualidade para imagens superexpostas, resultados gerados a partir de imagens subexpostas podem acarretar até mesmo em perda de contraste, em relação à imagem original.

3.3 Modelos de Classificação

Estas técnicas buscam expandir os diferentes aspectos do conteúdo LDR através da identificação ou classificação de diferentes partes da imagem como fontes de luz ou pontos de brilho especular.

3.3.1 Highlight Generation for HDR Monitors

Meylan et al. apresentaram um operador de tone mapping inverso (*inverse Tone Mapping Operator, iTMO*) com a tarefa específica de representar pontos com brilho especular em imagens LDR quando exibidas em monitores HDR [MEYLAN; DALY; SÜSS-TRUNK, 2006]. A idéia principal é detectar as áreas difusas e especulares da imagem (Figura 3.2) e expandi-las utilizando diferentes funções lineares. A detecção é baseada na suposição de que as áreas com brilho especular são pequenas e brilhantes, o que significa que o valor de luminância difusa máxima ω pode ser obtida como o valor máximo da luminância da imagem. No entanto, é necessário um maior processamento da imagem para o caso em que regiões difusas brancas estão próximas a regiões com brilho especular.



Figura 3.2: Classificação de áreas difusas e especulares no trabalho de MEYLAN; DALY; SÜSSSTRUNK [2006]. **Esquerda.** Imagem de entrada. **Centro.** Componentes difusas da imagem. **Direita.** Componentes classificadas como especulares.

Após o cálculo de ω , a imagem pode ser expandida utilizando a seguinte função:

$$f(I(x)) = \begin{cases} s_1 I(x) & \text{se } L(x) \leq \omega \\ s_1 \omega + s_2 (I(x) - \omega) & \text{caso contrário} \end{cases}$$

$$s_1 = \frac{\rho}{\omega} \quad s_2 = \frac{1-\rho}{1-\omega}$$

onde $L(x)$ corresponde à luminância da imagem no pixel x e ρ é a porcentagem da luminância do monitor HDR alocada à parte difusa da imagem. No entanto, uma aplicação global de f pode produzir artefatos de quantização em torno das áreas com brilho especular expandidas. Este problema é reduzido através do uso de um filtro passa-baixas nestas regiões.

Finalmente, uma série de experimentos psicofísicos foram executados para determinar o valor de ρ em f para o monitor Dolby DR-37P. Os resultados mostraram que, para cenas externas, os usuários preferiam um valor alto de ρ , o que significa uma pequena porcentagem de faixa dinâmica alocada para brilho especular, enquanto para cenas internas o resultado era oposto. Para cenas internas e externas com a mesma luminância de áreas difusas os usuários escolheram um valor baixo de ρ , para que uma maior faixa dinâmica fosse alocada para o brilho especular. A conclusão dos autores ao analisar os dados foi de que $\rho = 0.66$ é uma boa estimativa para um caso geral.

Este algoritmo tem uma tarefa específica, que é a de reproduzir áreas com brilho especular em um monitor HDR. O uso deste método para outras tarefas, como a expansão de vídeos, requer um maior processamento dos dados disponíveis e um classificador, que foi descrito em um experimento de avaliação feito pelos autores.

3.3.2 Enhancement of Bright Video Features for HDR Display

DIDYK et al. [2008] propõe um sistema interativo para aumentar o brilho de vídeos LDR, focando e mostrando resultados para conteúdo disponível em DVD. A idéia principal do sistema é classificar uma cena em três componentes: difusos, reflexões e fontes de luz, e então expandir apenas reflexões e fontes de luz. Os autores explicam que componentes difusos são difíceis de expandir sem gerar artefatos de visualização e que é provavelmente a intenção dos diretores dos vídeos de mostrá-los saturados, ao contrário de fontes de luz e reflexões.

O sistema consiste de três partes principais: pré-processamento, classificação e expansão de regiões com perda de informação. O passo de pré-processamento gera os dados necessários durante a classificação. Em particular, ele determina as regiões em que houve perda de informação devido a saturação utilizando um algoritmo *flood-fill*. Ao menos um canal precisa estar saturado (ter um valor acima de 230, para conteúdo em DVD) e valores de luminância precisam ser maiores que 222. Além disso, neste estágio o fluxo ótico é calculado, além de outras informações como estatísticas da imagem, feições geométricas e características das vizinhanças dos pixels.

Durante a classificação determinam-se quais áreas correspondem a luzes, reflexões e regiões difusas em um *frame*. Esta etapa se baseia em um conjunto de treinamento de 2000 regiões classificadas manualmente. Primeiramente, uma *support vector machine* [VAPNIK, 1995] com kernel $k(z, z') = \exp(-\gamma \|z - z'\|^2)$ executa uma classificação inicial das regiões. Em seguida, um algoritmo de *motion tracking* melhora a estimativa inicial, utilizando um classificador de vizinho mais próximo (*nearest neighbour*) baseado na seguinte métrica Euclideana:

$$d^2((z, x, t), (z', x', t')) = 50 \|z - z'\|^2 + \|x - x'\|^2 + 5(t - t')^2$$

onde z são características da região, x são coordenadas na imagem, e t é o número do *frame*. Estes passos permitem alcançar um erro de classificação de 12.6% em todas as regiões utilizadas nos testes. O rastreamento de regiões onde informação foi perdida devido a saturação utilizando compensação de movimento reduziu a porcentagem de objetos que requerem correção manual para 3%. Finalmente, o usuário pode supervisionar as regiões classificadas, corrigindo classificações erradas utilizando uma interface de usuário intuitiva.

Regiões com perda de informação são expandidas através da aplicação de uma curva não-linear adaptativa, que é calculada baseada em derivadas parciais da região armazenadas em um histograma H . A curva é definida como uma equalização de histograma nos valores inversos de H :

$$f(b) = k \sum_{j=2}^b (1 - H |j|) + t_2$$

onde t_2 é o menor valor de luminância para a região com perda de informação. k é um fator de escala que limita o valor de expansão máximo m (igual a 150% para fontes de luz e 125% para reflexões):

$$k = \frac{m - t_2}{\sum_{j=1}^N (1 - H |j|)}$$

onde N é o número de *bins* em H . Para evitar o surgimento de artefatos de contorno durante a expansão, o canal de luminância é filtrado com um filtro bilateral, separando este em uma imagem de detalhes e uma camada base, que vão ser fundidas posteriormente.

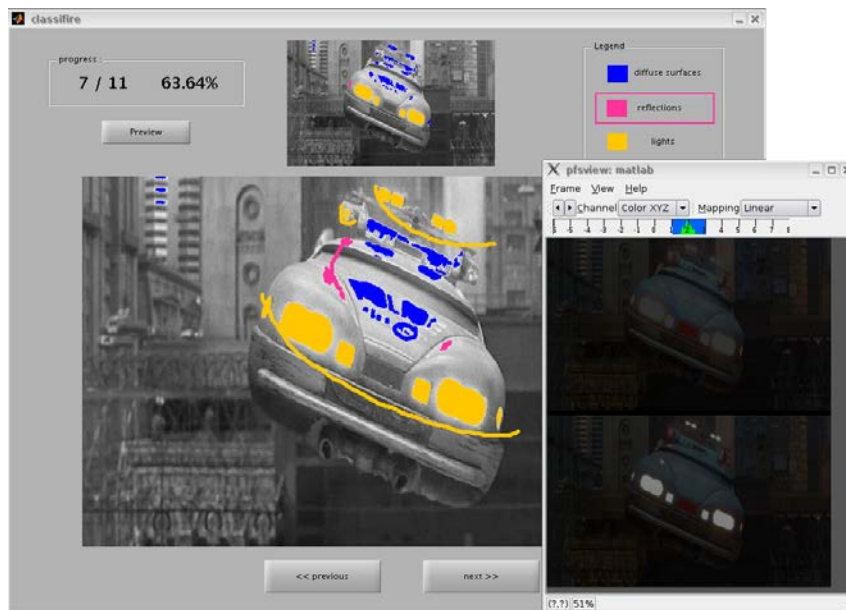


Figura 3.3: Interface do programa para classificação dos componentes de iluminação de uma imagem do trabalho de DIDYK et al. [2008].

Este método é considerado semi-automático, pois depende de intervenção do usuário para funcionamento pleno. A Figura 3.3 demonstra a interface do programa desenvolvido pelos autores.

3.4 Modelos de Mapas de Expansão

Estas técnicas utilizam algum tipo de guia para direcionar a expansão do conteúdo LDR, ao contrário de métodos globais. Estes guias podem ser chamados de mapas de expansão (*expand maps*) [BANTERLE et al., 2006, 2007, 2008] ou funções de aumento de brilho (*Brightness Enhancement Functions, BEFs*) [REMPEL et al., 2007; KOVALESKI; OLIVEIRA, 2009]

3.4.1 Non-Linear Expansion using Expand Maps

Uma abordagem geral para a expansão de conteúdo LDR para monitores HDR e iluminação baseada em imagens foi proposta em [BANTERLE et al., 2006, 2007]. Seus pontos chave são o uso de operadores de tone mapping reverso para expandir as intensidades presentes na imagem e a utilização de um campo de suavização para reconstruir as áreas perdidas por superexposição.

O primeiro passo é linearizar a imagem de entrada. Se a função de resposta da câmera é conhecida, sua inversa é aplicada ao sinal. Do contrário, métodos gerais podem ser aplicados, como [LIN et al., 2004; LIN; ZHANG; ZHANG, 2005], por exemplo. Em seguida, os valores de intensidade da imagem são expandidos, utilizando a inversa de um operador de tone mapping. Em sua implementação, a inversa do operador global de tone mapping de Reinhard et al. [REINHARD et al., 2002] foi utilizada. Esta escolha é devido ao operador possuir apenas dois parâmetros, o que permite que a expansão seja controlada

facilmente. Este operador de tone mapping reverso é definido como:

$$I'(x) = \frac{1}{2} I'_{max} I_{white} \left(I(x) - 1 + \sqrt{(1 - I(x))^2 + \frac{4}{I_{white}^2} I(x)} \right)$$

onde I'_{max} é a luminância máxima desejada da imagem expandida em cd/m^2 , e $I_{white} \in (1, +\infty)$ é um parâmetro que determina a forma da curva de expansão. Este valor é proporcional ao contraste, e os autores sugerem um valor de $I_{white} \approx I'_{max}$ para aumentar o contraste sem introduzir artefatos devido à expansão.

Em seguida, o mapa de expansão é calculado. O mapa de expansão é um campo representando uma versão de baixa frequência da imagem em áreas de alta luminância. Este mapa tem dois objetivos principais: o primeiro é reconstruir perfis de luminância perdidos em áreas superexpostas da imagem, enquanto o segundo é atenuar artefatos de quantização ou compressão que podem ter sido realçadas durante a expansão. O mapa de expansão foi implementado aplicando-se uma estimativa de densidade em amostras geradas utilizando amostragem de importância (*median-cut sampling* [DEBEVEC, 2005]). Finalmente, a imagem LDR expandida e a imagem original são combinadas usando interpolação linear, utilizando o mapa de expansão como peso da interpolação (Figura 3.4). Note que valores de luminância baixos são mantidos como na imagem original. Isto evita a compressão (para valores altos de L_{white}) de valores baixos, que poderia resultar em artefatos de contorno.

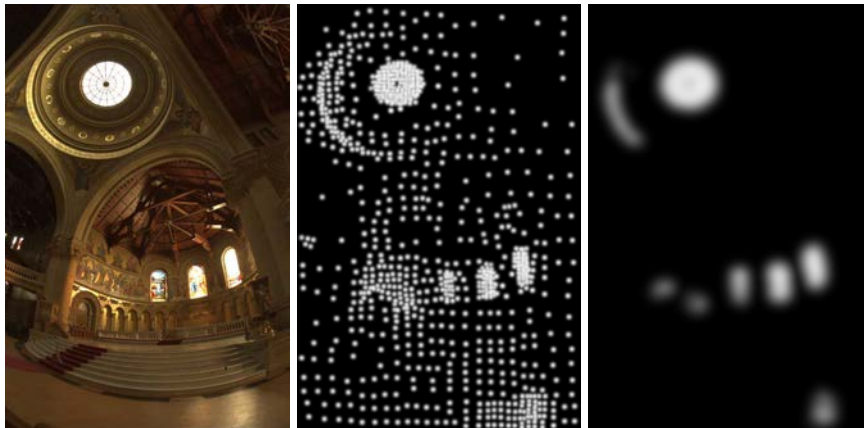


Figura 3.4: Cálculo do mapa de expansão de acordo com BANTERLE et al. [2006]. **Esquerda.** Imagem de entrada, em LDR. **Centro.** 1024 fontes de luz distribuídas na imagem, de acordo com o algoritmo de *median cut*. **Direita.** Mapa de expansão gerado pelo algoritmo.

Esta abordagem foi estendida para processar imagens e vídeos automaticamente em [BANTERLE et al., 2008]. Isso foi feito utilizando-se algoritmos de amostragem em 3D, estimação de densidade de volumes, transferência de arestas, e um número de heurísticas para determinar os parâmetros de cada componente do *framework*. Além disso, um mapa de expansão colorido foi empregado, permitindo a reconstrução de cores perdidas por saturação. O maior problema é sua velocidade, mas é possível alcançar um desempenho em tempo real em conteúdo de alta definição utilizando-se GPUs.

Este algoritmo apresenta uma solução geral para a visualização de conteúdo LDR em monitores HDR, e foi testado utilizando a métrica HDR-VDP [MANTIUK et al., 2005] para demonstrar sua eficiência em comparação com outros métodos. O maior problema

deste *framework* é que grandes áreas superexpostas podem não ser reconstruídas, produzindo áreas suaves em cor cinza nestes locais.

3.4.2 LDR2HDR

Uma técnica similar baseada em mapas de expansão (neste caso, chamadas de funções de aumento de brilho (*Brightness Enhancement Functions, BEFs*)) foi proposta em [REMPEL et al., 2007]. O objetivo desta técnica é a expansão em tempo real de vídeos em LDR.

O primeiro passo do algoritmo LDR2HDR é remover artefatos de compressão. Para isto utiliza-se um filtro bilateral simples. Técnicas mais avançadas para a remoção deste tipo de artefato são evitadas devido ao desempenho buscado pelos autores.

O próximo passo consiste em linearizar o sinal, utilizando uma função de gamma inversa. Uma vez que o sinal é linearizado, o contraste é aumentado de maneira otimizada para o monitor Dolby DR-37P. O contraste máximo foi limitado a $4000 : 1$ para evitar artefatos de visualização. Isto significa que o valor mínimo foi mapeado para 0.3 cd/m^2 , enquanto o máximo foi mapeado para 1200 cd/m^2 . Para aumentar o brilho em áreas claras, uma BEF é empregada. Esta função é calculada aplicando um limiar t aos dados de entrada que especifica o valor mínimo da luminância de um píxel para ser considerado "saturado" ($t = 254$ para imagens, $t = 230$ para vídeo). Em seguida, a função é filtrada utilizando um kernel Gaussiano com $\sigma = 150$. Para aumentar o contraste próximo a arestas da imagem, uma função de detecção de arestas é utilizada. Partindo dos píxels saturados (acima do limiar t), um algoritmo de *flood-fill* é aplicado até alcançar uma aresta, que é estimada utilizando gradientes. A máscara gerada por esta algoritmo é multiplicada pela função obtida anteriormente, limitando sua propagação (Figura 3.5).

Finalmente, a BEF é mapeada no intervalo $[1, \alpha]$, onde $\alpha = 4$ foi utilizado para o monitor Dolby DR-37P. Para aumentar a eficiência na geração da BEF, é possível utilizar-se de pirâmides Laplacianas [BURT, 1983], que são implementadas em GPU ou FPGA.

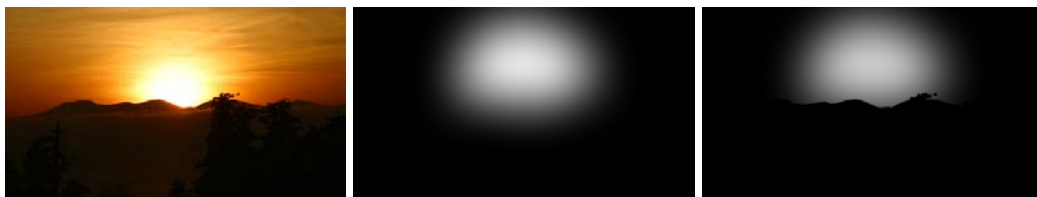


Figura 3.5: Exemplo de geração da função de aumento de brilho (BEF) por REMPEL et al. [2007]. **Esquerda.** Imagem de entrada, em LDR. **Centro.** Pixels saturados são borrados utilizando um kernel Gaussiano com $\sigma = 150$. **Direita.** BEF gerada após sua multiplicação pelo resultado do algoritmo de *flood-fill*.

Este algoritmo foi avaliado utilizando a métrica HDR-VDP [MANTIUK et al., 2005], comparando a imagem linearizada inicial com a imagem HDR gerada pelo algoritmo. Esta avaliação foi necessária para mostrar que o método proposto não introduz artefatos espaciais durante a expansão do conteúdo. Este algoritmo processa cada *frame* separadamente, o que pode resultar em conteúdo não temporalmente coerente, devido à natureza da geração da BEF.

3.4.3 High-quality Brightness Enhancement Functions

O trabalho descrito em [KOVALESKI; OLIVEIRA, 2009] busca aperfeiçoar a função de aumento de brilho (BEF) utilizada no algoritmo LDR2HDR [REMPEL et al., 2007]. No lugar de um algoritmo de *flood-fill* e operações morfológicas para gerar a BEF, este trabalho utiliza um *bilateral grid*, que aplica operações sensíveis a arestas inerentemente, para aumentar a qualidade da expansão (Figura 3.6).



Figura 3.6: Exemplos de BEFs geradas pelo algoritmo de KOVALESKI; OLIVEIRA [2009]. **Esquerda:** Imagens LDR de entrada. **Direita:** BEFs obtidas após processamento. Note os detalhes sutis capturados pela técnica, como as diferentes camadas de nuvens (primeira linha) e as varas de pescar (terceira linha), através da utilização de um *bilateral grid* para detectar as arestas da imagem.

Inicialmente a imagem de entrada I precisa ter suas intensidades linearizadas. Isso é feito da mesma maneira que em [REMPEL et al., 2007], onde uma curva gamma com $\gamma = 2.2$ é utilizada ($I_L = I^\gamma$). A partir de I_L geram-se duas novas imagens: L_1 e L_2 . L_1 é obtida pela conversão de I_L para o espaço de cores CIELAB [CIELAB] e mantendo-

se somente o canal L^* . L_2 também possui apenas um canal L^* mas, antes de converter a imagem I_L para o espaço de cores CIELAB, os valores dos canais RGB são alterados para que cada pixel contenha o maior valor entre R, G e B (ou seja, para cada canal RGB de um pixel, $p(R, G, B) = (C, C, C)$, onde $C = \max(R, G, B)$). A imagem L_2 é utilizada para criar um *bilateral grid* Γ , porém somente pixels de L_2 cujos valores correspondentes em I_L têm valores maiores que t , onde t depende do tipo de imagem ($t = 254$ para imagens, $t = 230$ para vídeos), são utilizados para preencher as células de Γ . Os demais pixels são inseridos em suas células com valor 0. Após a criação do *bilateral grid*, seu conteúdo é filtrado utilizando uma função Gaussiana, com $\sigma_s = 100$ e $\sigma_r = 30$. Em seguida, L_1 é utilizada para a operação de *slicing*. A BEF gerada por este algoritmo é obtida após converter a imagem resultante da operação de *slicing* para o espaço de cores RGB.

O algoritmo de expansão utiliza a BEF da mesma maneira que [REMPEL et al., 2007]: após linearizar os valores de intensidade da imagem de entrada (utilizando $\gamma = 2.2$), seu valor máximo é mapeado para 1200 cd/m^2 , enquanto seu ponto mínimo é mapeado para 0.3 cd/m^2 . A BEF, por sua vez, é mapeada para o intervalo $[1, \alpha]$, com $\alpha = 4$, e multiplicada pela imagem de entrada expandida, gerando a imagem final.

A Figura 3.6 mostra diversas BEFs geradas por esta técnica, que apresentam uma qualidade muito maior do que outras técnicas baseadas em mapas de expansão (como [REMPEL et al., 2007] e [BANTERLE et al., 2006]). No entanto, embora apresente bons resultados para imagens propriamente expostas e até subexpostas, esta técnica não obtém o mesmo desempenho para imagens superexpostas, provocando perda de detalhes nas BEFs geradas, o que compromete o resultado final e a experiência do usuário, como veremos no Capítulo 6.

3.5 Modelos Dependentes do Usuário

Como não é sempre possível recuperar o conteúdo HDR perdido utilizando técnicas automáticas, um método diferente, dependente do usuário foi proposto em [WANG et al., 2007], onde detalhes HDR podem ser adicionados a áreas que necessitam ser expandidas.

Os autores demonstraram os benefícios de um sistema que transfere detalhes para áreas super e subexpostas da imagem, combinando com um aumento de luminância. O processo foi chamado de *alucinação*, e seu sistema apresenta uma mistura entre métodos automáticos e dependentes de usuário.

O primeiro passo da alucinação é linearizar o sinal. Isto é obtido utilizando-se uma função de gamma inversa com $\gamma = 2.2$, que é o valor padrão para formatos de televisão e DVDs [ITU-BT.709]. Após este passo, a imagem é decomposta em um componente de iluminação e um componente de detalhes de textura. Isto se faz através da aplicação de um filtro bilateral à imagem I , obtendo uma versão filtrada I_f . Os detalhes da textura são obtidos com $I_d = I/I_f$. A luminância para o componente de iluminação é estimada utilizando uma interpolação linear de kernels Gaussianos elípticos. Inicialmente, um mapa de pesos w é calculado para cada pixel p :

$$w(p) = \begin{cases} \frac{C_{ue} - Y(p)}{C_{ue}} & \text{se } Y(p) \in [0, C_{ue}) \\ 0 & \text{se } Y(p) \in [C_{ue}, C_{oe}) \\ \frac{Y(p) - C_{oe}}{1 - C_{oe}} & \text{se } Y(p) \in [C_{oe}, 1] \end{cases}$$

onde $Y(p) = R(p) + 2G(p) + B(p)$, e C_{ue} e C_{oe} são os limites para pixels subexpostos e superexpostos, respectivamente. Os autores sugerem valores de 0.05 e 0.85 para C_{ue} e C_{oe} respectivamente. Em seguida, cada região superexposta é segmentada e adaptada com um

lóculo Gaussiano elíptico G , onde a variância do eixo é estimada utilizando a extensão da região, e o perfil é calculado utilizando um procedimento de otimização baseado em pixels não-superexpostos na aresta da região. A luminância é, então, combinada utilizando uma interpolação linear:

$$O(p) = w(p)G(p) + (1 - w(p)) \log_{10} Y(p)$$

Os detalhes de textura I_d são reconstruídos utilizando uma técnica de síntese de textura similar a [BERTALMÍO et al., 2003], onde o usuário pode selecionar uma área como a origem da textura. Esta síntese automática tem alguns limites quando algum tipo de entendimento da cena é necessário, e por este motivo uma ferramenta de *warping* é necessária. Isto permite que o usuário selecione uma região de origem e destino de uma textura, e pixels são então transferidos de um ponto a outro. Esta ferramenta é similar às ferramentas *stamp* e *healing* no programa Adobe Photoshop.

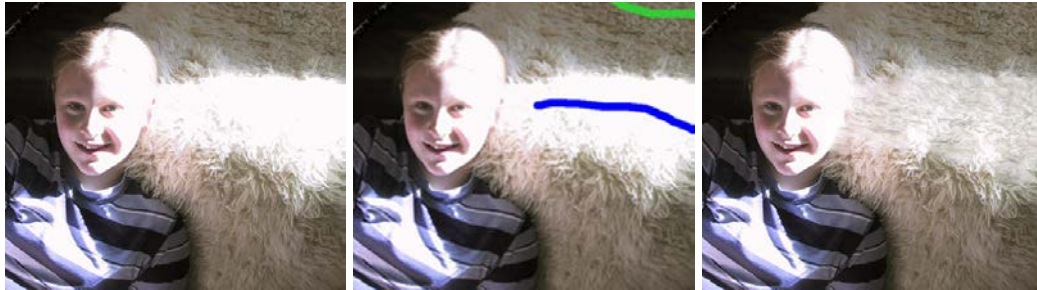


Figura 3.7: Exemplo de alucinação por WANG et al. [2007]. Os detalhes são transferidos através de intervenção do usuário, que define áreas de transferência na imagem. Neste caso, detalhes de textura provenientes da região indicada em verde foram transferidos para a região saturada indicada em azul, na imagem ao centro.

Finalmente, a imagem HDR é reconstruída combinando os componentes de detalhe e de iluminação através de edição de imagens baseadas em equações de Poisson [PÉREZ et al., 2003] para evitar divisas aparentes na transição de áreas superexpostas (Figura 3.7).

O problema principal desta técnica é que é dependente de um usuário e não automática, o que limita seu uso apenas para imagens.

4 AVALIANDO OPERADORES DE TONE MAPPING REVERSO

Em geral, operadores de tone mapping reverso são avaliados de acordo com testes psicofísicos que envolvem uma série de experimentos com diferentes usuários, geralmente com testes de escolha de imagens, em um monitor HDR. No entanto, estes testes refletem a preferência de cada pessoa, estando assim sujeitos à falta de treinamento e subjetividade.

Para avaliar algoritmos de tone mapping reverso objetivamente, podem-se utilizar métricas que avaliam a qualidade dos resultados em relação à imagem original. Estas técnicas buscam simular o sistema visual humano para compreender o que se passa durante a visualização de uma cena, retratando o estímulo em um mapa de diferenças visíveis entre as imagens que se deseja avaliar. Algumas das técnicas mais influentes para a avaliação de imagens de alta faixa dinâmica serão apresentadas a seguir.

4.1 Visual Difference Predictor

O *Visual Difference Predictor* (VDP) é um algoritmo que calcula a probabilidade de uma pessoa perceber diferenças entre duas imagens [DALY, 1993], onde uma delas é a imagem de referência e a outra uma imagem de teste, em que foram feitas alterações.

Para calcular estas diferenças, um modelo do sistema visual humano é utilizado. O modelo utilizado considera três principais variações de sensibilidade, sendo elas funções do nível de iluminação, frequência espacial e conteúdo do sinal.

As variações em função do nível de iluminação são primariamente devido às propriedades de adaptação à luz da retina, que é uma função não-linear da luminância. O passo de não-linearidade da amplitude (*amplitude nonlinearity*) do VDP descreve as variações como função da luminância de uma imagem, assim como as diferenças introduzidas com alterações nos níveis de iluminação. Este passo utiliza um modelo da rede neural da retina para sua computação. O modelo é modificado para que o nível de adaptação do olho para um pixel seja determinado apenas pelo valor daquele pixel. Embora dados fisiológicos indiquem que o sistema visual não consegue se adaptar a regiões pequenas de uma imagem, assume-se que o observador pode visualizar a imagem de qualquer distância, até mesmo a menor distância possível. Esta observação possibilita que este passo não necessite de nenhum parâmetro relacionado à frequência espacial, permitindo que estes parâmetros sejam modelados somente pelo próximo componente do VDP.

As variações em função da frequência espacial são devido à ótica do olho combinada com os circuitos neurais, e esses efeitos são combinados na função de sensibilidade ao contraste (*contrast sensitivity function*, CSF). Nossos componentes biológicos são altamente adaptáveis, o que faz com que a CSF mude em função da adaptação da luz, ruído,

cor, acomodação visual, excentricidade da cena e o tamanho da imagem. Todos os efeitos globais devido a mudanças de frequência espacial são modelados por esse componente, tanto aqueles que ocorrem devido à ótica e circuitos neurais como aqueles que ocorrem em outros locais do sistema visual. No VDP, a CSF também é afetada pelos parâmetros de distância de visualização, espaçamento entre pixels horizontal e vertical, além de todos os outros citados anteriormente (adaptação da luz, cor e outros). Embora não seja realisticamente possível considerar uma distância de visualização fixa, é válido afirmar que a distância visual ficará dentro de dois limites. Assim, a CSF é modelada em função de uma faixa de valores de distância, ao invés de um único valor.

Finalmente, variações como função do conteúdo do sinal são devido aos circuitos neurais pós-receptivos, e esses efeitos são chamados de mascaramento (*masking*). Estas variações ocorrem quando existe um fundo com intensidade suficiente para alterar a percepção em uma área da imagem, o que faz com que diferenças possam ser escondidas. Os parâmetros para definir os limites de diferença de intensidade são derivados da CSF. Um exemplo do resultado deste algoritmo encontra-se na Figura 4.1.



Figura 4.1: Detecção de diferenças pelo Visual Differences Predictor [DALY, 1993]. **Esquerda:** Imagem de referência. **Centro:** Imagem de teste, onde diferenças foram introduzidas. **Direita:** Diferenças detectadas pelo algoritmo.

4.2 HDR Visual Difference Predictor

MANTIUK et al. [2004] estendem o algoritmo do *Visual Differences Predictor* para trabalhar com imagens de alta faixa dinâmica, permitindo que seja possível comparar diferenças entre imagens HDR.

O algoritmo original do VDP utiliza o modelo de um fotorreceptor para simular a não-linearidade da resposta à luminância do olho humano. Porém, o uso deste modelo retorna valores arbitrários de resposta, que não estão relacionados fortemente com os valores de estudos de sensibilidade ao contraste. Assim, a CSF, que no trabalho de DALY [1993] mapeia os valores normalizados de contraste para unidades JND (*just noticeable difference*), utiliza um valor possivelmente incorreto para estimar estas unidades. Além disso, o modelo de fotorreceptor utiliza uma função sigmóide para a curva de resposta à luminância, o que assume perda de sensibilidade igual para níveis de luminância altos e baixos, embora o olho humano perca sensibilidade somente para níveis baixos de luminância. Assim, mesmo que esta simplificação seja aceitável para uma imagem LDR, isso pode levar a resultados errados devido aos valores encontrados em imagens HDR.

Este problema é abordado com o HDR-VDP, utilizando uma conversão dos valores de luminância para um espaço não-linear que utiliza unidades JND, que tem a propriedade

que adicionando-se ou subtraindo-se uma unidade neste espaço resulta em uma mudança mínima de contraste perceptível. A função que modela esta conversão simula a perda de sensibilidade para baixos níveis de luminância, e tem resposta logarítmica para níveis maiores.

No algoritmo de DALY [1993], a CSF é responsável não somente pela modelagem da perda de sensibilidade devido à frequência espacial e adaptação à luminância, mas também pela normalização do contraste em unidades JND. Para o HDR-VDP, esta normalização já foi executada no passo anterior. Assim, somente a modelagem da perda de sensibilidade é necessária.

No caso de imagens HDR, uma única CSF não pode ser utilizada, devido ao formato da função mudar significativamente de acordo com a adaptação à luminância presente em tais imagens. Assim, para cada pixel, um kernel de convolução diferente precisa ser utilizado. Porém, devido ao custo computacional desta solução, uma alternativa é utilizada: diversas filtragens da imagem após a transformada de Fourier são interpoladas linearmente de acordo com a luminância presente em cada pixel na imagem original.

Como o mascaramento de áreas não depende da resposta à luminância, esta etapa do VDP de DALY [1993] foi matida, exceto por uma alteração na normalização de alguns valores devido à maneira com que o HDR-VDP interpreta a intensidade em cada canal da imagem.



Figura 4.2: Detecção de diferenças pelo HDR-VDP [MANTIUK et al., 2004]. **Esquerda:** Imagem de referência. **Centro:** Imagem de teste, onde diferenças foram introduzidas. **Direita:** Resultado da detecção pelo algoritmo. Áreas em verde correspondem a uma probabilidade de detecção maior que 75%. Áreas em magenta correspondem a uma probabilidade de detecção maior que 95%.

Na saída do algoritmo, uma imagem com cores falsas é gerada. Áreas que possuem uma probabilidade de detecção maior que 75% são apresentadas em verde, e áreas com probabilidade de detecção maior que 95% são apresentadas em magenta (Figura 4.2).

4.3 HDR Visual Difference Predictor 2

O trabalho de MANTIUK et al. [2011] apresenta uma métrica de detecção de diferenças baseada em um novo modelo do sistema visual humano. Este modelo, derivado de novas medidas de sensibilidade ao contraste, busca simular uma ampla gama de condições de iluminação a que um usuário possa ser exposto, desde visão escotópica (noturna) a visão fotópica (diurna). Embora compartilhe suas origens com o HDR-VDP [MANTIUK et al., 2004], esta nova métrica e seus componentes constituem uma reformulação completa do algoritmo, em vez de uma mudança incremental.

Inicialmente, o algoritmo simula as distorções que podem ocorrer durante a passagem da luz pelo globo ocular. Ao observar uma cena, uma pequena porção de luz se dispersa

na córnea, lentes, interior do olho e retina [RITSCHER et al., 2009]. Esta dispersão causa uma poluição luminosa que reduz o contraste da imagem projetada na retina, e é especialmente pronunciada quando se observa cenas de alto contraste (HDR) contendo fontes de luz intensas. Este efeito se chama *disability glare* [VOS; BERG, 1999] e tem sido cuidadosamente medido através de estudos psicofísicos e métodos diretos.

Em seguida, calcula-se a probabilidade de um fotorreceptor captar um fóton de um dado comprimento de onda. Estes fotorreceptores também exibem uma resposta altamente não-linear à luz: a habilidade de ver uma grande faixa de valores de luminância deve-se principalmente à sua capacidade de regular a sensibilidade de acordo com a intensidade da luz que chega em nossos olhos. Este processo regulatório é chamado *luminance masking*, e é modelado através de uma CSF que varia com a frequência espacial e luminância da imagem. Muitos modelos visuais assumem um estado global de adaptação para uma imagem, o que não corresponde à maneira com que os fotorreceptores se comportam, especialmente no caso de cenas com grandes variações de intensidade de luz, como imagens HDR. O modelo proposto no HDR-VDP2 considera a natureza local do processo de adaptação.

Para calcular a resposta final da luz detectada por cones e bastonetes, simplesmente soma-se a resposta calculada para cada um destes fotorreceptores nos passos anteriores. Após, simulam-se mecanismos que são seletivos a faixas estreitas de frequências e orientações espaciais. Para imitar esta decomposição, que presume-se acontecer no córtex visual, modelos visuais comumente utilizam decomposições de imagens multiescala, como *wavelets* ou pirâmides. No caso do HDR-VDP2, utiliza-se *steerable pyramids* [SIMONCELLI; FREEMAN, 1995], que oferecem uma boa separação entre frequência e orientação espacial, além de evitar alguns artefatos que afetam a *Cortex Transform*, utilizadas pelo VDP [DALY, 1993] e HDR-VDP [MANTIUK et al., 2004].

Diferenças na detecção de contraste também podem ocorrer devido a diferentes fontes de ruído [DALY, 1990]. No HDR-VDP2, este fenômeno é modelado através da soma do ruído independente do sinal (devido à atividade neuronal) e ruído dependente do sinal (que causa uma menor visibilidade de detalhes em um fundo não uniforme). O ruído independente do sinal pode ser extraído da CSF, mas é necessário descontar seu componente ótico, que já foi utilizado em um passo anterior (mais especificamente, durante a simulação de *luminance masking*). Por sua vez, o ruído dependente do sinal, chamado de *visual* ou *contrast masking*, acontece devido à superposição de padrões com frequência e orientação espaciais similares, o que faz com que diferenças sejam mais difíceis de ser detectadas [FOLEY, 1994]. Três tipos de *contrast masking* são modelados: altas frequências na própria imagem (que podem esconder diferenças), orientações similares e/ou bandas de frequência adjacentes entre a imagem e a diferença introduzida.

Finalmente, os dados calculados nos passos anteriores são utilizados para calcular um mapa de diferenças entre as duas imagens de entrada, onde cada pixel do mapa corresponde à probabilidade de detecção de uma diferença na sua posição. Assim como os trabalhos anteriores, o mapa pode ser gerado com cores falsas para uma melhor visualização da saída do algoritmo (Figura 4.3).

4.4 Dynamic Range Independent Image Quality Metric

A *Dynamic Range Independent Image Quality Metric* [AYDIN et al., 2008], também conhecida como DRIM (*Dynamic Range Independent Metric*), é uma técnica capaz de detectar diferenças entre imagens com diferentes faixas dinâmicas. No entanto, ao invés



Figura 4.3: Detecção de diferenças pelo HDR-VDP2 [MANTIUK et al., 2011]. **Esquerda:** Imagem de referência. **Centro:** Imagem de teste, onde diferenças foram introduzidas. **Direita:** Resultado do algoritmo. Cores mais vibrantes correspondem a uma maior probabilidade de detecção.

de simplesmente relatar a probabilidade de detecção de uma diferença, estes dados são analisados para reportar apenas mudanças estruturais visíveis. É esta formulação que torna esta métrica independente de faixa dinâmica ou a pequenas mudanças na curva de tons.

Três diferentes classes de mudanças estruturais são detectadas por esta técnica (Figura 4.4):

Perda de contraste visível acontece quando contraste que era visível na imagem de referência não é mais detectado na imagem de teste. Esta mudança geralmente ocorre quando um operador de tone mapping comprime detalhes ao ponto que se tornam invisíveis.

Amplificação de contraste invisível acontece quando o contraste que era invisível na imagem de referência se torna visível na imagem de teste. Isto pode ser observado quando artefatos de contorno começam a aparecer devido a um aumento excessivo de contraste, ou detalhes que se tornam visíveis em áreas com pouco contraste após a aplicação de um operador de tone mapping reverso.

Reversão de contraste visível ocorre quando um contraste que é visível tanto na imagem de referência quanto na imagem de teste tem sua polaridade invertida. Isto pode ocorrer em áreas da imagem com muitas distorções, como artefatos de compressão salientes ou áreas com recortes (*clipping*).

De acordo com REMPEL et al. [2007], a amplificação de contraste invisível é uma característica desejável para operadores de tone mapping reverso, enquanto a perda de contraste é algo que deve ser evitado.

Antes de detectar estas mudanças estruturais, é necessário verificar se há uma mudança de contraste visível ou não. Para este passo, é necessário gerar um mapa de respostas normalizado perceptualmente, em que amplitudes iguais a 1 correspondem à probabilidade de detecção $P_{det} = 75\%$ (1 JND). Inicialmente, utilizava-se o HDR-VDP [MANTIUK et al., 2004] para a geração do mapa. No entanto, após a publicação do HDR-VDP2 [MANTIUK et al., 2011], este passou a ser utilizado.

No segundo passo, o mapa gerado anteriormente é separado em diferentes bandas com diferentes orientações e larguras espaciais, utilizando a *Cortex Transform* modificada por DALY [1993]. Então, para cada uma das diferentes distorções detectadas por esta técnica, calcula-se um mapa de probabilidades. Estes mapas são, finalmente, plotados em uma

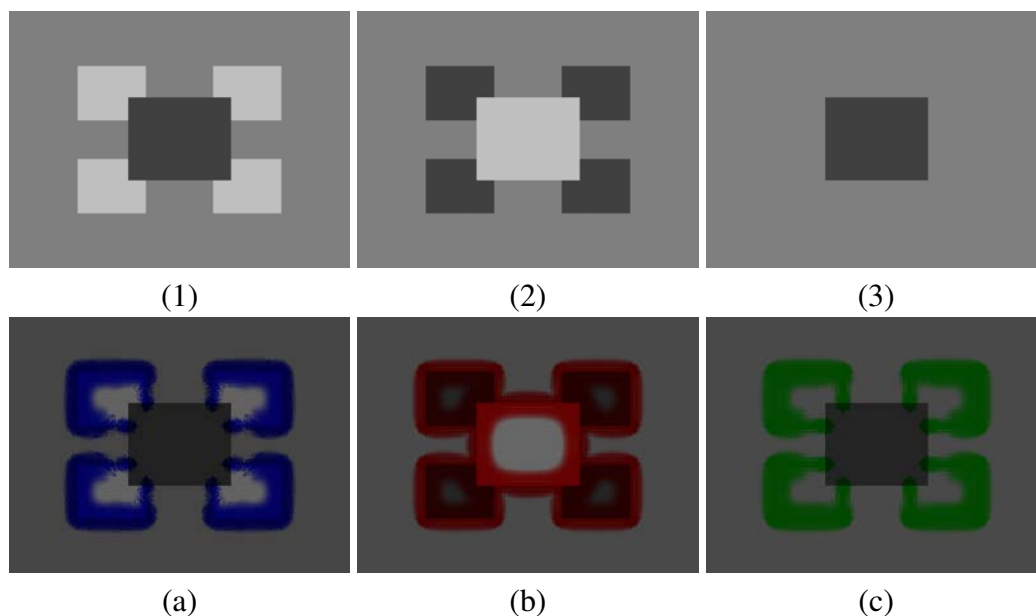


Figura 4.4: Exemplos de mudanças estruturais detectadas pelo DRIM [AYDIN et al., 2008]. **(a) Amplificação de contraste invisível.** Mudanças detectadas utilizando a imagem (3) como referência e (1) como teste. **(b) Reversão de contraste visível.** Mudanças detectadas utilizando a imagem (1) como referência e (2) como teste. **(c) Perda de contraste visível.** Mudanças detectadas utilizando a imagem (1) como referência e (3) como teste.

imagem de luminância da imagem de teste, para facilitar a visualização. Foram arbitrariamente escolhidas as cores verde para perda de contraste visível, azul para amplificação de contraste invisível, e vermelho para reversão de contraste visível (Figura 4.4).

É importante citar que, devido à sua formulação independente de faixa dinâmica, o DRIM não modela mascaramento por contraste (*contrast masking*), pois diferenças de contraste entre imagens não são relevantes ao resultado da métrica.

5 TONE MAPPING REVERSO PARA UMA AMPLA GAMA DE EXPOSIÇÕES

No trabalho de KOVALESKI; OLIVEIRA [2009], a função de aumento de brilho é gerada através da criação de um *bilateral grid* [CHEN; PARIS; DURAND, 2007] utilizando duas imagens diferentes, enquanto a operação de *slicing* é aplicada utilizando uma terceira imagem. Esta configuração provoca perda de detalhes nas BEFs, especialmente em imagens superexpostas, como mostra a Figura 5.1. Esta perda de detalhes reduz o contraste da imagem HDR obtida, como demonstraremos no Capítulo 6. Além disso, a técnica descrita em KOVALESKI; OLIVEIRA [2009] está restrita ao *bilateral grid*, o que limita sua longevidade frente a futuras técnicas de aceleração do filtro bilateral.

Neste capítulo mostramos como é possível melhorar o desempenho deste algoritmo para imagens superexpostas, bem como generalizá-lo para diferentes técnicas de aceleração do filtro bilateral. Em seguida, mostramos que é possível simplificar a técnica apresentada, devido ao seu comportamento similar a um filtro bilateral cruzado, e em consequência melhorar seu desempenho para condições de iluminação não ideais.

5.1 Generalização

O algoritmo utilizado em [KOVALESKI; OLIVEIRA, 2009] pode ser expresso, utilizando a mesma notação utilizada na Equação 2.5 do *bilateral grid*, como:

$$bf(I) = s_L(G_{\sigma_s, \sigma_r} \otimes g_M(T)) \quad (5.1)$$

onde cada pixel da imagem M contém o valor máximo entre os componentes RGB de I , L contém a imagem de luminância de I , e T contém todos os pixels de M que estão acima de um limiar t , que depende do conteúdo processado (254 para imagens, 230 para vídeos). $g_M(T)$ é um *bilateral grid* criado utilizando informações de arestas da imagem M , onde se armazenam valores da imagem T , G_{σ_s, σ_r} é uma Gaussiana 3D com parâmetro espacial σ_s e parâmetro de intensidade σ_r , e s_L corresponde à operação de *slicing* utilizando a imagem L . Finalmente, a imagem I tem seus valores linearizados utilizando uma curva gamma de 2.2 antes de gerar as imagens L , M e T .

Como o *bilateral grid* é uma generalização do *fast bilateral filter* [PARIS; DURAND, 2006], podemos utilizar as suas equações para reduzir a Equação 5.1 à equação básica do filtro bilateral (Equação 2.1).

Para realizarmos a derivação do equacionamento que suporta o nosso novo algoritmo de tone mapping reverso, iniciamos pela definição do filtro bilateral cruzado (Equa-



Figura 5.1: Demonstração da qualidade das BEFs geradas pela técnica proposta para imagens superexpostas. **Topo:** Imagens LDR superexpostas. **Segunda linha:** BEFs geradas utilizando o método de KOVALESKI; OLIVEIRA [2009]. **Baixo:** BEFs geradas com a técnica proposta. Observe a maior definição da BEF gerada utilizando a técnica proposta nas arestas em torno de áreas saturadas da imagem LDR, como a lateral do edifício na imagem `building04` e o grafite na imagem `graffiti04`.

ção 2.2):

$$I_p^c = \frac{1}{W_p^c} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) I_q \quad (5.2)$$

$$W_p^c = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) \quad (5.3)$$

Assim como qualquer média ponderada, as duas linhas da equação acima são similares. Enfatiza-se a similaridade multiplicando os dois lados da Equação 5.2 por W_p^c e reescrevendo-a utilizando vetores 2D:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) \begin{pmatrix} I_q \\ 1 \end{pmatrix} \quad (5.4)$$

G_{σ_s} e G_{σ_r} são funções Gaussianas, S é o domínio espacial, I é a imagem de entrada, E é a imagem contendo as arestas e I^b o resultado do filtro bilateral cruzado. Para manter

a propriedade que o filtro bilateral é uma média ponderada, atribui-se um peso $W_q = 1$ aos valores de entrada:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - E_q|) \begin{pmatrix} W_q I_q \\ W_q \end{pmatrix} \quad (5.5)$$

Ao atribuir um par $(W_q I_q, W_q)$ a cada pixel q , expressa-se os pixels filtrados como uma combinação linear dos pixels adjacentes. A divisão não foi removida, pois para acessar o valor correto precisa-se dividir a primeira coordenada (WI) pela segunda (W). Os autores chamam este vetor bidimensional de *intensidade homogênea*, devido à sua similaridade com as coordenadas homogêneas utilizadas em geometria projetiva. No entanto, embora a Equação 5.5 seja uma combinação linear, ela não define um filtro linear porque os pesos das Gaussianas dependem dos valores dos pixels.

O produto das Gaussianas espacial e de intensidade define uma Gaussiana no espaço 3D, produto do domínio (S) e intensidade dos pixels (R) da imagem. Como os pesos estão em 3D mas a soma da Equação 2.1 é somente sobre o domínio espacial (2D), não é possível definir uma convolução. Para isso ser possível, introduz-se uma nova dimensão ζ e define-se a intensidade I para cada ponto (x, y, ζ) . Com o símbolo de Kronecker $\delta(\zeta)\{\delta(0) = 1, \delta(\zeta \neq 0) = 0\}$ e R o intervalo em que a intensidade de um pixel é definida, reescreve-se a Equação 5.5 utilizando $\delta(\zeta - E_q)$, para que os termos sejam cancelados quando $\zeta \neq E_q$:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{q \in S} \sum_{\zeta \in R} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|E_p - \zeta|) \delta(\zeta - E_q) \begin{pmatrix} W_q I_q \\ W_q \end{pmatrix} \quad (5.6)$$

A Equação 5.6 é uma soma sobre o espaço 3D $S \times R$. Para as funções declaradas sobre este espaço utilizam-se letras minúsculas. O produto $G_{\sigma_s} G_{\sigma_r}$ define um kernel Gaussiano separável g_{σ_s, σ_r} em $S \times R$. Cria-se duas outras funções, i e w , para o resto da Equação 5.6:

$$g_{\sigma_s, \sigma_r} : (x \in S, \zeta \in R) \mapsto G_{\sigma_s}(\|x\|) G_{\sigma_r}(|\zeta|) \quad (5.7)$$

$$i : (x \in S, \zeta \in R) \mapsto I_x \quad (5.8)$$

$$w : (x \in S, \zeta \in R) \mapsto \delta(\zeta - E_x) W_x \quad (5.9)$$

Utilizando as definições 5.8 e 5.9, reescreve-se a parte da direita da Equação 5.6:

$$\delta(\zeta - E_q) \begin{pmatrix} W_q I_q \\ W_q \end{pmatrix} = \begin{pmatrix} \delta(\zeta - E_q) W_q I_q \\ \delta(\zeta - E_q) W_q \end{pmatrix} = \begin{pmatrix} w(q, \zeta) i(q, \zeta) \\ w(q, \zeta) \end{pmatrix} \quad (5.10)$$

E utilizando a definição 5.7, temos:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{(q, \zeta) \in S \times R} g_{\sigma_s, \sigma_r}(p - q, E_p - \zeta) \begin{pmatrix} w(q, \zeta) i(q, \zeta) \\ w(q, \zeta) \end{pmatrix} \quad (5.11)$$

Esta fórmula corresponde ao valor do ponto (p, E_p) de uma convolução entre g_{σ_s, σ_r} e a função bidimensional (wi, w) :

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \left[g_{\sigma_s, \sigma_r} \otimes \begin{pmatrix} wi \\ w \end{pmatrix} \right] (p, E_p) \quad (5.12)$$

Com a equação acima, introduz-se as funções i^c e w^c , expressando o filtro bilateral cruzado como uma convolução seguida por operações não-lineares (*slicing* + divisão):

$$\mathbf{linear:} \quad (w^c i^c, w^c) = g_{\sigma_s, \sigma_r} \otimes (wi, w) \quad (5.13)$$

$$\mathbf{n\~ao-linear:} \quad I_p^c = \frac{w^c(p, E_p) i^c(p, E_p)}{w^c(p, E_p)} \quad (5.14)$$

Como desejamos utilizar uma imagem diferente para a operação de *slicing*, avaliamos as funções $w^c i^c$ e w^c no ponto (p, A_p) , onde A é uma terceira imagem.

$$(w^c i^c, w^c) = g_{\sigma_s, \sigma_r} \otimes (wi, w) \quad (5.15)$$

$$I_p^c = \frac{w^c(p, E_p) i^c(p, E_p)}{w^c(p, E_p)} \Rightarrow \frac{w^c(p, A_p) i^c(p, A_p)}{w^c(p, A_p)} \quad (5.16)$$

Isto corresponde ao valor do ponto (p, A_p) em uma convolução entre g_{σ_s, σ_r} e a função (wi, w) :

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \left[g_{\sigma_s, \sigma_r} \otimes \begin{pmatrix} wi \\ w \end{pmatrix} \right] (p, A_p) \quad (5.17)$$

que pode ser reescrita de acordo com a Equação 5.11:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{(q, \zeta) \in S \times R} g_{\sigma_s, \sigma_r}(p - q, A_p - \zeta) \begin{pmatrix} w(q, \zeta) i(q, \zeta) \\ w(q, \zeta) \end{pmatrix} \quad (5.18)$$

Utilizando as definições 5.7, 5.8 e 5.9, juntamente com a Equação 5.10, podemos representar a equação acima por:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{q \in S} \sum_{\zeta \in R} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - \zeta|) \delta(\zeta - E_q) \begin{pmatrix} W_q I_q \\ W_q \end{pmatrix} \quad (5.19)$$

Eliminando o símbolo de Kronecker (δ), utilizado para representar o filtro bilateral com uma nova dimensão, temos:

$$\begin{pmatrix} W_p^c I_p^c \\ W_p^c \end{pmatrix} = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - E_q|) \begin{pmatrix} W_q I_q \\ W_q \end{pmatrix} \quad (5.20)$$

Finalmente, como $W_q = 1$, pode-se representar esta equação da mesma forma que o filtro bilateral:

$$I_p^c = \frac{1}{W_p^c} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - E_q|) I_q \quad (5.21)$$

$$W_p^c = \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|A_p - E_q|)$$

A Equação 5.21 torna possível utilizar o trabalho de KOVALESKI; OLIVEIRA [2009] com diferentes técnicas de filtro bilateral. No entanto, ainda são necessárias modificações específicas à técnica de aceleração utilizada, o que pode ser muito trabalhoso e nem sempre possível. Na próxima seção demonstramos que é possível eliminar uma das imagens para facilitar o uso de outros algoritmos de aceleração do filtro bilateral, e que esta mudança promove a geração de uma BEF mais apropriada para imagens com diferentes condições de exposição.

5.2 Simplificação

A Equação 5.21, com seus parâmetros substituídos de acordo com o trabalho de KOVALESKI; OLIVEIRA [2009], corresponde a:

$$\begin{aligned} \text{Bef}_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|L_p - M_q|) T_q \\ W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|L_p - M_q|) \end{aligned} \quad (5.22)$$

onde L , M e T são as mesmas imagens descritas na Equação 5.1: cada pixel da imagem M contém o valor máximo entre os componentes RGB de I , L contém a imagem de luminância de I , e T contém todos os pixels de M que estão acima de um limite t , que depende do conteúdo processado (254 para imagens, 230 para vídeos). A imagem I tem seus valores linearizados utilizando uma curva gamma de 2.2 antes de gerar as imagens L , M e T .

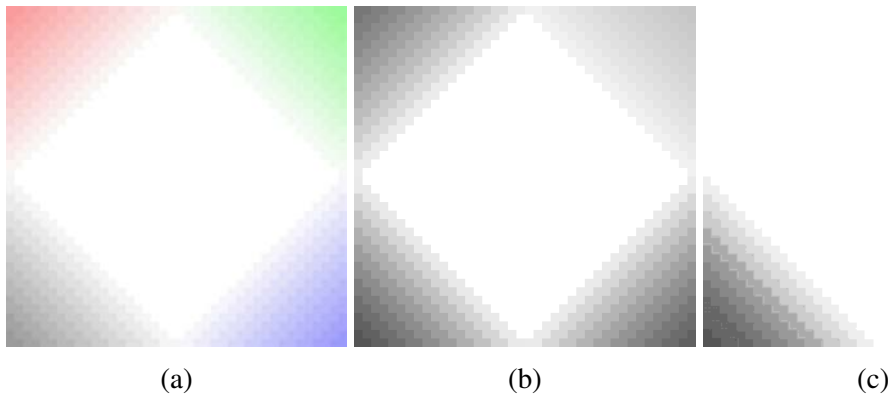


Figura 5.2: Exemplo de imagens de suporte para a técnica descrita em KOVALESKI; OLIVEIRA [2009]. **(a)** Imagem sintética de entrada. **(b)** Imagem em tons de cinza da imagem de entrada. **(c)** Imagem contendo apenas o valor máximo entre os canais R, G e B de cada pixel. Observe a falta de detalhes nas áreas contendo gradientes coloridos. Para a técnica proposta nesta dissertação, apenas a imagem (b) é utilizada.

É possível observar que a imagem de máximos, utilizada durante a filtragem na técnica descrita em KOVALESKI; OLIVEIRA [2009], não tem uma relação direta com o conteúdo da imagem de entrada, como podemos ver na Figura 5.2. Estruturas muito iluminadas na cena podem possuir um valor máximo similar, e o uso deste valor acarreta em perda de informação de arestas e detalhes na imagem (Figura 5.2c), levando a erros na geração da BEF e a funções menos precisas. Este problema é acentuado em imagens superexpostas, onde os valores máximos entre os canais de um pixel podem ser similares em grandes áreas da imagem. Isto pode ser observado na Figura 5.3, onde a imagem sintética exibida na Figura 5.2a foi utilizada para gerar imagens HDR utilizando a técnica proposta e o método descrito por KOVALESKI; OLIVEIRA [2009]. Como a técnica proposta utiliza apenas a imagem em tons de cinza (Figura 5.2b) durante a filtragem, a BEF gerada contém mais detalhes em comparação com o resultado da técnica de KOVALESKI; OLIVEIRA [2009]. Isto resulta em uma imagem HDR com maior ampliação de contraste entre as arestas presentes na imagem de entrada.

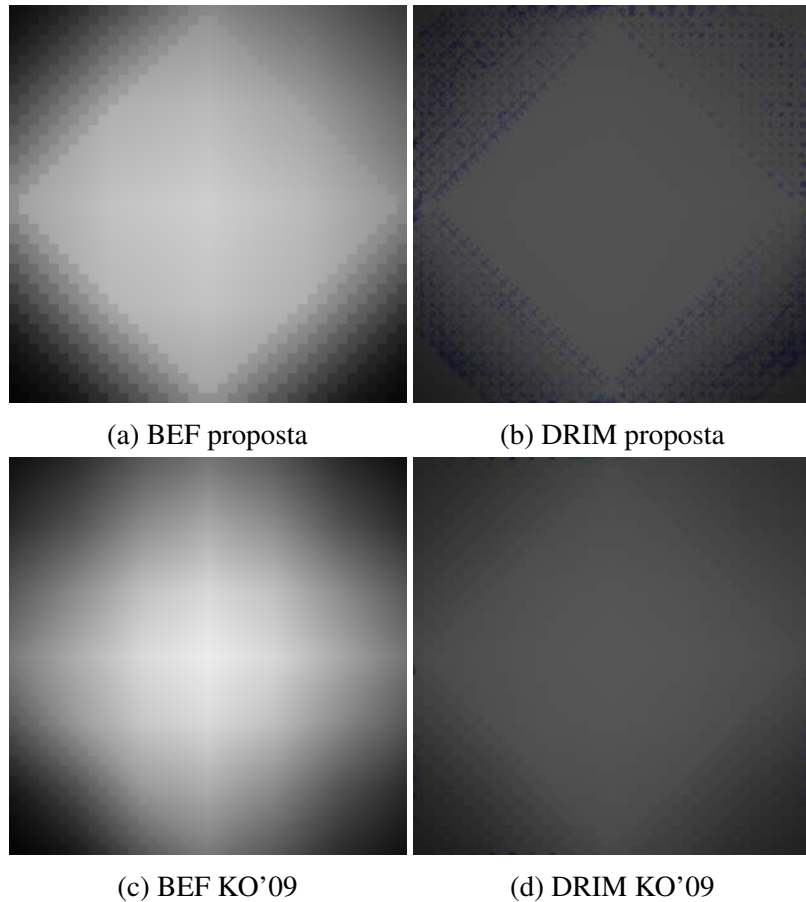


Figura 5.3: Vantagens da utilização de uma imagem de tons de cinza no lugar da imagem de valor máximo entre canais de um pixel. Estas imagens foram geradas utilizando a Figura 5.2a como entrada. Observe os detalhes mantidos na BEF da técnica proposta (imagem (a)), em comparação com a BEF do trabalho de KOVALESKI; OLIVEIRA [2009] (imagem (c)). Isto acontece devido ao uso da imagem em tons de cinza (Figura 5.2b) no lugar da imagem de máximos (Figura 5.2c) durante a filtragem. Esta alteração promove uma maior ampliação no contraste entre as arestas presentes na imagem de entrada, como demonstra o resultado do DRIM nas imagens (b), para a técnica proposta, e (d), para a técnica de KOVALESKI; OLIVEIRA [2009], com a cor azul.

Por este motivo, substituímos a imagem de máximos por outra cópia da imagem de luminância na Equação 5.21:

$$\begin{aligned}
 Bef_p^b &= \frac{1}{W_p^b} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|L_p - L_q|) T_q \\
 W_p^b &= \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|L_p - L_q|)
 \end{aligned} \tag{5.23}$$

onde L corresponde à imagem de entrada em tons de cinza e os pixels de T são iguais a 1 se o pixel correspondente na imagem de entrada possui pelo menos um canal de cor acima do limite t .

A equação resultante é equivalente à do filtro bilateral cruzado (Equação 2.2). Este filtro, que processa os dados de entrada de acordo com arestas de uma segunda imagem, é implementado facilmente por diversas técnicas de aceleração do filtro bilateral, o que simplifica o uso de nossa técnica.



Figura 5.4: Comparação entre espaços de cor RGB e CIELAB, considerando tanto casos de imagens escuras (esquerda e direita) quanto de imagens claras. **Topo:** Imagens de entrada. **Segunda linha:** BEFs geradas utilizando o filtro bilateral com o espaço de cor RGB. **Terceira linha:** BEFs geradas utilizando o filtro bilateral com o espaço de cor CIELAB. **Quarta linha:** Resultado do DRIM, comparando os dois resultados anteriores. Note que não há diferenças significativas detectadas pelo algoritmo.

No trabalho de KOVALESKI; OLIVEIRA [2009], as imagens são transformadas de RGB para o espaço de cor CIELAB (ou $L^*a^*b^*$) antes de serem introduzidas no *bilateral grid*. No entanto, apenas o canal L^* é utilizado durante a filtragem e geração da BEF. A diferença entre o canal L^* normalizado e a imagem de luminância (L) é muito pequena, e não é possível detectar diferenças significativas na imagem HDR gerada pelas BEFs calculadas em CIELAB ou RGB (Figura 5.4). No intuito de reduzir a complexidade e melhorar o desempenho do algoritmo, a transformação entre RGB e CIELAB foi eliminada da etapa de criação das BEFs, o que facilita ainda mais a utilização da nova técnica descrita nesta dissertação.

Além destas mudanças, os parâmetros do filtro bilateral foram modificados. No trabalho de KOVALESKI; OLIVEIRA [2009], os parâmetros eram $\sigma_s = 100$ e $\sigma_r = 0.30$. Na técnica proposta, utilizamos $\sigma_s = 150$, que em um monitor HDR de 37" com resolução de 1920×1080 corresponde a 1.2 graus à uma distância de visualização de 3 metros. Desta maneira, o espectro do filtro contém principalmente baixas frequências angulares, de 0.5 ciclos ou menos, para as quais o sistema visual humano não é muito sensível [REMPEL et al., 2007]. Observamos, também, que a utilização de um desvio padrão menor para a Gaussiana de intensidades auxilia na detecção de arestas em imagens superexpostas. Portanto, utilizamos $\sigma_r = 0.25$ para nossos resultados, valor obtido através de experimentação com diferentes imagens.

Como é possível ver nas imagens das Figuras 5.5 e 5.6, as BEFs geradas com diferentes métodos de filtro bilateral são extremamente similares. As diferenças que ocorrem são apenas devido à maneira como o algoritmo acelera a obtenção da imagem filtrada, mas não são significativas a ponto de comprometer a qualidade das imagens HDR resultantes.

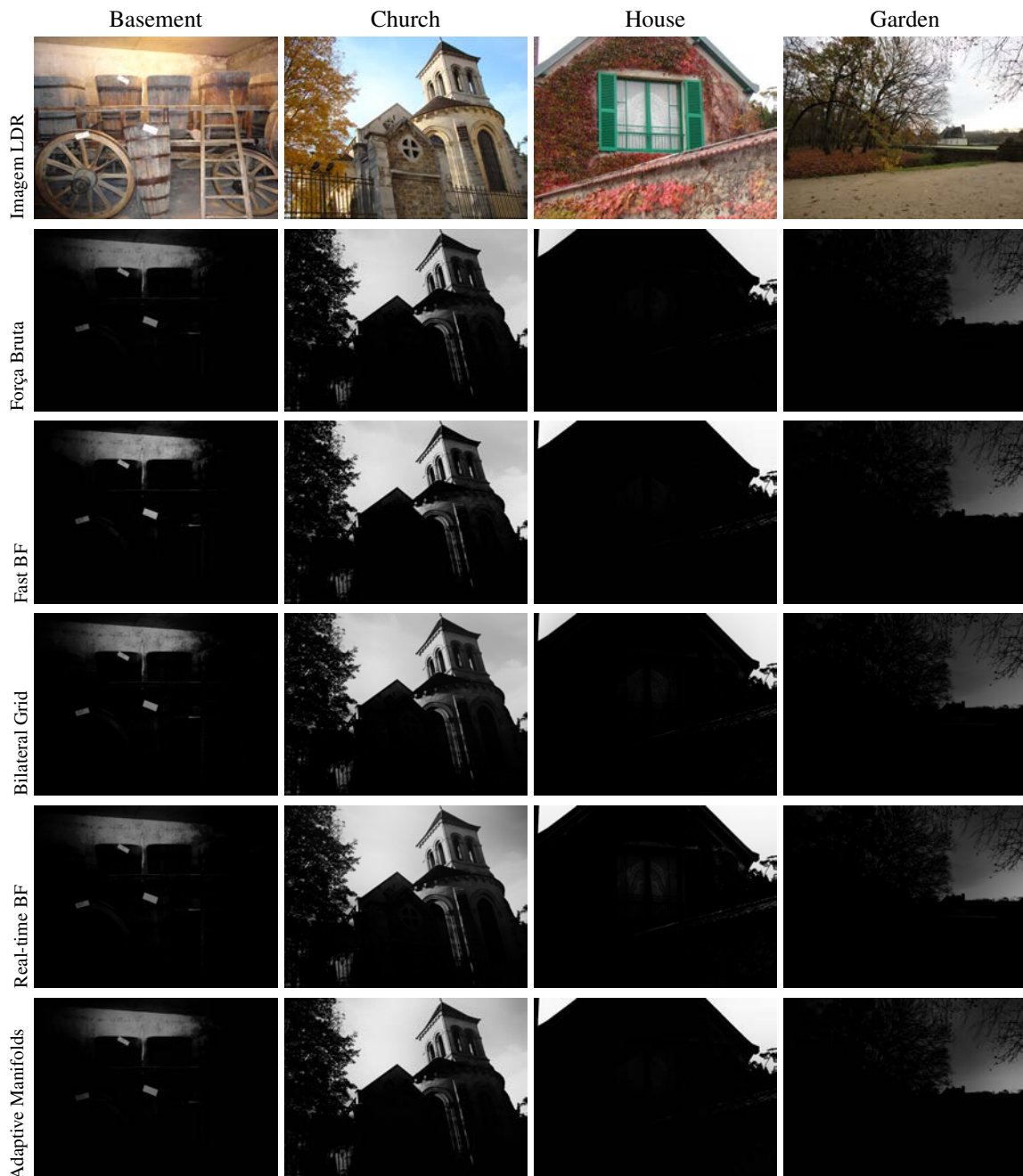


Figura 5.5: Demonstração de BEFs obtidas com a técnica descrita nesta dissertação explorando diferentes técnicas de filtragem bilateral. **Topo:** Imagens de entrada. **Segunda linha:** BEFs geradas utilizando um filtro bilateral implementado usando uma abordagem de força bruta. **Terceira linha:** BEFs geradas com o algoritmo *fast bilateral filter* [PARIS; DURAND, 2006]. **Quarta linha:** BEFs geradas através do *bilateral grid* [CHEN; PARIS; DURAND, 2007]. **Quinta linha:** BEFs geradas utilizando o *real time bilateral filter* [YANG; TAN; AHUJA, 2009]. **Última linha:** BEFs geradas com a técnica de *adaptive manifolds* [GASTAL; OLIVEIRA, 2012].

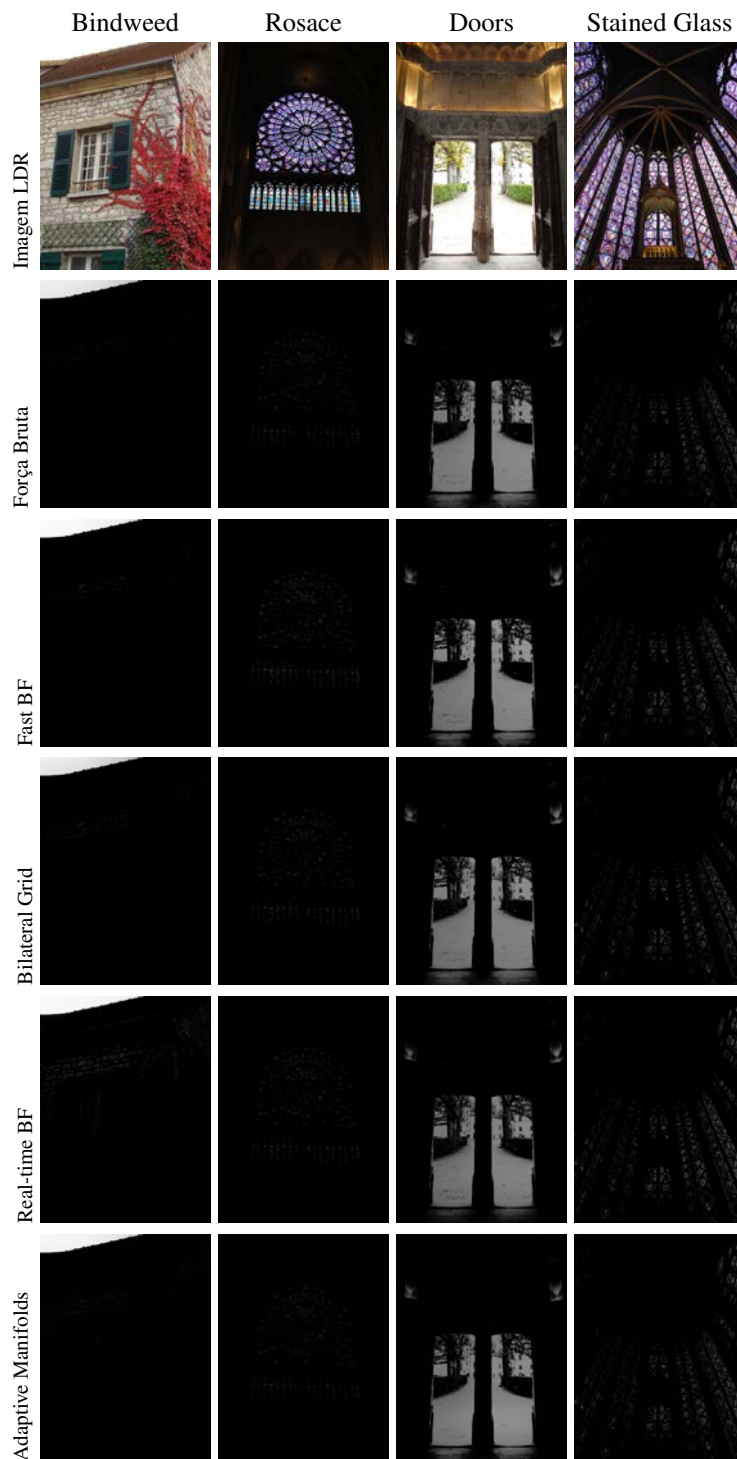


Figura 5.6: Demonstração de BEFs obtidas com a técnica descrita nesta dissertação explorando diferentes técnicas de filtragem bilateral. **Topo:** Imagens de entrada. **Segunda linha:** BEFs geradas utilizando um filtro bilateral implementado usando uma abordagem de força bruta. **Terceira linha:** BEFs geradas com o algoritmo *fast bilateral filter* [PARIS; DURAND, 2006]. **Quarta linha:** BEFs geradas através do *bilateral grid* [CHEN; PARIS; DURAND, 2007]. **Quinta linha:** BEFs geradas utilizando o *real time bilateral filter* [YANG; TAN; AHUJA, 2009]. **Última linha:** BEFs geradas com a técnica de *adaptive manifolds* [GASTAL; OLIVEIRA, 2012].

6 RESULTADOS

MASIA et al. [2009a] compararam os resultados de seu operador de tone mapping reverso contra três outros operadores: LDR2HDR [REMPEL et al., 2007], BANTERLE et al. [2006] e escala linear de contraste [AKYÜZ et al., 2007]. Para estas comparações, eles utilizaram diversas imagens capturadas com diferentes tempos de exposição. Em seu trabalho, mostraram que seu operador tem um desempenho melhor que as técnicas comparadas para imagens superexpostas, com um desempenho similar para imagens apropriadamente expostas ou subexpostas. No entanto, seu estudo não considerou o operador de KOVALESKI; OLIVEIRA [2009].

A fim de validar a técnica descrita nesta dissertação, comparamos nossos resultados com aqueles gerados por MASIA; GUTIERREZ [2011] e por KOVALESKI; OLIVEIRA [2009]. Como o trabalho de Masia e Gutierrez não utiliza uma função de aumento de brilho (BEF), utilizaremos a técnica de avaliação de imagens DRIM (introduzida na seção 4.4) para comparar imagens HDR geradas por cada uma destas técnicas.

Para gerar imagens HDR, usamos o mesmo algoritmo utilizado por REMPEL et al. [2007] e KOVALESKI; OLIVEIRA [2009]: a imagem de entrada I tem seus valores linearizados utilizando-se uma curva gamma de 2.2 ($I_l = I^{2.2}$) que, de acordo com [REMPEL et al., 2007], é utilizada em formatos padrões de vídeo e televisão. A partir da imagem linearizada I_l e utilizando a técnica descrita na Seção 5.2, geramos uma BEF B , que deve ter seus valores escalados para o intervalo $[1..\alpha]$, onde α é um valor dependente do monitor utilizado. Para o monitor Dolby DR-37P e para nossos testes, utilizamos um valor $\alpha = 4$. Em seguida, a imagem linearizada I_l tem seus valores escalados de acordo com a capacidade do monitor: utilizamos um valor de 0.3 cd/m^2 para o ponto preto e 1200 cd/m^2 para o ponto branco do monitor. A imagem HDR final é obtida através da multiplicação ponto a ponto da nossa BEF B pela imagem de entrada I_l , cujos valores foram escalados de acordo com a capacidade do monitor utilizado.

No trabalho de MASIA; GUTIERREZ [2011], três diferentes equações são apresentadas para calcular o valor γ utilizado por sua técnica. A fim de comparar seus resultados com os nossos, utilizamos sua terceira equação, pois, de acordo com os autores, apresenta uma maior robustez que as outras. Embora já apresentada na Seção 3.2.3, resumimos sua técnica a seguir.

Inicialmente, a imagem é linearizada de acordo com o espaço de cores sRGB, aplicando a seguinte equação a cada um de seus canais:

$$I_l = \begin{cases} \frac{I_{srgb}}{12.92} & I_{srgb} \leq 0.04045 \\ \left(\frac{I_{srgb}+a}{1+a}\right)^{2.4} & I_{srgb} > 0.04045 \end{cases} \quad (6.1)$$

A luminância da imagem é calculada a partir dos valores sRGB linearizados, obtidos

com a Equação 6.1:

$$L = 0.2126R + 0.7152G + 0.0722B \quad (6.2)$$

onde L é normalizado entre $[0..1]$.

Em seguida, uma série de estatísticas são calculadas, iniciando pela média geométrica da luminância da imagem:

$$L_H = \exp\left(\frac{1}{N} \sum_{i=1}^N \log(L(i) + \epsilon)\right) \quad (6.3)$$

onde ϵ é um valor muito pequeno, para evitar singularidades quando $L(i) = 0$.

O valor chave (*key*) k_1 da imagem é calculado, que corresponde à intensidade média da imagem, ignorando 1% dos maiores e menores valores dos pixels de L :

$$k_1 = \frac{\log L_H - \log L_{min}}{\log L_{max} - \log L_{min}} \quad (6.4)$$

Finalmente, a porcentagem de pixels sobreexpostos p_{se} é calculada, definindo pixels sobreexpostos como aqueles onde $L \cdot 255 \geq 254$. Utilizando estas estatísticas, o valor γ é calculado através da seguinte equação:

$$\gamma = 2.4379 + 0.2319 \log L_H - 1.1228k_1 + 0.0085p_{se} \quad (6.5)$$

Para gerar suas imagens HDR, simplesmente aplicamos o valor γ calculado à imagem linearizada (I_l^γ) e expandimos seus valores linearmente, até os limites do monitor utilizado (para nossos testes, utilizamos 0.3 cd/m^2 para o ponto preto e 4800 cd/m^2 para o ponto branco da imagem).

Para nossos testes de comparação, escolhemos as imagens utilizadas em [MASIA et al., 2009a], disponibilizadas em [MASIA et al., 2009b], pois apresentam diversas situações e estruturas sub e superexpostas. De acordo com MASIA et al. [2009a], as imagens *building* (Figura 6.1), *graffiti* (Figura 6.2) e *sunset* (Figura 6.3) correspondem a séries claras (superexpostas), enquanto as imagens *car* (Figura 6.6), *pencils* (Figura 6.7) e *flowers* (Figura 6.8) correspondem a séries escuras (subexpostas). Para minimizar diferenças devido à técnica de aceleração do filtro bilateral entre a técnica proposta e [KOVALESKI; OLIVEIRA, 2009], foi utilizado o bilateral grid [CHEN; PARIS; DURAND, 2007] para gerar nossas BEFs. Também incluímos as representações em LDR do resultado de cada operador de tone mapping reverso, utilizando o algoritmo de tone mapping de REINHARD et al. [2002], para uma melhor representação do resultado das técnicas.

Pode-se ver, na série clara de imagens (Figuras 6.1, 6.2 e 6.3), que os resultados obtidos pela nossa técnica são equivalentes aos resultados obtidos pelo algoritmo de MASIA; GUTIERREZ [2011], e relativamente melhores que as imagens resultantes do trabalho de KOVALESKI; OLIVEIRA [2009]. É possível notar, para todas as técnicas, um aumento de contraste em áreas onde há uma virtual perda de detalhe (como o grafite na Figura 6.2, *graffiti03-04*, e o poste de luz na Figura 6.1, *building04*), demonstrados pela cor azul nas imagens do DRIM.

No entanto, é possível notar grandes áreas de perda de contraste (em verde, no resultado do DRIM) e reversão de contraste (em vermelho) por parte do trabalho de MASIA; GUTIERREZ [2011] na série escura de imagens (Figuras 6.6, 6.7 e 6.8), enquanto os resultados de KOVALESKI; OLIVEIRA [2009] e os da técnica proposta não apresentam

perda de contraste, e demonstram uma menor quantidade de reversão de contraste. Isto pode ser notado especialmente na Figura 6.6, imagens `car01-02`, e Figura 6.8, imagens `flowers01-02`. Além do resultado do DRIM, é possível observar problemas nas imagens de *tonemap* obtidas a partir das imagens HDR resultantes do trabalho de MASIA; GUTIERREZ [2011], o que não acontece para a técnica proposta. O operador de *tone mapping* utilizado foi a técnica local proposta por REINHARD et al. [2002].

Em geral, os resultados obtidos com a técnica proposta nesta dissertação apresentam desempenho superior ao da técnica de KOVALESKI; OLIVEIRA [2009]. Isto demonstra que as modificações propostas pela técnica aqui apresentada são efetivas para melhorar a BEF obtida, resultando em uma imagem com menos artefatos indesejados e maior contraste.

Pode-se afirmar, com base nestes resultados, que o algoritmo apresentado em [MASIA; GUTIERREZ, 2011] não é adequado para cenas que apresentam uma grande faixa dinâmica e que contêm grandes áreas com pixels subexpostos. A técnica proposta, por outro lado, apresenta ótimos resultados nestes casos, e se mantém estável em todas as imagens com diferentes tempos de exposição.

Também demonstramos o uso de nossa técnica para criar vídeos HDR. O algoritmo é aplicado independentemente em cada frame do vídeo. Podemos observar, nas Figuras 6.4 e 6.5, que as BEFs geradas são suaves e não apresentam descontinuidades entre frames. Finalmente, o resultado do DRIM é consistente com as imagens obtidas anteriormente.

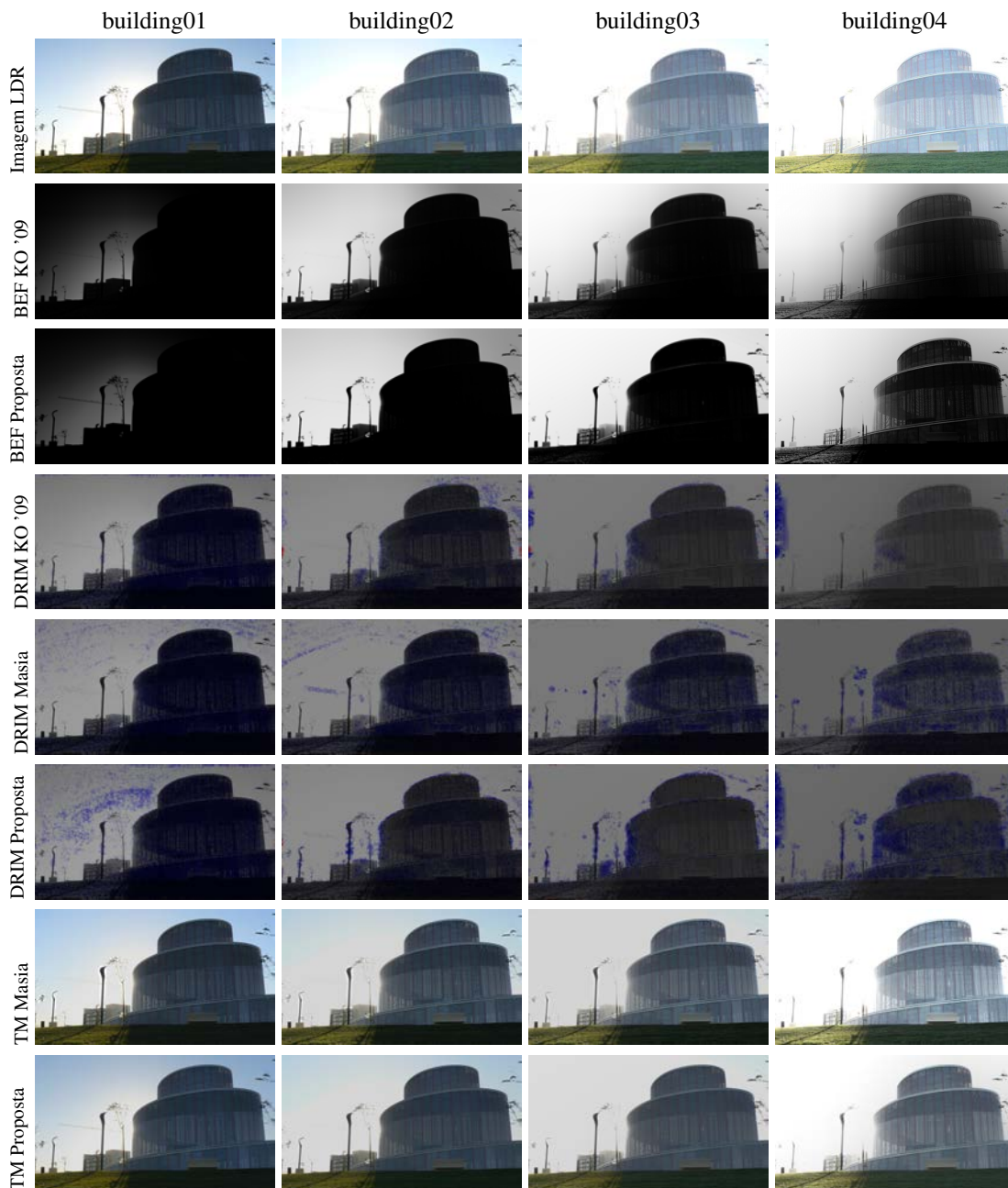


Figura 6.1: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série *building*. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. A cor azul, no DRIM, representa ampliação de contraste invisível. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta.

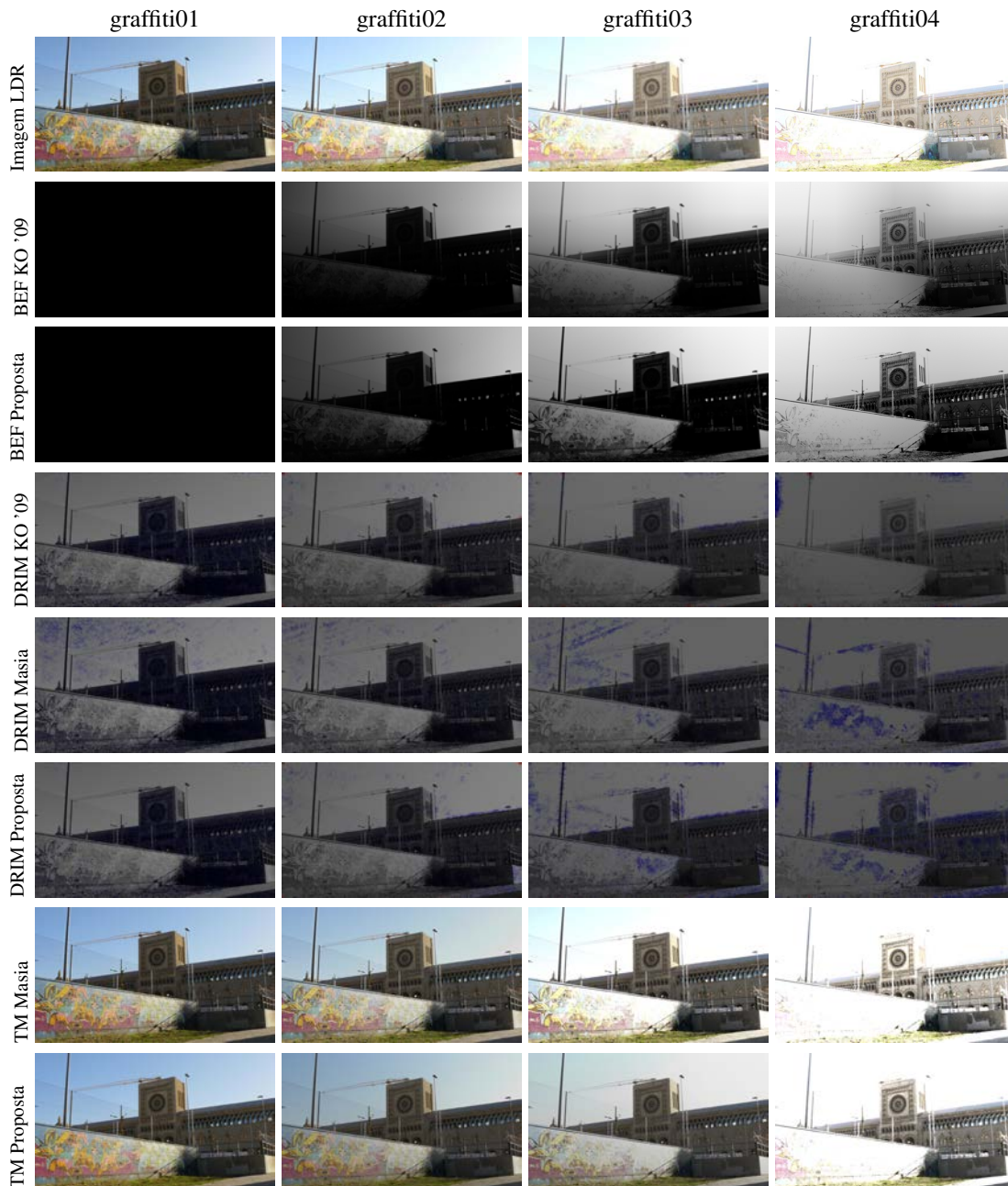


Figura 6.2: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série *graffiti*. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. A cor azul, no DRIM, representa ampliação de contraste invisível. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta.

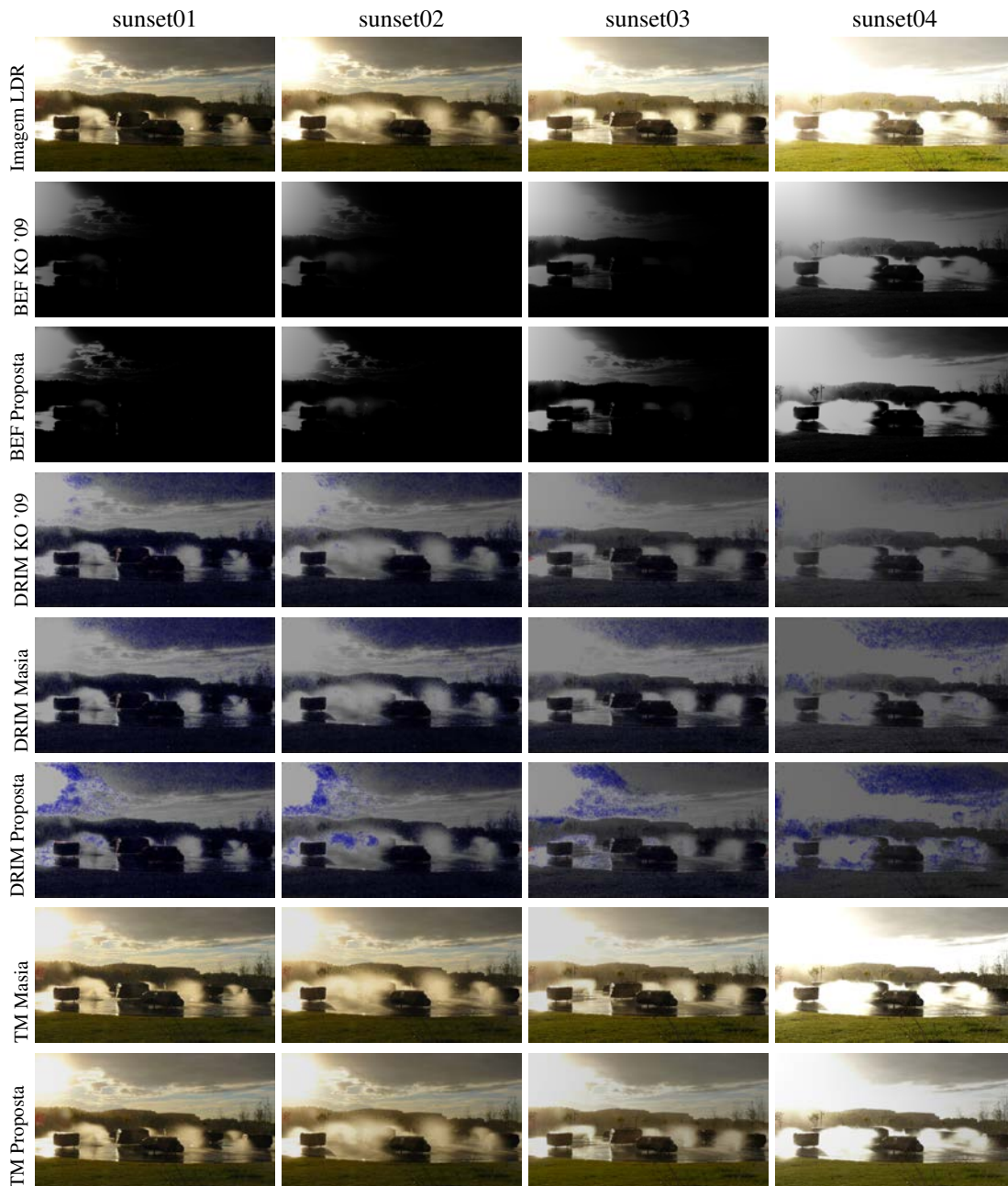


Figura 6.3: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série *graffiti*. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. A cor azul, no DRIM, representa ampliação de contraste invisível. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta.

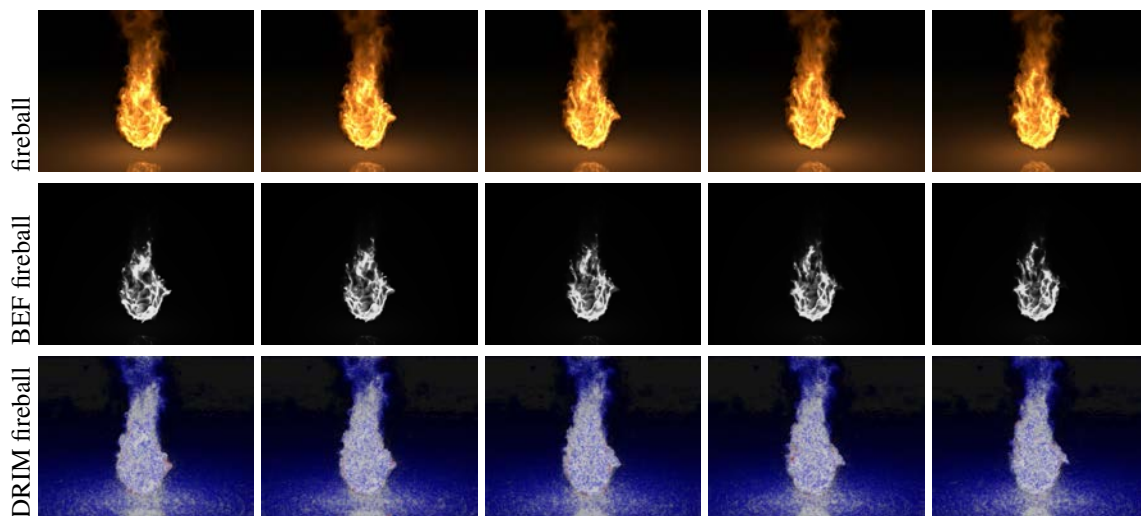


Figura 6.4: Aplicação da técnica proposta em vídeos. **Cima:** Cinco frames consecutivos do vídeo *fireball*. **Meio:** BEFs geradas para cada um dos frames individualmente. Observe que não há descontinuações entre cada BEF. **Baixo:** Resultado do DRIM para cada frame. Note a amplificação de contraste invisível nas áreas claras e escuras da imagem.

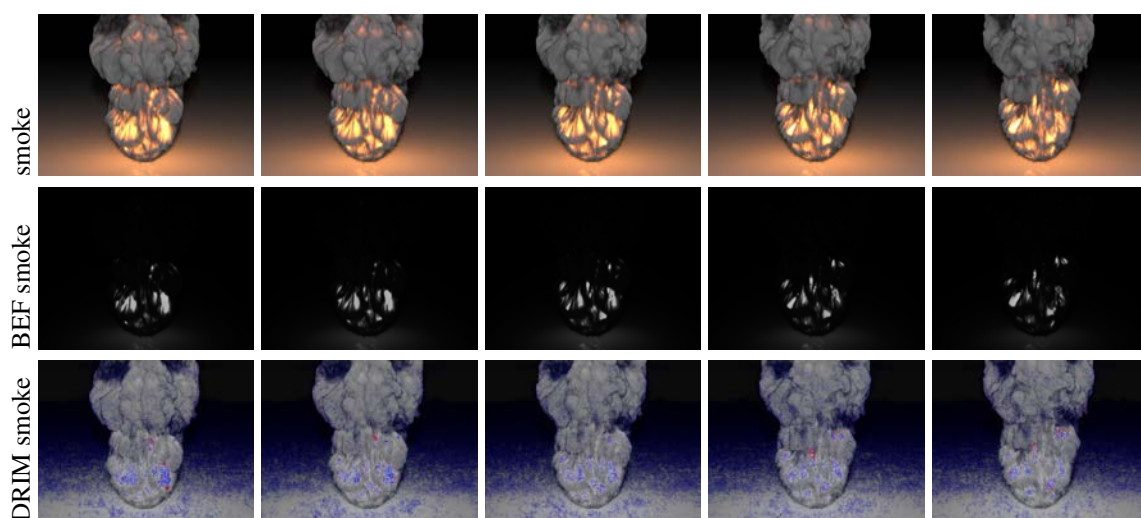


Figura 6.5: Aplicação da técnica proposta em vídeos. **Cima:** Cinco frames consecutivos do vídeo *smoke*. **Meio:** BEFs geradas para cada um dos frames individualmente. Observe que não há descontinuações entre cada BEF. **Baixo:** Resultado do DRIM para cada frame. Note a amplificação de contraste invisível nas áreas claras e escuras da imagem.

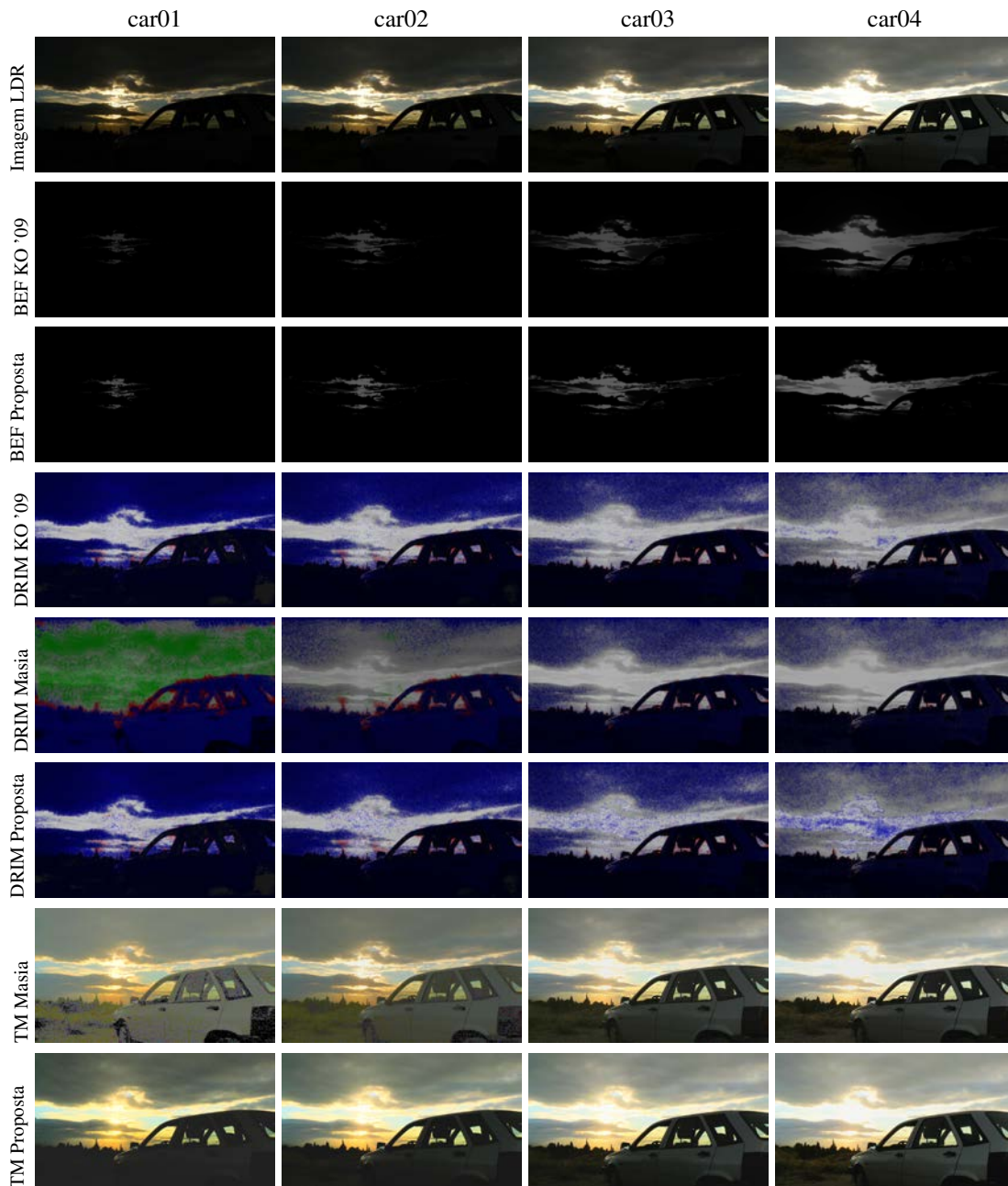


Figura 6.6: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série *car*. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta. Observe as grandes áreas com perda de contraste (em verde) e reversão de contraste (em vermelho) no resultado do DRIM para Masia e Gutierrez, e o resultado anômalo do operador de *tone mapping* para sua técnica, especialmente nas imagens mais escuras.

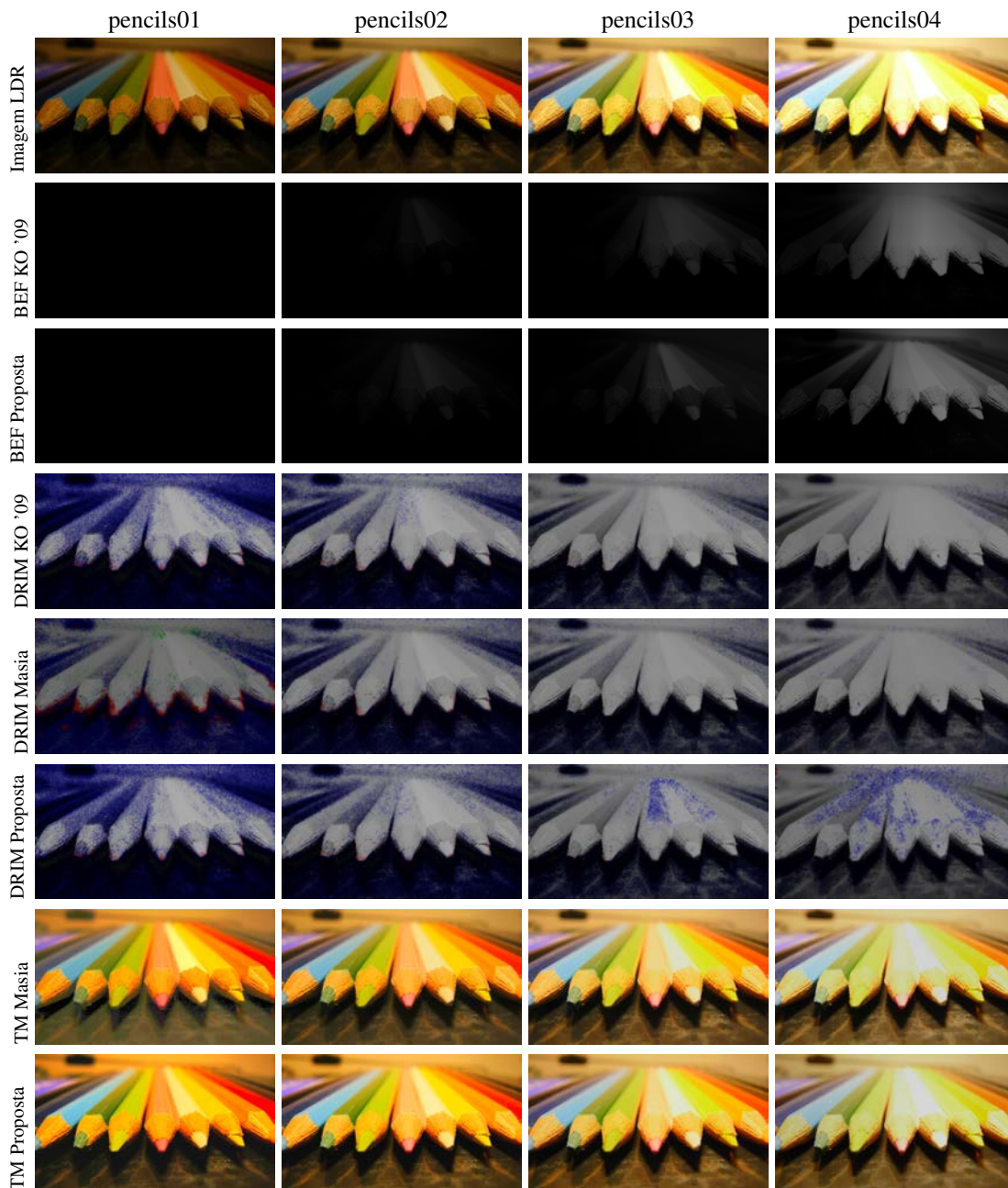


Figura 6.7: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série `pencils`. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta. Observe as áreas com perda de contraste (em verde) e reversão de contraste (em vermelho) no resultado do DRIM para Masia e Gutierrez, especialmente nas imagens mais escuras.

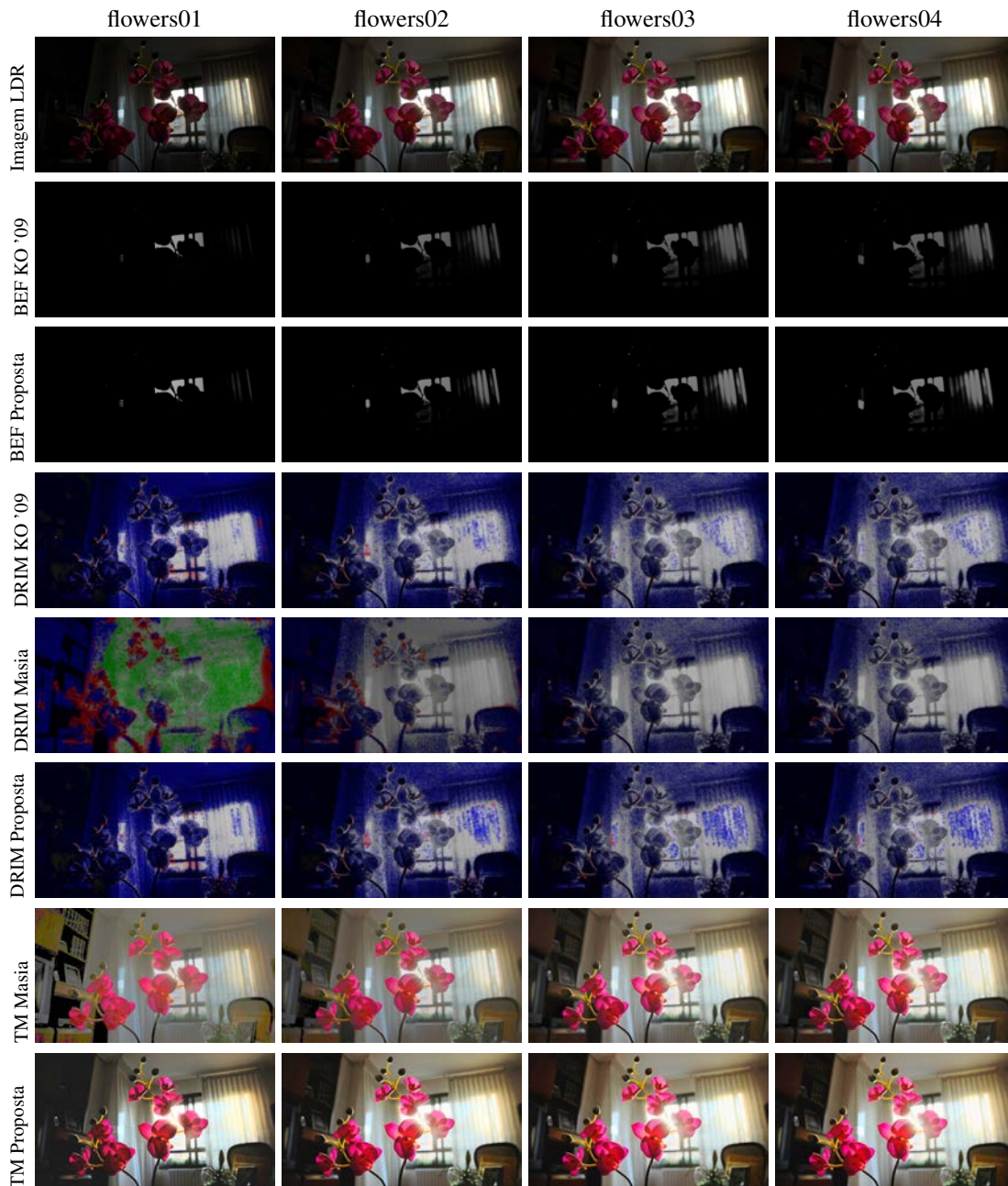


Figura 6.8: Comparação entre os resultados obtidos por Masia e Gutierrez, Kovaleski e Oliveira e a técnica proposta. **Topo:** Imagem de entrada, série *flowers*. Cada imagem foi obtida com um tempo de exposição progressivamente maior. **Segunda linha:** BEF, técnica proposta. **Terceira linha:** BEF, Kovaleski e Oliveira. **Quarta linha:** Resultado do DRIM para Kovaleski e Oliveira. **Quinta linha:** Resultado do DRIM para Masia e Gutierrez. **Sexta linha:** Resultado do DRIM para a técnica proposta. **Sétima linha:** Imagem de *tonemap* do resultado obtido por Masia e Gutierrez, utilizando a técnica de REINHARD et al. [2002]. **Última linha:** Imagem de *tonemap* do resultado obtido pela técnica proposta. Observe as grandes áreas com perda de contraste (em verde) e reversão de contraste (em vermelho) no resultado do DRIM para Masia e Gutierrez, e o resultado anômalo do operador de *tone mapping* para sua técnica, especialmente nas imagens mais escuras.

7 CONCLUSÃO

Esta dissertação apresentou um novo operador de tone mapping reverso para imagens e vídeos, capaz de suportar uma alta gama de condições de exposição. Estes operadores são importantes para a recuperação e demonstração de mídia legada em monitores capazes de exibir uma faixa dinâmica muito maior que o de dispositivos de visualização convencionais. O suporte a diferentes condições de exposição torna a técnica proposta adequada não somente para conteúdo obtido profissionalmente, mas também para imagens e vídeos de uso pessoal.

Através do uso do filtro bilateral cruzado, produzimos funções de aumento de brilho (BEFs) suaves e que preservam as arestas presentes na imagem original. Nossa técnica pode ser utilizada diretamente com qualquer método de aceleração do filtro bilateral, liberando o usuário das limitações de uma implementação específica.

Foram realizadas comparações objetivas entre nossos resultados e aqueles produzidos pelos operadores de MASIA; GUTIERREZ [2011] e KOVALESKI; OLIVEIRA [2009] para diversas imagens com diferentes tempos de exposição utilizando a métrica de comparação de imagens DRIM [AYDIN et al., 2008]. Estes experimentos mostram que nosso método é o único capaz de realçar detalhes imperceptíveis em imagens que contemplam uma ampla gama de exposições, variando desde imagens subexpostas até imagens superexpostas. Nossa técnica tem um desempenho superior ao operador de MASIA; GUTIERREZ, que foi projetado para imagens superexpostas, mas que não tem bom desempenho para imagens subexpostas. Ele também produz resultados superiores aos do método de KOVALESKI; OLIVEIRA, que não apresenta bons resultados para imagens superexpostas. Além disso, demonstramos que nossa técnica pode ser usada para gerar vídeos HDR.

Por expandir a gama de exposições de imagens e vídeos capazes de serem realçadas e por permitir ao usuário a flexibilidade de escolher entre qualquer técnica atual (ou futura) de aceleração de filtro bilateral, este novo operador de tone mapping reverso amplia significativamente o repertório de ferramentas de imagens HDR disponível. Nosso método tem o potencial de permitir novas e criativas aplicações HDR até então indisponíveis devido à falta de um operador de tone mapping reverso capaz de lidar com uma grande gama de exposições.

REFERÊNCIAS

- ADAMS, A.; BAEK, J.; DAVIS, M. A. Fast High-Dimensional Filtering Using the Permutohedral Lattice. **Comput. Graph. Forum**, [S.l.], v.29, n.2, p.753–762, 2010.
- AKYÜZ, A. O. et al. Do HDR displays support LDR content?: a psychophysical evaluation. **ACM Trans. Graph.**, [S.l.], v.26, n.3, July 2007.
- AKYÜZ, A. O.; REINHARD, E. Color appearance in high-dynamic-range imaging. **J. Electronic Imaging**, [S.l.], v.15, n.3, p.033001, 2006.
- AURICH, V.; WEULE, J. Non-Linear Gaussian Filters Performing Edge Preserving Diffusion. In: DAGM-SYMPOSIUM. **Anais...** [S.l.: s.n.], 1995. p.538–545.
- AYDIN, T. O. et al. Dynamic range independent image quality assessment. **ACM Trans. Graph.**, [S.l.], v.27, n.3, p.69:1–69:10, Aug. 2008.
- BANTERLE, F. et al. Inverse tone mapping. In: GRAPHITE. **Anais...** [S.l.: s.n.], 2006. p.349–356.
- BANTERLE, F. et al. A framework for inverse tone mapping. **The Visual Computer**, [S.l.], v.23, n.7, p.467–478, 2007.
- BANTERLE, F. et al. Expanding low dynamic range videos for high dynamic range applications. In: SPRING CONFERENCE ON COMPUTER GRAPHICS, 24. **Proceedings...** [S.l.: s.n.], 2008. p.33–41. (SCCG '08).
- BANTERLE, F. et al. A Psychophysical Evaluation of Inverse Tone Mapping Techniques. **Comput. Graph. Forum**, [S.l.], v.28, n.1, p.13–25, 2009.
- BANTERLE, F. et al. High Dynamic Range Imaging and Low Dynamic Range Expansion for Generating HDR Content. **Comput. Graph. Forum**, [S.l.], v.28, n.8, p.2343–2367, 2009.
- BERTALMÍO, M. et al. Simultaneous Structure and Texture Image Inpainting. In: CVPR (2). **Anais...** [S.l.: s.n.], 2003. p.707–712.
- BITTECH. **BrightSide DR37-P HDR display**. Disponível em: <http://www.bittech.net/hardware/2005/10/04/brightside_hdr_edr/1>. Acesso em: oct. 2012.
- BUADES, A.; COLL, B.; MOREL, J.-M. A Non-Local Algorithm for Image Denoising. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR'05) - VOLUME 2 - VOLUME 02, 2005., Washington, DC, USA. **Proceedings...** IEEE Computer Society, 2005. p.60–65. (CVPR '05).

BURT, P. The Laplacian pyramid as a compact image code. In: IEEE TRANSACTIONS ON COMMUNICATIONS. **Anais...** [S.l.: s.n.], 1983.

CGSD. **Computer Graphics Systems Development Corp – Gamma Correction Home Page**. Disponível em: <<http://www.cgsd.com/papers/gamma.html>>. Acesso em: oct. 2012.

CHEN, J.; PARIS, S.; DURAND, F. Real-time edge-aware image processing with the bilateral grid. **ACM Transactions on Graphics**, [S.l.], v.26, n.3, p.103, 2007.

CIELAB. **ISO 11664-4 2008 (CIE S 014-4/E 2007) Colorimetry – Part 4 – CIE 1976 L*a*b* Colour space**. Disponível em: <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52497>. Acesso em: apr. 2013.

CIEXYZ. **ISO 11664-1 2007 (CIE S 014-1/E 2006) Colorimetry – Part 1 – CIE standard colorimetric observers**. Disponível em: <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52495>. Acesso em: apr. 2013.

DALY, S. The visible differences predictor: an algorithm for the assessment of image fidelity. In: WATSON, A. B. (Ed.). **Digital images and human vision**. Cambridge, MA, USA: MIT Press, 1993. p.179–206.

DALY, S. J. Application of a noise-adaptive contrast sensitivity function to image data compression. **Optical Engineering**, [S.l.], v.29, n.8, p.977–987, 1990.

DALY, S. J.; FENG, X. Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system. **Proceedings of Color Imaging VIII: Processing, Hardcopy, and Applications**, [S.l.], p.455–466, 2003.

DALY, S. J.; FENG, X. Decontouring: prevention and removal of false contour artifacts. **Human Vision and Electronic Imaging IX**, [S.l.], p.130–149, 2004.

DEBEVEC, P. A median cut algorithm for light probe sampling. In: ACM SIGGRAPH 2005 POSTERS, New York, NY, USA. **Anais...** ACM, 2005. (SIGGRAPH '05).

DEBEVEC, P. E.; MALIK, J. Recovering high dynamic range radiance maps from photographs. In: SIGGRAPH. **Anais...** [S.l.: s.n.], 1997. p.369–378.

DIDYK, P. et al. Enhancement of Bright Video Features for HDR Displays. **Comput. Graph. Forum**, [S.l.], v.27, n.4, p.1265–1274, 2008.

DOLBY. **Dolby Signs Agreement to Acquire BrightSide**. Disponível em: <<http://investor.dolby.com/ReleaseDetail.cfm?ReleaseID=231363>>. Acesso em: oct. 2012.

DRAGO, F. et al. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. In: COMPUTER GRAPHICS FORUM. **Anais...** [S.l.: s.n.], 2003. v.22, p.419–426.

DURAND, F.; DORSEY, J. Fast bilateral filtering for the display of high-dynamic-range images. In: SIGGRAPH. **Anais...** [S.l.: s.n.], 2002. p.257–266.

EISEMANN, E.; DURAND, F. Flash photography enhancement via intrinsic relighting. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2004. p.673–678.

FATTAL, R. et al. Gradient domain high dynamic range compression. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2002. p.249–256.

FERWERDA, J. A. et al. A model of visual masking for computer graphics. In: SIGGRAPH. **Anais...** [S.l.: s.n.], 1997. p.143–152.

FOLEY, J. M. Human luminance pattern-vision mechanisms: masking experiments require a new model. **J. Opt. Soc. Am. A**, [S.l.], v.11, n.6, p.1710–1719, Jun 1994.

GASTAL, E. S. L.; OLIVEIRA, M. M. Adaptive Manifolds for Real-Time High-Dimensional Filtering. **ACM TOG**, [S.l.], v.31, n.4, p.33:1–33:13, 2012. Proceedings of SIGGRAPH 2012.

HOEFFLINGER, B. **High-Dynamic-Range (HDR) Vision**: microelectronics, image processing, computer graphics (springer series in advanced microelectronics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

ITU-BT.709. **BT.709 – Parameter values for the HDTV standards for production and international programme exchange**. Disponível em: <<http://www.itu.int/rec/R-REC-BT.709/en>>. Acesso em: oct. 2012.

KANG, S. B. et al. High dynamic range video. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2003. p.319–325.

KOVALESKI, R. P.; OLIVEIRA, M. M. High-quality brightness enhancement functions for real-time reverse tone mapping. **The Visual Computer**, [S.l.], v.25, n.5-7, p.539–547, 2009.

LANDIS, H. **Production-ready global illumination**. 2002.

LEDDA, P. et al. A local model of eye adaptation for high dynamic range images. In: AFRIGRAPH. **Anais...** [S.l.: s.n.], 2004. p.151–160.

LIN, S. et al. Radiometric Calibration from a Single Image. In: CVPR (2). **Anais...** [S.l.: s.n.], 2004. p.938–945.

LIN, S.; ZHANG, L.; ZHANG, L. Determining the Radiometric Response Function from a Single Grayscale Image. In: CVPR (2). **Anais...** [S.l.: s.n.], 2005. p.66–73.

MANN, S. Compositing multiple pictures of the same scene. In: ANNUAL IMAGING SCIENCE & TECHNOLOGY CONFERENCE, 46. **Proceedings...** [S.l.: s.n.], 1993.

MANTIUK, R. et al. Perception-motivated high dynamic range video encoding. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2004. p.733–741.

MANTIUK, R. et al. Predicting visible differences in high dynamic range images - model and its calibration. In: HUMAN VISION AND ELECTRONIC IMAGING X, IS&T/SPIE'S 17TH ANNUAL SYMPOSIUM ON ELECTRONIC IMAGING. **Anais...** [S.l.: s.n.], 2005. p.204–214.

MANTIUK, R. et al. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2011. p.40.

MASIA, B. et al. Evaluation of reverse tone mapping through varying exposure conditions. **ACM Transactions on Graphics**, [S.l.], v.28, n.5, 2009.

MASIA, B. et al. **Evaluation of reverse tone mapping through varying exposure conditions – Supplemental material**. Disponível em: <http://igl.ethz.ch/projects/rTMO/eval_rTMO_supp.rar>. Acesso em: oct. 2012, 2009.

MASIA, B.; GUTIERREZ, D. **Multilinear Regression for Gamma Expansion of Overexposed Content**. [S.l.]: Universidad de Zaragoza (Tech Report RR-03-11), 2011.

MEYLAN, L.; DALY, S.; SÜSSTRUNK, S. The reproduction of specular highlights on high dynamic range displays. In: IN PROC. OF THE 14TH COLOR IMAGING CONFERENCE. **Anais...** [S.l.: s.n.], 2006.

MITSUNAGA, T.; NAYAR, S. K. Radiometric Self Calibration. In: CVPR. **Anais...** [S.l.: s.n.], 1999. p.1374–1380.

NAYAR, S. K.; BRANZOI, V. Adaptive Dynamic Range Imaging: optical control of pixel exposures over space and time. In: ICCV. **Anais...** [S.l.: s.n.], 2003. p.1168–1175.

OH, B. M. et al. Image-based modeling and photo editing. In: COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 28., New York, NY, USA. **Proceedings...** ACM, 2001. p.433–442. (SIGGRAPH '01).

PANOSCAN. **MK-3 Panoramic Digital Camera Information**. Disponível em: <<http://www.panoscan.com/MK3/>>. Acesso em: oct. 2012.

PARIS, S.; DURAND, F. A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach. In: ECCV (4). **Anais...** [S.l.: s.n.], 2006. p.568–580.

PATTANAIK, S. N. et al. A multiscale model of adaptation and spatial vision for realistic image display. In: SIGGRAPH '98. **Proceedings...** [S.l.: s.n.], 1998. p.287–298.

PÉREZ, P. et al. Poisson image editing. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2003. p.313–318.

PETSCHNIGG, G. et al. Digital photography with flash and no-flash image pairs. In: ACM TRANSACTIONS ON GRAPHICS. **Anais...** [S.l.: s.n.], 2004. p.664–672.

PHAM, T. Q.; VLIET, L. J. van. Separable bilateral filtering for fast video preprocessing. In: ICME. **Anais...** [S.l.: s.n.], 2005. p.454–457.

PORIKLI, F. Constant time O(1) bilateral filtering. In: COMPUTER VISION AND PATTERN RECOGNITION, 2008. CVPR 2008. IEEE CONFERENCE ON. **Anais...** [S.l.: s.n.], 2008. p.1–8.

REINHARD, E. et al. Photographic tone reproduction for digital images. **ACM Transactions on Graphics**, New York, NY, USA, v.21, p.267–276, July 2002.

REINHARD, E. et al. **High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting (2. ed.)**. [S.l.]: Elsevier (Morgan Kaufmann), 2005.

REMPEL, A. G. et al. Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs. **ACM Transactions on Graphics**, [S.l.], v.26, n.3, p.39, 2007.

RITSCHHEL, T. et al. Temporal Glare: real-time dynamic simulation of the scattering in the human eye. **Comput. Graph. Forum**, [S.l.], v.28, n.2, p.183–192, 2009.

ROBERTS, L. G. Picture Coding Using Pseudo-Random Noise. **IRE Transactions on Information Theory**, [S.l.], v.IT-8, n.2, p.145–154, Feb. 1962.

ROBERTSON, M. A. et al. Dynamic Range Improvement Through Multiple Exposures. In: ICIP (3). **Anais...** [S.l.: s.n.], 1999. p.159–163.

ROBERTSON, M. A. et al. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. In: J. ELECTRONIC IMAGING. **Anais...** [S.l.: s.n.], 2003. p.219–228.

SEETZEN, H. et al. High dynamic range display systems. **ACM Transactions on Graphics**, New York, NY, USA, v.23, p.760–768, August 2004.

SEETZEN, H. et al. Guidelines for contrast, brightness, and amplitude resolution of displays. **Society for Information Display (SID) Digest**, [S.l.], p.1229–1233, 2006.

SIMONCELLI, E. P.; FREEMAN, W. T. The Steerable Pyramid: a flexible architecture for multi-scale derivative computation. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. **Anais...** IEEE Signal Processing Society, 1995. p.444–447.

SPHEROCAM. **SPHERON VR – SpheroCam HDR**. Disponível em: <<http://www.spheron.com/en/spheron-cgi/products/spherocam-hdr.html>>. Acesso em: oct. 2012.

TOMASI, C.; MANDUCHI, R. Bilateral Filtering for Gray and Color Images. In: ICCV. **Anais...** [S.l.: s.n.], 1998. p.839–846.

TUMBLIN, J.; RUSHMEIER, H. Tone Reproduction for Realistic Images. **IEEE Comput. Graph. Appl.**, Los Alamitos, CA, USA, v.13, n.6, p.42–48, Nov. 1993.

VAPNIK, V. N. **The nature of statistical learning theory**. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

VOS, J.; BERG, T. van der. **Report on disability glare**. 1999.

WANG, L. et al. High Dynamic Range Image Hallucination. In: RENDERING TECHNIQUES. **Anais...** [S.l.: s.n.], 2007. p.321–326.

WARD, G. et al. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. In: IEEE TRANS. VIS. COMPUT. GRAPH. **Anais...** [S.l.: s.n.], 1997. p.291–306.

WEISS, B. Fast median and bilateral filtering. **ACM Transactions on Graphics**, [S.l.], v.25, n.3, p.519–526, 2006.

DOBBS FERRY, N. (Ed.). **The New Zone System Manual**. [S.l.]: Morgan & Morgan, 1976.

WINNEMÖLLER, H.; OLSEN, S. C.; GOOCH, B. Real-time video abstraction. **ACM Trans. Graph.**, New York, NY, USA, v.25, n.3, p.1221–1226, July 2006.

WYCKOFF, C. W.; FEIGENBAUM, S. A. An Experimental Extended Exposure Response Film. **Society of Photographic Instrumentation Engineers**, [S.l.], v.1, p.117–125, 1962.

YANG, Q.; TAN, K.-H.; AHUJA, N. Real-time $O(1)$ bilateral filtering. In: **CVPR. Anais...** [S.l.: s.n.], 2009. p.557–564.

YOSHIDA, A. et al. Analysis of reproducing real-world appearance on displays of varying dynamic range. **Computer Graphics Forum**, [S.l.], v.25, n.3, p.415–426, 2006.