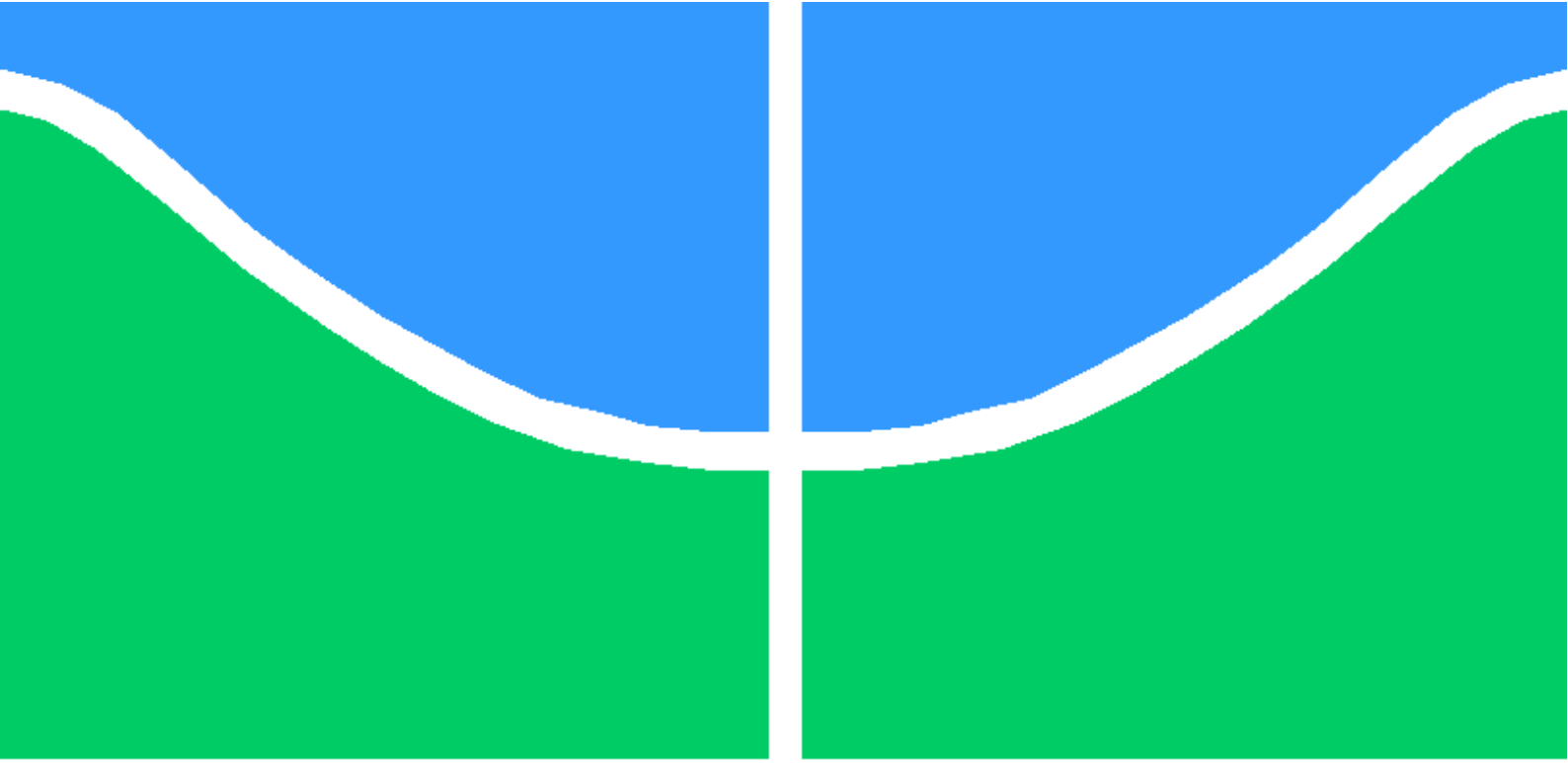


TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO

TÉCNICAS DE SUPER-RESOLUÇÃO PARA SISTEMAS DE VIDEO DE MÚLTIPLAS VISTAS EM RESOLUÇÃO MISTA

Diogo Caetano Garcia

Brasília, agosto de 2012

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASILIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

TÉCNICAS DE SUPER-RESOLUÇÃO PARA SISTEMAS DE VIDEO DE MÚLTIPLAS VISTAS EM RESOLUÇÃO MISTA

Diogo Caetano Garcia

ORIENTADOR: Ricardo Lopes de Queiroz

TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO

Publicação: PPGEA.TD 059/2012
Brasília/DF: Agosto-2012

UNIVERSIDADE DE BRASILIA
Faculdade de Tecnologia

TESE DE DOUTORADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E AUTOMAÇÃO

TÉCNICAS DE SUPER-RESOLUÇÃO PARA SISTEMAS DE VIDEO DE MÚLTIPLAS VISTAS EM RESOLUÇÃO MISTA

Diogo Caetano Garcia

Relatório submetido ao Departamento de Engenharia Elétrica como requisito parcial para obtenção do grau de Doutor em Engenharia Elétrica

Banca Examinadora

Ricardo Lopes de Queiroz, Ph.D. _____
CIC/UnB, Orientador

Francisco Assis de Oliveira Nascimento, Dr. _____
ENE/UnB, Examinador interno

Eduardo Antônio Barros da Silva, Ph.D. _____
COPPE/UFRJ, Examinador externo

Alexandre Zaghetto, Dr. _____
CIC/UnB, Examinador externo

João Luiz Azevedo de Carvalho, Ph.D. _____
ENE/UnB, Examinador interno

FICHA CATALOGRÁFICA

GARCIA, DIOGO CAETANO
Técnicas de Super-Resolução para Sistemas de Video
de Múltiplas Vistas em Resolução Mista. [Distrito Federal] 2012.
xxvii, 134p., 297 mm (ENE/FT/UnB, Doutorado, Telecomunicações
Processamento de Sinais, 2012). Tese de Doutorado.
Universidade de Brasília. Faculdade de Tecnologia.
Departamento de Engenharia Elétrica.

1. Super-resolução baseada em exemplos
2. Múltiplas vistas
3. Resolução mista
4. Estimação de movimento
5. Mapas de profundidade
6. Compressão de video

I. ENE/FT/UnB
II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

GARCIA, D. C. (2012). Técnicas de Super-Resolução para Sistemas de Video de Múltiplas Vistas em Resolução Mista. Tese de Doutorado em Engenharia de Sistemas Eletrônicos e Automação, Publicação PPGEA.TD - 059/2012, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 134p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Diogo Caetano Garcia.

TÍTULO DA TESE DE DOUTORADO: Técnicas de Super-Resolução para Sistemas de Video de Múltiplas Vistas em Resolução Mista.

GRAU / ANO: Doutor / 2012

Diogo Caetano Garcia
SQN 210 Bloco E Apto. 605
70.862-050 Brasília - DF - Brasil.

Dedicatória

À minha esposa e aos meus pais.

Diogo Caetano Garcia

Agradecimentos

Aos meus pais, Sílvio e Edwiges, minha irmã, Luísa, e meus avós, Fabiano, Zélia, Osvaldo e Ieda, por desde sempre me darem a base emocional, material e intelectual que permitiu-me chegar até aqui.

À minha esposa, Janaína, por nos abrir as portas do mundo, e por enriquecer minha vida com seu amor e carinho.

Aos amigos de infância Renê, Rosenberg, Matheus, Juliano, Ybiti, Bellini, Gabriel, Tuca e Ian pela deliciosa formação em paralelo.

Aos amigos Bruno, Camilo, Chaffim, Eduardo, Fernanda, Jorge, Karen, Mintsu, Rafael, Renan, Tiago e Zaghetto pelas risadas, reuniões, festas, "lanchates"e o riquíssimo ambiente de trabalho, em todos os aspectos.

Ao meu orientador, Ricardo Lopes de Queiroz, pela formação intelectual, pelo apoio, e pelo prazer em trabalhar dentro do estado da arte.

Diogo Caetano Garcia

RESUMO

Sequências de múltiplas vistas emergiram recentemente, gerando uma série de aplicações imersivas, tais como como televisões 3D, telas autoestereoscópicas e televisão de ponto-de-vista livre. Em compensação, surgem considerações técnicas, tais como o aumento das taxas de transmissão e da complexidade computacional, em uma escala muito maior do que grande parte dos sistemas de transmissão atuais está preparada para suportar. Uma alternativa viável para muito sistemas é a codificação em resolução mista, amparada por diversos estudos que indicam que a visão binocular não é afetada quando uma das vistas é mais borrada que a outra. O sistema visual humano compensa a falta de detalhes com os detalhes da outra vista, tornando a visão estéreo subjetivamente muito próxima ao resultado obtido quando não se borra uma das vistas. Em compensação, esta arquitetura não é viável para sistemas de ponto-de-vista livre, pois os usuários podem escolher ver a vista borrada em um dado momento. A presente tese propõe três métodos de super-resolução para sequências de múltiplas vistas em resolução mista, nos quais as vistas em resolução normal são utilizadas para recuperar detalhes de alta frequência nas vistas em tamanho reduzido. Diversos testes com sequências reais e sintéticas, realizados com e sem codificação H.264/AVC, mostram ganhos objetivos de qualidade significativos para os métodos propostos, recuperando detalhes de alta frequência para as vistas em tamanho reduzido.

ABSTRACT

Multiview sequences recently emerged, generating a number of immersive applications such as 3D TV, auto-stereoscopic screens ad free-viewpoint TV. On the other hand, several technical considerations emerge, like data-rate and computational-complexity growth, in a much larger scale than many current transmission systems can bear. A viable alternative is mixed resolution coding, based on studies that indicate that binocular vision is not affected when one of the views is blurrier than the other. The human visual system compensates the lack of details with the details in the other view, making stereo vision subjectively very close to the results obtained when one of the views is not blurred. However, this system is not viable for free-viewpoint systems, as users may choose the blurred view at any given time. The present thesis proposes three super-resolution methods for mixed-resolution multiview sequences, using full-resolution views to recover high-frequency details for the low-resolution views. Several tests with real and synthetic sequences, made with and without H.264/AVC coding, show significant objective quality gains for the proposed method, recovering high-frequency details for the low-resolution views.

SUMÁRIO

1 Introdução 1
1.1 Contexto 1
1.2 Apresentação do problema e justificativa 2
1.3 Métodos propostos 2
1.4 Organização da tese 3

2 Representação de video em uma vista 5
2.1 Introdução 5
2.2 Representação de cenas naturais e sintéticas 5
2.3 Mudança de resolução de imagens 8
2.3.1 Decimação 9
2.3.2 Interpolação 11
2.4 Super-resolução de imagens 11
2.4.1 Super-resolução por combinação de múltiplas imagens 12
2.4.2 Super-resolução baseada em exemplos 14
2.5 Padrão H.264/AVC de compressão de video 16
2.5.1 Predição 17
2.5.2 Transformação e escalonamento 20
2.5.3 Codificação e decodificação de entropia 21
2.5.4 Reconstrução e outros processos 21
2.6 Métricas de qualidade de video 21
2.6.1 Métricas subjetivas de qualidade de video 22
2.6.2 Métricas objetivas de qualidade de video 22

3 Representação de video em múltiplas vistas 25
3.1 Introdução 25
3.2 Representação de múltiplas vistas 25
3.2.1 Geometria epipolar 27
3.2.2 Mapas de profundidade 28
3.3 Compressão de múltiplas vistas 30
3.3.1 Compressão de video em múltiplas vistas 31
3.3.2 Compressão de mapas de profundidade 32
3.3.3 Teoria da supressão binocular 32

3.3.4 Codificação de múltiplas vistas em resolução mista 34
3.4 Aplicações de videos em múltiplas vistas 35

4 Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade 39
4.1 Introdução 39
4.2 Arquitetura em consideração 39
4.3 Solução proposta 40
4.3.1 Extração de altas frequências 40
4.3.2 Combinação de altas frequências 43
4.4 Resultados experimentais 43
4.4.1 Testes sem codificação H.264/AVC 46
4.4.2 Testes com codificação H.264/AVC 50

5 Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade 59
5.1 Introdução 59
5.2 Arquitetura em consideração 59
5.3 Solução proposta 60
5.3.1 Extração de altas frequências 60
5.3.2 Combinação de altas frequências 63
5.4 Resultados experimentais 64
5.4.1 Testes sem codificação H.264/AVC 64
5.4.2 Testes com codificação H.264/AVC 70

6 Super-resolução de mapas de profundidade em baixa resolução 79
6.1 Introdução 79
6.2 Arquitetura em consideração 79
6.3 Solução proposta 80
6.3.1 Projeção dos mapas 81
6.3.2 Preenchimento de buracos 82
6.3.3 Filtragem de mapas de profundidade 82
6.4 Resultados experimentais 83
6.4.1 Testes sem codificação H.264/AVC 84
6.4.2 Testes com codificação H.264/AVC 88

7 Conclusões 99
7.1 Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade 99
7.2 Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade 100
7.3 Super-resolução de mapas de profundidade em baixa resolução 100
7.4 Comparações entre arquiteturas 101

7.5 Considerações finais ... 102

REFERÊNCIAS BIBLIOGRÁFICAS ... 103

Anexos ... 109

I Resultados experimentais ... 111

I.1 Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade ... 111

I.1.1 Resultados sem codificação ... 111

I.1.2 Resultados com codificação ... 114

I.2 Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade ... 119

I.2.1 Resultados sem codificação ... 119

I.2.2 Resultados com codificação ... 119

I.3 Super-resolução de mapas de profundidade em baixa resolução ... 119

I.3.1 Resultados sem codificação ... 119

I.3.2 Resultados com codificação ... 130

LISTA DE FIGURAS

2.1 Exemplos de cenas: (a) natural; (b) sintética. 6
2.2 Amostragem espacial e temporal de uma sequência de video..... 6
2.3 Componentes do espaço de cores RGB: (a) imagem original; (b) vermelho; (c) verde; (d) azul. 7
2.4 Componentes do espaço de cores Y:Cr:Cb 4:2:0: (a) imagem final, após ao aumento da resolução de Cr e Cb e da conversão de Y:Cr:Cb para RGB; (b) Y; (c) Cb; (d) Cr. 9
2.5 Conceitos básicos de aumento e redução de imagens: (a) decimação $I^D(u,v)$ de uma imagem $I(u,v)$ de tamanho 8×8 por um fator 4; (b) interpolação $I^I(u,v)$ de uma imagem $I(u,v)$ de tamanho 2×2 por um fator 4; (c) operação geral de decimação: a imagem $I(u,v)$ é convoluída com o filtro $H_D(u,v)$, gerando a versão passa-baixas $I_{PB1}(u,v)$, e depois decimada por um fator M, gerando uma versão de $I(u,v)$ em menor resolução; (d) operação geral de interpolação: a imagem $I(u,v)$ é interpolada por um fator M e depois convoluída com o filtro $H_I(u,v)$, gerando uma versão de $I(u,v)$ em maior resolução. 10
2.6 Super-resolução por combinação de múltiplas imagens. 13
2.7 Super-resolução baseada em exemplos: uma imagem interpolada $\mathbf{I}_0^B$ recebe informações de alta frequência a partir de uma imagem em alta resolução $\mathbf{I}_j$ separada em versões com componentes de baixa e alta frequência, $\mathbf{I}_j^B$ e $\mathbf{I}_j^A$. 15
2.8 Super-resolução baseada em exemplos para uma sequência real: (a) Bloco de tamanho 16×16 em $\mathbf{I}_O^B$; (b) Bloco correspondente ao anterior em $\mathbf{I}_O^A$ (objetivo do algortimo); (c) Janela de busca de tamanho 176×176 em $\mathbf{I}_j^B$; (d) Altas frequências em $\mathbf{I}_j^A$, correspondentes à janela anterior; (e) Bloco de tamanho 16×16 em $\mathbf{I}_j^B$ mais similar a (a); (f) Alta frequência correspondente a (e), utilizada para super-resolver $\mathbf{I}_O^B$. Foi adicionado às Figs (b), (d) e (f) o valor 128 (metade da resolução em 8 *bits*), para que estas pudessem ser apresentadas corretamente, já que $\mathbf{I}_O^A$ e $\mathbf{I}_j^A$ possuem valores negativos. 16
2.9 Extração do bloco atual de baixa frequência em $\mathbf{I}_O^B$ em conjunto com seus vizinhos espaciais de alta frequência em $\hat{\mathbf{I}}_O^A$, a serem utilizados na comparação com os candidatos correspondentes no banco de dados. 17
2.10 Esquema típico de um *codec* H.264/AVC: (a) codificador; (b) decodificador. 18
2.11 Direções de predição Intra para blocos de tamanho 4×4. 19
2.12 Ilustração dos conceitos básicos de estimação de movimento. 20

3.1 Geometria da câmera de modelo *pinhole*. 26

3.2 Relação entre os pontos $\mathbf{m}$ e $\mathbf{M}$ por similaridade de triângulos. 26
3.3 Geometria epipolar: duas câmeras, A e B, capturam uma cena a partir de dois pontos de vista, e o ponto $\mathbf{M}$ é mapeado para os pontos $\mathbf{m}_A$ e $\mathbf{m}_B$ respectivos às câmeras. 27
3.4 Conversão de profundidades $\zeta(u, v)$ entre 1 e 10 metros para representação em 8 *bits*: valores de $\zeta(u, v)$ mais próximos à câmera são amostrados com maior precisão. . 29
3.5 Detalhes de cenas com mapas de profundidade típicos: (a) componente de luminância da sequência sintética *Venus*, vista 6, instante 0; (b) mapa de profundidade correspondente; (c) componente de luminância da sequência real *Ballet*, vista 1, instante 1; (d) mapa de profundidade correspondente. 30
3.6 Conceitos de codificação em múltiplas vistas: (a) Quadros de uma sequência de video em múltiplas vistas; (b) Predição do quadro 1 da vista 1 usando os quadros 0 da vista 1 e 1 da vista 0 como referência. 31
3.7 Detalhes da síntese da vista 6 da sequência *Lovebird1*, quadro 0, utilizando diferentes mapas de profundidade: (a) mapa de profundidade original da vista 4; (b) vista 6 sintetizada com as vistas 4 e 8 originais (PSNR de $25,39$ dB e MSSIM de $80,57\%$); (c) mapa de profundidade da vista 4 decodificado após compressão com o padrão H.264/AVC, modo Intra, $QP = 40$; (d) após compressão das vistas 4 e 8 com o padrão H.264/AVC, modo Intra, $QP = 22$ para cor e $QP = 40$ para os mapas, obtém-se a vista 6 sintetizada (PSNR de 25.40 dB e MSSIM de $80,95\%$). 33
3.8 Ilustração da teoria da supressão binocular com detalhes da sequência *Teddy*, vistas 2 e 6, quadro 0. Cruzando os olhos até uma imagem coincidir com a outra, o leitor pode obter uma impressão tridimensional da cena, percebendo pouca diferença subjetiva entre as cenas tridimensionais geradas nas três linhas, apesar de elas terem sido geradas por diferentes pares estereoscópicos. As Figs. (a), (c) e (e) apresentam a versão original da vista 6 da sequência, e as Figs. (b), (d) e (f) apresentam diferentes versões da vista 2. A Fig. (b) apresenta a versão original da vista 2. A Fig. (d) apresenta a versão da vista 2 após a decimação e a interpolação por 2, utilizando o filtro de Lanczos, com PSNR de $30,19$ dB e MSSIM de $88,98\%$ em relação ao original. A Fig. (f) apresenta a versão da vista 2 após a compressão Intra H.264/AVC, com PSNR de 30.01 dB e MSSIM de $82,95\%$ em relação ao original. ... 37
3.9 Arquitetura de codificação de múltiplas vistas em resolução mista. 38

4.1 Arquitetura de codificação de múltiplas vistas em resolução mista alternada sem mapas de profundidade. 40
4.2 Extração de altas frequências do quadro de referência $\mathbf{I}_R$: (a) com base na correlação com a versão de baixa frequência de $\mathbf{I}_R$; (b) com base na correlação com a versão de alta frequência de $\mathbf{I}_R$. 41
4.3 Quadros de exemplo das sequências reais testadas: (a) *Ballet*; (b) *Breakdancers*; (c) *Pantomime*; (d) *Cafe*; (e) *Lovebird1*; (f) *Poznan Street*; (g) *Newspaper*. 44
4.4 Quadros de exemplo das sequências sintéticas testadas: (a) *Barn1*; (b) *Barn2*; (c) *Bull*; (d) *Map*; (e) *Poster*; (f) *Sawtooth*; (g) *Venus*; (h) *Cones*; (i) *Teddy*; (j) *Room3D*. 45

4.5 Detalhes da sequência *Newspaper*, vista 4, quadro 1, apresentando pouca quantidade de altas frequências: (a) $\mathbf{I}_O$ em resolução 1024×768; (b) $\mathbf{I}_O^B$ em resolução 1024×768 ($40,85$ dB); (c) $\mathbf{I}_O$ em resolução 512×384; (d) $\mathbf{I}_O^B$ em resolução 512×384 ($33,62$ dB). 46

4.6 Resultados sem codificação para a componente de luminância da sequência *Cafe*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 13$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$; (2) $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_{O1}$; (4) $\mathbf{I}_{O2}^{A}{}' + \mathbf{I}_O^B$; (5) $\mathbf{I}_{O2}^{A}{}'' + \mathbf{I}_O^B$; (6) $\hat{\mathbf{I}}_{O2}$; (7) $\mathbf{I}_{O3}^{A}{}' + \mathbf{I}_O^B$; (8) $\mathbf{I}_{O3}^{A}{}'' + \mathbf{I}_O^B$; (9) $\hat{\mathbf{I}}_{O3}$; (10) $\mathbf{I}_{O4}^{A}{}' + \mathbf{I}_O^B$; (11) $\mathbf{I}_{O4}^{A}{}'' + \mathbf{I}_O^B$; (12) $\hat{\mathbf{I}}_{O4}$; (13) $\hat{\mathbf{I}}_O$. 48

4.7 Resultados sem codificação para a componente de luminância da sequência *Sawtooth*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1-3$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$; (2) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_O$. 49

4.8 Detalhes da sequência *Cafe*, vista 3, quadro 2, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR $= 36,53$ dB / MSSIM×100 $= 97,36\%$); (b) $\mathbf{I}_{O3}^{A}{}' + \mathbf{I}_O^B$ ($43,12$ dB / $98,02\%$); (c) $\mathbf{I}_{O3}^{A}{}'' + \mathbf{I}_O^B$ ($42,13$ dB / $97,30\%$); (d) $\hat{\mathbf{I}}_{O3}$ ($43,67$ dB / $98,11\%$); (e) $\hat{\mathbf{I}}_O$ ($46,01$ dB / $98,72\%$); (f) $\mathbf{I}_O$..... 51

4.9 Detalhes da sequência *Teddy*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR $= 29,69$ dB / MSSIM×100 $= 89,19\%$); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ ($30,33$ dB / $90,70\%$); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ ($30,20$ dB / $88,72\%$); (d) $\hat{\mathbf{I}}_O$ ($31,00$ dB / $91,69\%$); (e) $\mathbf{I}_O$. 54

4.10 Detalhes da sequência *Newspaper*, vista 4, quadro 2, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR $= 27,19$ dB / MSSIM×100 $= 81,46\%$); (b) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$ ($37,47$ dB / $97,09\%$); (c) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$ ($29,93$ dB / $88,11\%$); (d) $\hat{\mathbf{I}}_{O1}$ ($37,58$ dB / $97,16\%$); (e) $\hat{\mathbf{I}}_O$ ($40,48$ dB / $98,60\%$); (f) $\mathbf{I}_O$..... 55

4.11 Detalhes da sequência *Cones*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR $= 28,41$ dB / MSSIM×100 $= 84,76\%$); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ ($28,23$ dB / $82,76\%$); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ ($28,67$ dB / $84,27\%$); (d) $\hat{\mathbf{I}}_O$ ($28,83$ dB / $84,64\%$); (e) $\mathbf{I}_O$. 56

4.12 Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Poznan Street*, vista 4: (a) taxa e PSNR, $M = 2$ (ganho médio de $1,61$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $2,11\%$); (c) taxa e PSNR, $M = 4$ (ganho médio de $2,16$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $5,66\%$)..... 57

4.13 Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Cones*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$ (ganho médio de $0,32$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $0,64\%$); (c) taxa e PSNR, $M = 4$ (perda média de $0,19$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $1,15\%$)..... 58

5.1 Arquitetura de codificação de múltiplas vistas em resolução mista alternada com mapas de profundidade. 60
5.2 Extração de altas frequências para sequências em múltiplas vistas. Uma estimativa de alta frequência $\mathbf{I}_O^{A\prime}$ para a vista O é gerada a partir de uma vista adjacente em alta resolução $\mathbf{I}_R$, e dos mapas de profundidade correspondentes, $\mathbf{D}_O$ e $\mathbf{D}_R$. 61
5.3 Teste de consistência entre os mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$: (a) projeção do ponto $\mathbf{m}_O$ para o ponto $\mathbf{m}_R$, utilizando as Eqs. (3.4) e (3.5); (b) posição em ponto fixo $\mathbf{m}_R$ no mapa $\mathbf{D}_R$, e posições inteiras mais próximas $\mathbf{m}_{R1}$, $\mathbf{m}_{R2}$, $\mathbf{m}_{R3}$ e $\mathbf{m}_{R4}$; (c) re-projeção de $\mathbf{m}_{R2}$ de $\mathbf{D}_R$ para $\mathbf{D}_O$, caindo dentro de um raio de 1 *pixel* e resultando em uma projeção válida; (d) re-projeção de $\mathbf{m}_{R2}$ de $\mathbf{D}_R$ para $\mathbf{D}_O$, caindo fora de um raio de 1 *pixel* e resultando em uma projeção inválida. 62
5.4 Erosão do mapa $\mathbf{I}'_{VAL}$ de consistência entre mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$. Considerando *pixels* brancos como $I'_{VAL}(u,v) = 1$ e *pixels* pretos como $I'_{VAL}(u,v) = 0$, para cada *pixel* $I'_{VAL}(u,v)$, avalia-se os *pixels* $I'_{VAL}(u-1,v)$, $I'_{VAL}(u,v-1)$, $I'_{VAL}(u,v)$, $I'_{VAL}(u+1,v)$ e $I'_{VAL}(u,v+1)$. Se todos estes forem iguais a 1, $I''_{VAL}(u,v) = 1$; caso contrário, $I''_{VAL}(u,v) = 0$. 63
5.5 Quadros de exemplo dos mapas de profundidade das sequências reais testadas: (a) *Ballet*; (b) *Breakdancers*; (c) *Pantomime*; (d) *Cafe*; (e) *Lovebird1*; (f) *Poznan Street*; (g) *Newspaper*. 66
5.6 Quadros de exemplo dos mapas de profundidade das sequências sintéticas testadas: (a) *Barn1*; (b) *Barn2*; (c) *Bull*; (d) *Map*; (e) *Poster*; (f) *Sawtooth*; (g) *Venus*; (h) *Cones*; (i) *Teddy*; (j) *Room3D*. 67
5.7 Resultados sem codificação para a componente de luminância da sequência *Pantomime*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1-13$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$; (2) $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_{O1}$; (4) $\mathbf{I}_{O2}^{A}{}' + \mathbf{I}_O^B$; (5) $\mathbf{I}_{O2}^{A}{}'' + \mathbf{I}_O^B$; (6) $\hat{\mathbf{I}}_{O2}$; (7) $\hat{\mathbf{I}}_O$. 68
5.8 Resultados sem codificação para a componente de luminância da sequência *Venus*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1-3$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$; (2) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_O$. 69
5.9 Detalhes da sequência *Ballet*, vista 1, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 33,65 dB / MSSIM×100 = 94,40%); (b) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$ (36,35 dB / 95,88%); (c) $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_O^B$ (36,23 dB / 95,96%); (d) $\hat{\mathbf{I}}_{O1}$ (36,42 dB / 96,14%); (e) $\hat{\mathbf{I}}_O$ (37,57 dB / 96,92%); (f) $\mathbf{I}_O$. 71
5.10 Detalhes da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 28,47 dB / MSSIM×100 = 86,24%); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ (35,81 dB / 97,40%); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ (35,50 dB / 97,40%); (d) $\hat{\mathbf{I}}_O$ (35,74 dB / 97,44%); (e) $\mathbf{I}_O$. 72

5.11 Detalhes dos *pixels* advindos de valores de profundidade válidos: (a) $\mathbf{I}'_{VAL}$ e $\mathbf{I}'_{O|R}$ sobrepostos para a sequência *Venus*, vista 6, quadro 0; (b) $\mathbf{I}''_{VAL}$ e $\mathbf{I}''_{O|R}$ sobrepostos para a mesma sequência; (c) $\mathbf{I}'_{VAL}$ e $\mathbf{I}'_{O|R}$ sobrepostos para a sequência *Ballet*, vista 1, quadro 1; (d) $\mathbf{I}''_{VAL}$ e $\mathbf{I}''_{O|R}$ sobrepostos para a mesma sequência. 73

5.12 Detalhes da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{I}^B_O$ (PSNR = 28,53 dB / MSSIM×100 = 93,01%); (b) $\mathbf{I}^A_{O1}{}' + \mathbf{I}^B_O$ (35,77 dB / 96,87%); (c) $\mathbf{I}^A_{O1}{}'' + \mathbf{I}^B_O$ (33,95 dB / 96,37%); (d) $\hat{\mathbf{I}}_{O1}$ (35,36 dB / 96,76%); (e) $\hat{\mathbf{I}}_O$ (36,05 dB / 97,09%); (f) $\mathbf{I}_O$ 74

5.13 Detalhes da sequência *Poster*, vista 6, quadro 0, $M = 4$: (a) $\mathbf{I}^B_O$ (PSNR = 22,65 dB / MSSIM×100 = 57,59%); (b) $\hat{\mathbf{I}}'_O$ (31,93 dB / 96,34%); (c) $\hat{\mathbf{I}}''_O$ (31,53 dB / 96,10%); (d) $\hat{\mathbf{I}}_O$ (31,82 dB / 96,28%); (e) $\mathbf{I}_O$ 76

5.14 Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Pantomime*, vista 39, quadro 1: (a) taxa e PSNR, $M = 2$ (ganho médio de 3,37 dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de 0,98%); (c) taxa e PSNR, $M = 4$ (ganho médio de 6,75 dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de 6,35%). 77

5.15 Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Poster*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$ (ganho médio de 4,30 dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de 14,89%); (c) taxa e PSNR, $M = 4$ (ganho médio de 5,79 dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de 34,12%). 78

6.1 Arquitetura de codificação de múltiplas vistas com mapas de profundidade em baixa resolução. 80

6.2 Super-resolução de um mapa de profundidade $\mathbf{D}^D_O$ em baixa resolução utilizando três mapas $\mathbf{D}^D_{R1}$, $\mathbf{D}^D_{R2}$ e $\mathbf{D}^D_{R3}$, também em baixa resolução. 81

6.3 Nuvem de três pontos $\{\mathbf{m}''_{R1}, \mathbf{m}''_{R2}, \mathbf{m}''_{R3}\}$ projetados de posições $\{\mathbf{m}_{R1}, \mathbf{m}_{R2}, \mathbf{m}_{R3}\}$ em vistas adjacentes. $\{\mathbf{m}_{O1}, \mathbf{m}_{O2}, \mathbf{m}_{O3}, \mathbf{m}_{O4}\}$ representam posições inteiras próximas na vista O. 82

6.4 Resultados sem codificação para a componente de luminância da sequência *Poznan Street*. São apresentados os ganhos de $\hat{\mathbf{I}}_O$ em relação a $\mathbf{I}^B_O$ utilizando os mapas de profundidade super-resolvidos com o método proposto neste Capítulo. Os ganhos baseiam-se na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros de $\hat{\mathbf{I}}_O$ e $\mathbf{I}^B_O$: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 5$ indicam os ganhos utilizando os seguintes mapas de profundidade: (1) $\mathbf{D}_r$; (2) $\mathbf{D}^B_r$; (3) $\mathbf{D}^{B,MED}_r$; (4) $\hat{\mathbf{D}}_{r1}$; (5) $\hat{\mathbf{D}}_r$. 84

6.5 Resultados sem codificação para a componente de luminância da sequência *Barn1*. São apresentados os ganhos de $\hat{\mathbf{I}}_O$ em relação a $\mathbf{I}_O^B$ utilizando os mapas de profundidade super-resolvidos com o método proposto neste Capítulo. Os ganhos baseiam-se na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros de $\hat{\mathbf{I}}_O$ e $\mathbf{I}_O^B$: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 5$ indicam os ganhos utilizando os seguintes mapas de profundidade: (1) $\mathbf{D}_r$; (2) $\mathbf{D}_r^B$; (3) $\mathbf{D}_r^{B,MED}$; (4) $\hat{\mathbf{D}}_{r1}$; (5) $\hat{\mathbf{D}}_r$. 85

6.6 Detalhes da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR = 28,53 dB / MSSIM×100 = 93,01%); (b) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^B$ (37,64 dB / 97,33%); (c) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^{B,MED}$ (37,56 dB / 97,31%); (d) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_{O1}$ (37,74 dB / 97,35%); (e) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_O$ (37,85 dB / 97,38%); (f) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O$ (36,05 dB / 97,09%). 87

6.7 Detalhes dos mapas de profundidade da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{D}_O^B$ (PSNR = 46,85 dB / MSSIM×100 = 98,65%); (b) $\mathbf{D}_O^{B,MED}$ (45,87 dB / 98,59%); (c) $\hat{\mathbf{D}}_{O1}$ (45,62 dB / 98,52%); (d) $\hat{\mathbf{D}}_O$ (45,45 dB / 98,47%); (e) $\mathbf{D}_O$. 88

6.8 Detalhes da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 28,47 dB / MSSIM×100 = 86,24%); (b) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^B$ (35,13 dB / 97,21%); (c) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^{B,MED}$ (35,17 dB / 97,23%); (d) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_O$ com todos os mapas de referência (35,10 dB / 97,22%); (e) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O$ (35,74 dB / 97,44%); (f) $\mathbf{I}_O$..... 89

6.9 Detalhes dos mapas de profundidade da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{D}_O^B$ (PSNR = 45,98 dB / MSSIM×100 = 99,38%); (b) $\mathbf{D}_O^{B,MED}$ (35,35 dB / 99,39%); (c) $\hat{\mathbf{D}}_O$ com todos os mapas de referência (35,30 dB / 99,34%); (d) $\mathbf{D}_O$.... 90

6.10 Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Pantomime*, vista 39, quadro 1: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$. 94

6.11 Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Cafe*, vista 3: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$. 95

6.12 Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Venus*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$. 96

6.13 Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Bull*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$..... 97

LISTA DE TABELAS

4.1 Sequências utilizadas..... 47

5.1 Sequências utilizadas..... 65

6.1 Tempo médio de codificação e de super-resolução para os testes de super-resolução com os mapas de profundidade processados, em segundos, para $M = 2$..... 92

6.2 Tempo médio de codificação e de super-resolução para os testes de super-resolução com os mapas de profundidade processados, em segundos, para $M = 4$..... 93

I.1 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 112

I.2 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 113

I.3 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 115

I.4 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências sintética, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 116

I.5 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 117

I.6 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 118

I.7 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 120

I.8 Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 121

I.9 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 122

I.10 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 123

I.11 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 124

I.12 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 125

I.13 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 126

I.14 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 127

I.15 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 128

I.16 Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 129

I.17 Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), sem codificação, para a componente de luminância de todas as sequências, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 131

I.18 Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), sem codificação, para a componente de luminância de todas as sequências, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 132

I.19 Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), com codificação, para a componente de luminância de todas as sequências, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 133
I.20 Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), com codificação, para a componente de luminância de todas as sequências, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. 134

LISTA DE SÍMBOLOS

Siglas

3D	Tridimensional
AVC	*Advanced Video Coding*
CABAC	*Context-Adaptive Binary Arithmetic Coding*
DCT	*Discrete Cosine Transform*
DSCQS	*Double Stimulus Continuous Quality Scale*
ISO/IEC	*International Organisation for Standardisation / International Electrotechnical Commission*
ITU-R	*International Telecommunication Union - Radiocommunications Sector*
MB	Macrobloco
MOS	*Mean Opinion Score*
MRC	*Mixed-Resolution Coding*
MSE	*Mean Square Error*
MSSIM	*Mean Structural Similarity Index*
MVC	*Multiview Video Coding*
PSNR	*Peak Signal-to-Noise Ratio*
QP	*Quantization Parameter*
RGB	*Red*, *Green* and *Blue*
SSD	*Sum of Squared Differences*
SSIM	*Structural Similarity Index*
VLC	*Variable Length Coding*
Y:Cb:Cr	Luminância e crominâncias

Capítulo 1

Introdução

1.1 Contexto

A representação digital de cenas em múltiplas vistas tem apresentado grande desenvolvimento na última década, graças a diversos avanços nas tecnologias de aquisição e apresentação de video 3D, e ao surgimento de processadores suficientemente velozes para lidar com a complexidade computacional requerida [1]. Tudo isso aumentou o interesse da indústria pela área, contribuindo ainda mais para desenvolvimento de novas técnicas e produtos, tais como televisões 3D, telas autoestereoscópicas, televisão de ponto-de-vista livre e teleconferências imersivas.

A pesquisa em sequências de múltiplas vistas ainda possui uma série de questões não resolvidas. Por exemplo, o cálculo completo de correspondências entre câmeras, ou estimação de disparidade, ainda é um tópico de estudo bastante ativo, sendo crucial em outras áreas de pesquisa relacionadas, tais como a síntese de vistas e o reconhecimento de objetos e de gestos humanos [2]. Outro problema importante é a taxa de dados necessária para representar sequências em múltiplas vistas, sendo ainda proibitivamente alta para diversas aplicações, tais como a transmissão em tempo real via *internet* ou redes móveis [1].

De forma geral, o acréscimo de câmeras para representar uma cena acarreta em um aumento de dados transmitidos. Utilizando técnicas de compressão de video para única vista, este aumento é aproximadamente proporcional. Técnicas mais avançadas de compressão reduzem esta taxa de dados em 20% na média, de forma que a codificação de 10 câmeras simultâneas, por exemplo, requer uma taxa aproximadamente 8× maior do que a taxa resultante para uma única câmera, para a qual grande parte das redes e sistemas atuais foi projetada [1] [3].

O acréscimo de câmeras também introduz um aumento na complexidade de codificação de video. As técnicas mais avançadas de compressão de múltiplas vistas utilizam imagens de vistas adjacentes para melhorar a predição, que é o processo mais oneroso da codificação em termos de complexidade computacional. Métodos mais avançados de codificação, tais como a interpolação e a extrapolação de vistas para predição, aumentam significativamente a complexidade computacional [1] [4].

1.2 Apresentação do problema e justificativa

Tendo em vista as considerações anteriores, foram desenvolvidas diversas técnicas para reduzir a taxa de transmissão e/ou a complexidade de compressão de sequências de video em múltiplas vistas, preferivelmente mantendo a qualidade subjetiva e/ou objetiva das mesmas. Uma área amplamente pesquisada é a codificação de múltiplas vistas em resolução mista [5] [6] [7] [8] [9] [10].

Esta técnica baseia-se em uma característica intrínseca ao sistema visual humano, descrita pela teoria da supressão binocular [11]. Basicamente, a visão binocular é pouco afetada quando uma das vistas é mais borrada em relação a outra, pois o sistema visual dos seres humanos utiliza a vista em melhor qualidade para compensar a qualidade final da visão estéreo. Diversos estudos indicam que nem todo tipo de degradação é compensado pelo sistema visual humano, sendo uma das alternativas viáveis a redução de resolução de uma das vistas [12] [13] [14].

Uma imagem reduzida raramente é completamente recuperada quando aumentada de volta às suas dimensões originais, tornando-se borrada em relação à imagem original. Na visão estéreo, este tipo de degradação é compensado pelo sistema visual humano. A imagem reduzida, por sua vez, acarreta em uma redução na taxa de dados a serem transmitidos e na complexidade da sua codificação, tornando o formato de múltiplas vistas em resolução mista viável a uma série de aplicações, tais como video 3D para celulares e câmeras [5] [6]. Em compensação, este formato não é viável para aplicações de ponto-de-vista livre [15], em que o usuário escolhe de qual posição ele deseja ver a cena, pois no momento em que ele escolher uma vista que foi reduzida, perceberá a queda de qualidade na imagem visualizada. Torna-se necessário obter uma forma de recuperar os detalhes das imagens reduzidas.

1.3 Métodos propostos

Neste trabalho, apresenta-se três métodos para recuperar detalhes de imagens reduzidas em sistemas de múltiplas vistas em resolução mista, baseado nas informações das vistas transmitidas em resolução normal. Estas vistas são em geral altamente correlacionadas com as vistas em resolução reduzida, possuindo informações correspondentes de alta frequência. Os métodos propostos não se excluem, mas atendem a diferentes arquiteturas, de acordo com as necessidades de cada sistema.

No primeiro método, desenvolve-se uma técnica de super-resolução de múltiplas vistas em resolução mista sem utilizar mapas de profundidade [16]. A partir de uma ou mais imagens de referência, extrai-se múltiplas informações de alta frequência através de métodos de estimação e compensação de movimento e/ou disparidade, e em seguida combina-se estas diversas informações em uma única imagem de alta frequência.

No segundo método, desenvolve-se uma técnica de super-resolução de múltiplas vistas em resolução mista utilizando mapas de profundidade [17] [18]. Estes são analisados e utilizados para projetar informações de alta frequência para a imagem reduzida, a partir de uma ou mais

imagens de referência. Utiliza-se também um método similar de combinação de informações de alta frequência para gerar uma imagem final super-resolvida.

No terceiro método, desenvolve-se uma técnica de super-resolução para mapas de profundidade em baixa resolução, que são posteriormente utilizados pelo segundo método para super-resolver múltiplas vistas em resolução mista. Deste forma, os mapas em baixa resolução contribuem para a redução de taxa e de complexidade na codificação de múltiplas vistas em resolução mista, e o método proposto melhora a qualidade das vistas em baixa resolução após a decodificação.

Foram realizados testes com uma série de sequências de video em múltiplas vistas, a fim de verificar a qualidade de cada um dos métodos propostos.

1.4 Organização da tese

A tese desenvolve-se em sete capítulos, sendo o primeiro deles a presente introdução. Os Capítulos 2 e 3 constituem a revisão bibliográfica, discutindo os principais termos empregados ao longo desta tese. O Capítulo 2 trata dos seguintes conceitos de vídeo em uma única vista: representação, mudança de resolução, compressão e métricas de qualidade. Em seguida, o Capítulo 3 detalha os seguintes conceitos de vídeo em múltiplas vistas: relações geométricas entre as diversas câmeras, representação de informação tridimensional através de mapas de profundidade, compressão e aplicações práticas de video em múltiplas vistas. Os Capítulos 4, 5 e 6 apresentam os métodos propostos pela presente tese, e o Capítulo 7 discute as principais conclusões e as sugestões para trabalhos futuros.

Capítulo 2

Representação de video em uma vista

2.1 Introdução

Neste Capítulo, são apresentados os conceitos mais importantes relativos à representação de sequências de video em formato digital, tais como amostragem espacial e temporal, representação matricial e espaços de cores. Em seguida, discute-se o procedimento básico de mudança de resolução de imagens, que será fundamental nos Capítulos por vir, e o conceito de super-resolução de imagens, que procura acrescentar detalhes de alta frequência a imagens redimensionadas. A compressão de video é abordada a seguir, com ênfase no padrão H.264/AVC. Por fim, apresentam-se métricas de qualidade de video com as quais este trabalho quantificará os resultados obtidos mais adiante.

2.2 Representação de cenas naturais e sintéticas

Uma cena natural capturada em video é geralmente composta por diversos objetos de diferentes formatos, texturas, profundidades, cores, brilhos e iluminações, dentre outras características, como mostra a Fig. 2.1(a) [19]. Uma cena sintética, por sua vez, procura emular um cena real ou apresentar uma cena virtual, e é geralmente desenvolvida com métodos de computação gráfica. Um exemplo de cena sintética pode ser visto na Fig. 2.1(b) [2]. Tanto em cenas naturais como em sintéticas, é comum observar mudanças espaciais e temporais no posicionamento de objetos, bem como alterações em texturas, na iluminação da cena e na localização da câmera [20].

Para se representar cenas naturais e sintéticas de forma digital, é necessário amostrá-las no espaço e no tempo, como indica a Fig. 2.2. Para isso, são obtidas fotografias digitais das cenas (quadros) entre intervalos regulares, o que envolve a amostragem regular de cada instante em um plano de imagem, geralmente retangular. O video digital consiste em uma série de amostras espaço-temporais, denominadas *pixels* (do inglês *picture element*, ou elemento de imagem).

Uma sequência de video é armazenada na forma de um conjunto de matrizes, uma para cada quadro. Em linguagem matricial, definimos $\mathbf{I}_n$ como o n-ésimo quadro de uma sequência de video, e $I_n(u, v)$ como o *pixel* na u-ésima coluna e v-ésima linha deste quadro.

(a) (b)

Figura 2.1: Exemplos de cenas: (a) natural; (b) sintética.

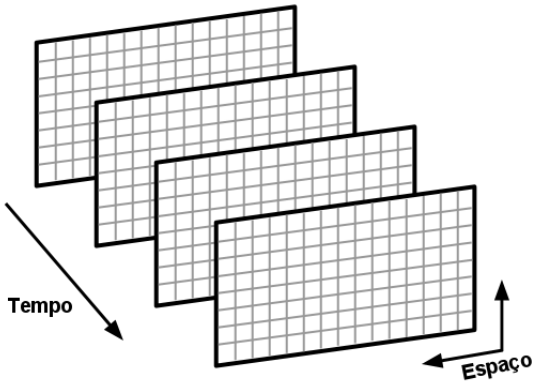


Figura 2.2: Amostragem espacial e temporal de uma sequência de video.

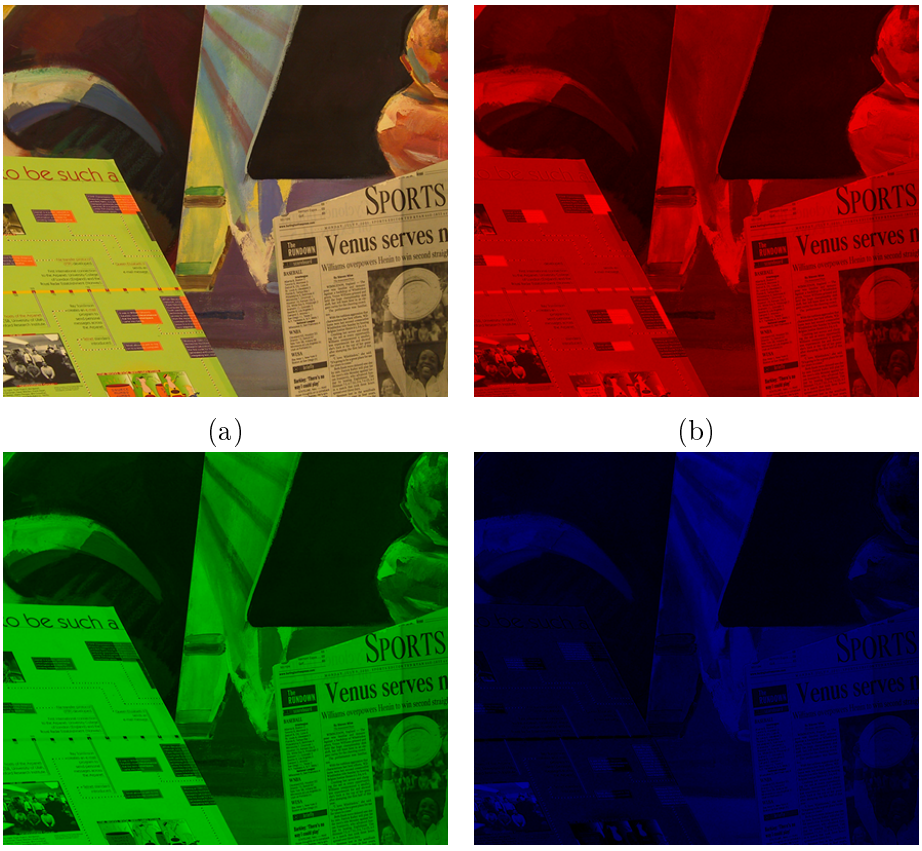
O número de amostras em cada quadro (resolução) influencia a qualidade visual do mesmo de forma proporcional. O número de quadros por segundo (taxa de quadros, ou *frame rate*) influencia a percepção do movimento: acima de 25 quadros por segundo, o movimento é aparentemente natural, e abaixo dessa taxa, geralmente percebe-se mudanças mais abruptas nas posições dos objetos. A resolução e a taxa de quadros afetam proporcionalmente a quantidade inicial de dados necessários para a representação de um video digital.

Cada *pixel* em um quadro descreve o brilho ou a cor de uma amostra. Uma imagem monocromática requer somente um valor por *pixel*, que representa o brilho (ou luminância) da amostra, enquanto imagens coloridas requerem pelo menos três valores por *pixel*. Cada um desses valores é representado com um número finito de *bits*, sendo bastante comum a representação por 8 *bits*. O espaço de cores é o método utilizado para representar brilho ou cor em cada amostra de uma imagem.

No espaço de cores RGB, um *pixel* é representado por três valores, relativos à proporção de cores vermelha, verde e azul em uma amostra (do inglês, *Red, Green and Blue*). Estas são consideradas

cores primárias aditivas, pois uma grande gama de cores pode ser criada a partir da combinação das três. A captura de uma imagem neste espaço de cores requer a filtragem da cena para cada cor separadamente. Já a apresentação desta cena requer telas coloridas que apresentem as proporções de cada cor, que naturalmente se misturam para o observador a uma distância normal da tela.

Na prática, cada componente no espaço de cores é representada por uma matriz diferente. Sendo assim, uma imagem colorida é representada no espaço de cores RGB por três matrizes, como mostra a Fig. 2.3.



(a) (b)

(c) (d)

Figura 2.3: Componentes do espaço de cores RGB: (a) imagem original; (b) vermelho; (c) verde; (d) azul.

O sistema visual humano apresenta resolução maior para luminância do que para cores. Assim, se a luminância puder ser extraída das informações de cor, as demais componentes de cor podem ser representadas com menor resolução do que a componente de luminância, diminuindo a quantidade de dados necessários na representação da imagem sem evidente prejuízo subjetivo na qualidade da mesma.

O espaço de cores Y:Cr:Cb procura representar imagens coloridas desta maneira. Para um dado *pixel*, Y é o chamado componente de luminância, que é calculado a partir de uma média ponderada dos componentes R, G e B de cada *pixel*:

$$Y = k_r R + k_g G + k_b B, \tag{2.1}$$

onde k_r, k_g e k_b são pesos constantes.

As informações de cor são dadas pela diferença entre as cores R, G e B e a luminância Y, e são denominadas crominâncias:

$$\begin{aligned} Cr &= R - Y \\ Cg &= G - Y. \\ Cb &= B - Y \end{aligned} \tag{2.2}$$

A princípio, o espaço de cores Y:Cr:Cb necessita de quatro componentes para representar um amostra, Y, Cr, Cg e Cb. Contudo, é possível obter as componentes R,G e B a partir de Y, Cr e Cb, motivo pelo qual o espaço de cores chama-se Y:Cr:Cb.

De acordo com a recomendação BT.601 da ITU-R (*International Telecommunication Union - Radiocommunications sector*), a conversão do espaço RGB para Y:Cr:Cb deve ser feita da seguinte forma:

$$\begin{aligned} Y &= 0,299R + 0,587G + 0,114B \\ Cb &= 0,564(B - Y) + 128 \\ Cr &= 0,713(R - Y) + 128 \end{aligned} \quad . \tag{2.3}$$

Desta maneira, caso as componentes R, G e B estiverem contidas em valores entre 0 e 255 (ou seja, 8 *bits* de precisão), as componentes Y, Cb e Cr também terão valores dentro dessa faixa, após o arredondamento dos resultados obtidos pela Eq. 2.3 [21].

A vantagem deste espaço de cores é que o sistema visual humano é mais sensível à luminância do que às cores, permitindo que as crominâncias Cr e Cb sejam representadas em menor resolução, sem prejuízo visual aparente, como mostra a Fig. 2.4. Assim, reduz-se a quantidade de dados necessários para representar cada quadro, e o video como um todo, por conseguinte. Quando as componentes Y, Cr e Cb são representadas com resoluções idênticas, o formato de amostragem é chamado 4:4:4. Quando as componentes Cr e Cb possuem metade das resoluções vertical e horizontal de Y, o formato de amostragem é chamado 4:2:0.

2.3 Mudança de resolução de imagens

Operações de aumento e redução na resolução de imagens são comuns em uma série de aplicações, tais como a adaptação de uma mesma sequência de video para diferentes receptores (televisão digital, computador e telefone celular, por exemplo), e a conversão do espaço de cores RGB para o espaço Y:Cb:Cr 4:2:0, como na Fig. 2.4. A redução e o aumento na resolução de imagens são obtidos através dos processos de decimação e interpolação, respectivamente [22].

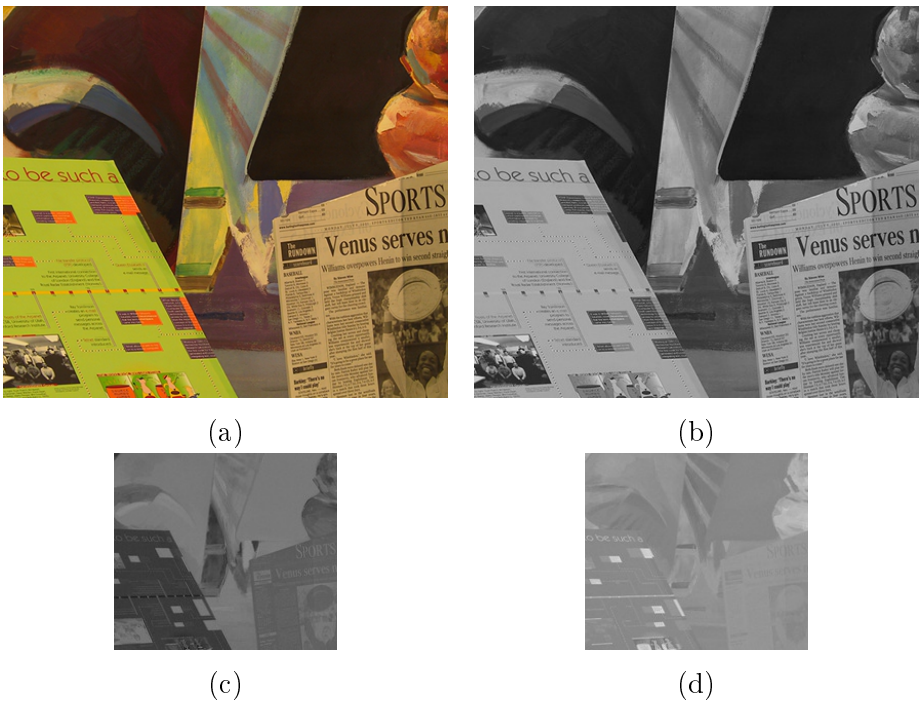


Figura 2.4: Componentes do espaço de cores Y:Cr:Cb 4:2:0: (a) imagem final, após ao aumento da resolução de Cr e Cb e da conversão de Y:Cr:Cb para RGB; (b) Y; (c) Cb; (d) Cr.

2.3.1 Decimação

O processo de decimação de uma imagem por um fator M consiste em manter uma a cada M colunas e uma a cada M linhas da mesma. Definindo u e v como sendo as coordenadas de largura e comprimento de uma dada imagem, cada *pixel* da imagem original é indicado por $I(u, v)$, e cada *pixel* da imagem decimada por M é indicado por $I^D(u, v)$. Assim, temos:

$$I^D(u, v) = I(uM, vM). \tag{2.4}$$

A Fig. 2.5(a) ilustra este conceitos para uma imagem de tamanho 8×8, e um fator $M = 4$.

Aplicando a transformada de Fourier sobre a Eq. (2.4), prova-se que $I^D(u, v)$ sofrerá do problema de superposição espectral (ou *aliasing*, em inglês) se a largura de faixa da transformada discreta de Fourier de $I(u, v)$ estiver fora do intervalo de frequências $\left[-\frac{\pi}{M}, \frac{\pi}{M}\right]$. Para evitar que isso aconteça, o processo de decimação é geralmente precedido por uma filtragem que preserve o intervalo $\left[-\frac{\pi}{M}, \frac{\pi}{M}\right]$, e anule todo o resto (um filtro passa-baixas, no caso). A operação geral de decimação por um fator M, incluindo o filtro passa-baixas $\mathbf{H}_D$, é ilustrada na Fig. 2.5(c).

Um filtro $\mathbf{H}_D$ bastante popular para a decimação é o filtro de Lanczos, que é definido em duas dimensões da seguinte forma:

$$H_D(u, v) = L(u)L(v), \tag{2.5}$$

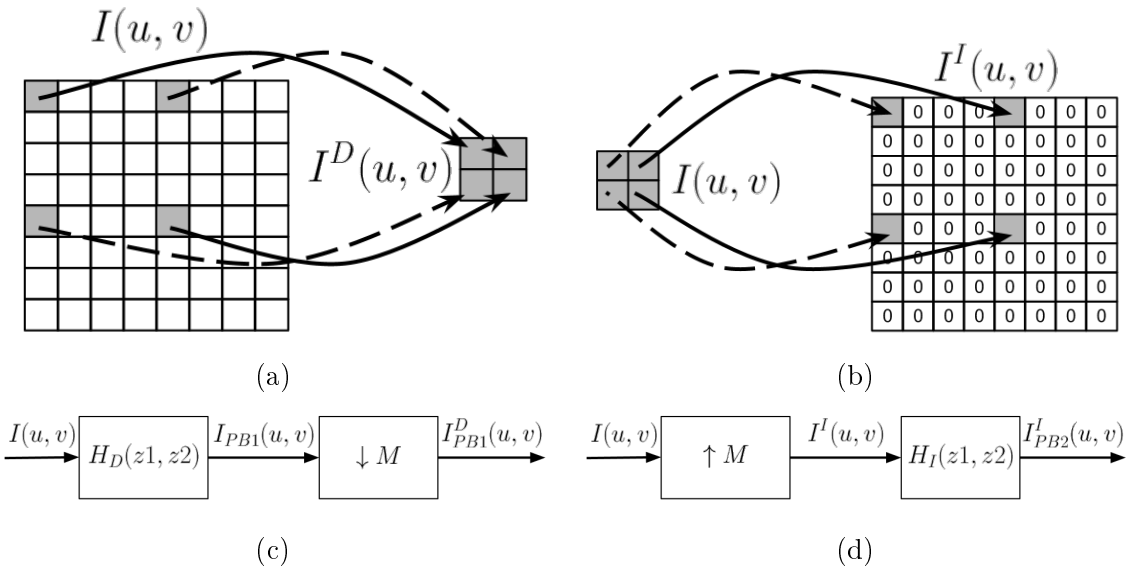


Figura 2.5: Conceitos básicos de aumento e redução de imagens: (a) decimação $I^D(u,v)$ de uma imagem $I(u,v)$ de tamanho 8×8 por um fator 4; (b) interpolação $I^I(u,v)$ de uma imagem $I(u,v)$ de tamanho 2×2 por um fator 4; (c) operação geral de decimação: a imagem $I(u,v)$ é convoluída com o filtro $H_D(u,v)$, gerando a versão passa-baixas $I_{PB1}(u,v)$, e depois decimada por um fator M, gerando uma versão de $I(u,v)$ em menor resolução; (d) operação geral de interpolação: a imagem $I(u,v)$ é interpolada por um fator M e depois convoluída com o filtro $H_I(u,v)$, gerando uma versão de $I(u,v)$ em maior resolução.

onde a função $L()$ é definida como:

$$L(\alpha) = \begin{cases} sinc(\alpha)sinc(\alpha/a) & -a < \alpha < a \\ 0 & \text{caso contrário} \end{cases} . \qquad (2.6)$$

$$sinc(\alpha) = \frac{\sin(\pi\alpha)}{\pi\alpha}$$

O fator a é um valor inteiro positivo, geralmente igual a 2 ou 3, que controla o tamanho de $L()$. Fazendo a amostragem desta função com $a = 3$, escolhendo 12 *taps* e normalizando para oferecer ganho unitário na faixa de passagem, temos a seguinte realização de $L(\alpha)$ para $M = 2$:

$$\begin{aligned} L(\alpha) = [&0,0036891,\ 0,0150561,\ -0,0339986,\ -0,0666373,\ 0,1355053,\ 0,4463854,\\ &0,4463854,\ 0,1355053,\ -0,0666373,\ -0,0339986,\ 0,0150561,\ 0,0036891]. \end{aligned} \qquad (2.7)$$

Para realizar $M = 4$, basta aplicar o processo de decimação com este filtro duas vezes; para $M = 8$, três vezes, e assim por diante.

2.3.2 Interpolação

O processo de interpolação de uma imagem por um fator M consiste em acrescentar $M-1$ zeros entre cada linha e $M-1$ zeros entre cada coluna da imagem, como mostra a Fig. 2.5(b). Definindo a imagem interpolada por M como sendo $I^I(u,v)$, então temos:

$$I^I(u,v) = \begin{cases} I(u/M, v/M), & u = jM \text{ e } v = kM,\ j,k \in \mathbb{Z} \\ 0, & \text{caso contrário} \end{cases} . \tag{2.8}$$

Aplicando a transformada de Fourier sobre a Eq. (2.8), prova-se que $I^I(u,v)$ deverá ser posteriormente filtrada por um filtro passa-baixas para evitar o problema de *aliasing*. O filtro, denominado $H_I(u,v)$, deverá ter ganho igual a M no intervalo de frequências $\left[-\frac{\pi}{M}, \frac{\pi}{M}\right]$, e ganho nulo em todas as outras frequências. A operação geral de interpolação por um fator M, incluindo o filtro passa-baixas $\mathbf{H}_I$, é ilustrado na Fig. 2.5(d).

O filtro Lanczos definido nas Eqs. (2.5) e (2.6) pode ser utilizado para realizar o filtro $\mathbf{H}_I$. Fazendo a amostragem da função $L(\alpha)$ com $a = 3$, escolhendo 12 *taps* e normalizando para oferecer ganho igual a M na faixa de passagem, temos o mesmo filtro da Eq. 2.7 multiplicado por M.

2.4 Super-resolução de imagens

A interpolação de uma imagem geralmente cria uma versão borrada da imagem, e em maior resolução. Isto é inerente ao próprio processo de interpolação, já que as amostras desconhecidas são inicialmente feitas iguais a zero, de acordo com a Eq. (2.8), e a imagem é depois filtrada por um filtro passa-baixas, como indicado na Fig. 2.5(d). Para se obter dados de alta frequência para a imagem interpolada, são necessárias mais informações da cena.

Técnicas de super-resolução procuram acrescentar estas informações de alta frequência a uma imagem interpolada a partir de uma ou mais imagens disponíveis [23] [24] [25] [26]. Pressupõe-se que a simples interpolação dessas imagens não seja suficiente para apresentar detalhes da cena de interesse com nitidez, fazendo-se necessário processar e combinar as diversas imagens disponíveis. É possível que todas as imagens estejam em baixa resolução, de forma que a super-resolução seja realizada através da combinação dessas imagens, ou que a imagem de interesse em baixa resolução seja processada a partir de um banco de imagens em alta resolução.

Estas técnicas de super-resolução encontram uma série de aplicações práticas. É possível, por exemplo, realçar regiões de interesse em uma imagem ou vídeo destinados a diversas atividades: forenses (realce de rostos e placas em videos de câmeras de segurança, por exemplo); médicas (tomografia computadorizada e ressonância magnética); científicas e militares (imagens astronômicas e de satélites). Além disso, pode-se utilizar a super-resolução a fim de reduzir a quantidade de sensores de imageamento de câmeras digitais, tanto para reduzir custos quanto para reaproveitar e melhorar sistemas em baixa resolução já existentes. Outra aplicação interessante para a super-resolução é a conversão de formatos de video em diferentes resoluções.

2.4.1 Super-resolução por combinação de múltiplas imagens

Imagens de alta resolução podem ser criadas a partir da combinação de diversas imagens em baixa resolução [24], no que se costuma chamar super-resolução por combinação de múltiplas imagens. Estas imagens de referência podem ser obtidas através de várias fotografias tiradas com uma mesma câmera em diferentes pontos de vista, através de várias câmeras sincronizadas e posicionadas em diferentes pontos de vista, ou através de uma sequência de video.

Esta forma de super-resolução constitui um processo inverso, em que se estima a fonte de informação a partir dos dados observados. Mais especificamente, a partir de V imagens $\mathbf{Y}_k$ em resolução $N_1 \times N_2$, $k \in \{1, ..., V\}$, deseja-se estimar a imagem $\mathbf{X}$, em resolução $L_1N_1 \times L_2N_2$. Sendo um processo inverso, é necessário modelar a relação entre as entradas $\mathbf{Y}_k$ e a saída $\mathbf{X}$, o que é geralmente feito da seguinte maneira:

$$\mathbf{Y}_k = \mathrm{IM}(\mathbf{X}) + \mathbf{n}_k, \tag{2.9}$$

onde a função IM() representa o sistema de imageamento, ou de captura das imagens, e $\mathbf{n}_k$ representa o sinal de ruído inerente a este sistema. O objetivo da super-resolução é minimizar alguma função de custo entre a estimativa $\mathrm{IM}(\hat{\mathbf{X}})$ e as imagens de entrada $\mathbf{Y}_k$, tais como a solução via mínimos quadrados,

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} J(\mathbf{X}) = \arg\min_{\mathbf{X}} \sum_{k=1}^{V} \|\mathrm{IM}(\mathbf{X}) - \mathbf{Y}_k\|_2^2, \tag{2.10}$$

e a solução langrangiana via mínimos quadrados regularizados, que utiliza uma função $\rho(\mathbf{X})$ para penalizar soluções má-formadas [25],

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \left\{ \sum_{k=1}^{V} \|\mathrm{IM}(\mathbf{X}) - \mathbf{Y}_k\|_2^2 + \lambda\rho(\mathbf{X}) \right\}. \tag{2.11}$$

A escolha adequada da função IM() é fundamental para o bom desempenho desta forma de super-resolução, e costuma-se modelar o sistema de imageamento da seguinte maneira. Primeiramente, representa-se todo o sistema em ordem lexicográfica: $\mathbf{X}$ é substituído por $\mathbf{X}_v = [X(0,0), X(0,1), ..., X(L_1N_1, L_2N_2)]$, $\mathbf{Y}_k$ é substituído por $\mathbf{Y}_{k,v}$, e $\mathbf{n}_k$, por $\mathbf{n}_{k,v}$. O sistema de imageamento mais completo é dado por:

$$\mathbf{Y}_{k,v} = \mathbf{D}\mathbf{B}_k\mathbf{M}_k\mathbf{X}_v + \mathbf{n}_{k,v}, \tag{2.12}$$

onde $\mathbf{D}$ é a matriz $(N_1N_2)^2 \times L_1N_1L_2N_2$ de decimação, $\mathbf{B}_k$ é a matriz $L_1N_1L_2N_2 \times L_1N_1L_2N_2$ de desfoque por movimento, tanto da câmera quanto dos objetos da cena, e $\mathbf{M}_k$ é a matriz $L_1N_1L_2N_2 \times L_1N_1L_2N_2$ de distorção de ótica, devido ao movimento global ou local de objetos e do fundo da cena (translação, rotação, entre outros).

Torna-se necessário estimar o movimento presente na cena das imagens de referência para a imagem de interesse, o que permite compensar o efeitos das matrizes $\mathbf{B}_k$ e $\mathbf{M}_k$. Espera-se que as diversas imagens em baixa resolução possuam deslocamentos em nível de *subpixel*, isto é, as imagens devem estar deslocadas por frações de *pixel*. Caso contrário, cada imagem possuirá a mesma informação, só que em posições inteiras diferentes.

A Fig. 2.6 ilustra os principais estágios da super-resolução por combinação de múltiplas imagens: registro das imagens, combinação e restauração. A primeira etapa corresponde à estimação do deslocamento relativo entre as diversas imagens em baixa resolução, o que permite definir as contribuições em nível de *subpixel* de cada uma das imagens de referência. Como esse deslocamento relativo é arbitrário, é necessário empregar interpolação não-uniforme para combinar estas imagens em posições bem definidas, gerando uma imagem com resolução maior. Por fim, esta imagem final é restaurada, para remover eventuais ruídos e borramentos.

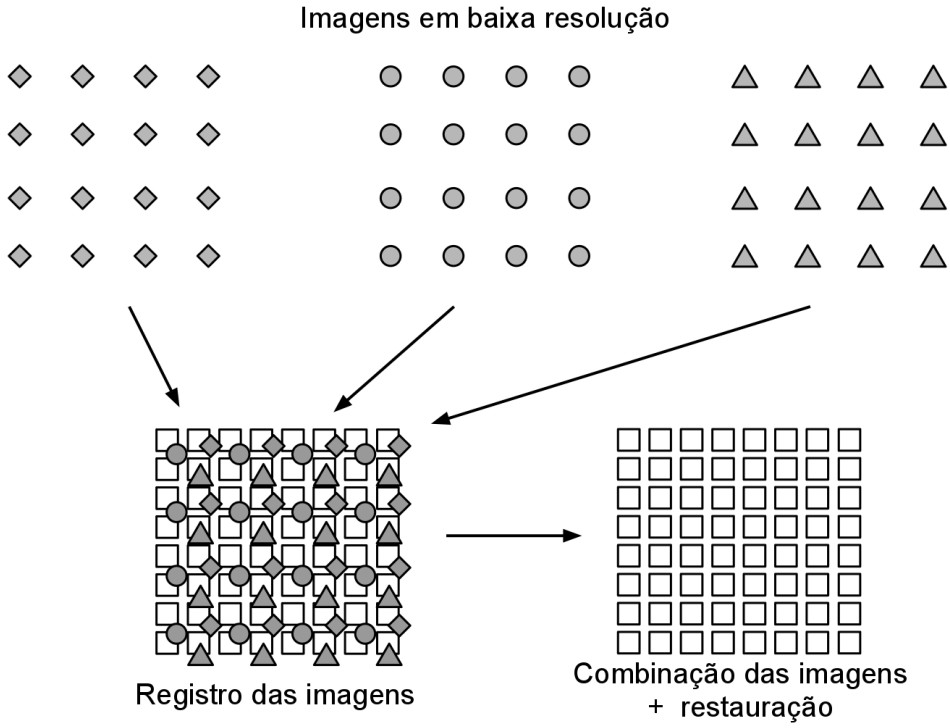


Figura 2.6: Super-resolução por combinação de múltiplas imagens.

Os diferentes métodos propostos para este tipo de super-resolução dizem respeito a questões estruturais, tais como o domínio de aplicação (espacial ou temporal), refinamentos ao modelo de observação considerado e detalhes de cada um dos estágios. O registro das imagens, por exemplo, pode ser feito baseado em blocos ou em *pixels*, por análise de fluxo ótico, dentre outros métodos, e os resultados desta etapa podem ser filtrados de maneira robusta, a fim de retirar resultados muito discrepantes (ou *outliers*, em inglês). Já o estágio de combinação pode ser feito de forma direta ou iterativa, e o estágio de restauração depende do modelo de observação considerado.

É possível também adotar uma perspectiva estocástica ao processo. Empregando-se métodos de estimação *Bayesiana*, a função de densidade de probabilidade *a posteriori* $P(\mathbf{X}|\mathbf{Y}_1, ..., \mathbf{Y}_p)$ pode ser aproximada, de forma que a solução é dada por:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} P(\mathbf{X}|\mathbf{Y}_1, ..., \mathbf{Y}_p), \tag{2.13}$$

e aplicando o teorema de Bayes e a função logarítmica, obtém-se:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} \left\{\log P(\mathbf{Y}_1, ..., \mathbf{Y}_p|\mathbf{X}) + \log P(\mathbf{X})\right\}, \tag{2.14}$$

onde $P(\mathbf{X})$ e $P(\mathbf{Y}_1, ..., \mathbf{Y}_p|\mathbf{X})$ representam conhecimento *a priori* do modelo de imagem e das densidades condicionais, respectivamente. Esta informação *a priori* leva a resultados estáveis, dado que a modelo aplicado seja adequado.

2.4.2 Super-resolução baseada em exemplos

A super-resolução baseada em exemplos [26] acrescenta informações de alta frequência a uma imagem interpolada utilizando uma base de dados com imagens em alta resolução. A princípio, não há relação direta entre a base de dados e a imagem de entrada, de forma que o algoritmo pode atender a uma ampla gama de imagens em baixa resolução, de posse de um grande banco de dados. Assim, esta técnica de super-resolução permite super-resolver uma única imagem, o que não é possível com as técnicas da Seção anterior. Além disso, evita-se capturar a cena a partir de diferentes pontos de vistas ou instantes.

Esta técnica baseia-se no fato que a coleção de imagens que capturam cenas reais, por exemplo, possui variabilidade muito menor do que a coleção de imagens aleatórias. Tais regularidades foram exploradas em estudos de modelagem dos primeiros estágios de processamento dos sistemas visuais de mamíferos [27] [28], e elas são aproveitadas nesta técnica de super-resolução.

Na super-resolução baseada em exemplos, gera-se para cada imagem em alta resolução $\mathbf{I}_j$ uma versão das componentes de baixa frequência, $\mathbf{I}_j^B$, e uma versão das componentes de alta frequência, $\mathbf{I}_j^A$, onde $\mathbf{I}_j^A = \mathbf{I}_j - \mathbf{I}_j^B$. Além disso, considera-se que a imagem interpolada $\mathbf{I}_0$ corresponde somente às componentes de baixa frequência, ou seja, $\mathbf{I}_0 = \mathbf{I}_0^B$. Isto permite obter uma relação entre a imagem de entrada $\mathbf{I}_0$ e as imagens de referência $\mathbf{I}_j$.

$\mathbf{I}_0^B$, $\mathbf{I}_j^B$ e $\mathbf{I}_j^A$ são divididas em blocos regulares, e procura-se em $\mathbf{I}_j^B$ pelo bloco mais semelhante a cada um dos blocos de $\mathbf{I}_0^B$. De acordo com a posição do bloco escolhido em $\mathbf{I}_j^B$, extrai-se o bloco em $\mathbf{I}_j^A$. Este bloco de alta frequência é então somado a $\mathbf{I}_0$, acrescentando detalhes que não puderam ser obtidos através da interpolação. A Fig. 2.7 ilustra este processo.

A Fig. 2.8 ilustra detalhes dos resultados da super-resolução baseada em exemplos de segundo quadro da sequência *Ballet* [19], vista 0, utilizando o primeiro quadro da mesma vista como referência. As Figs. 2.8(a) e (b) apresentam blocos correspondentes, de tamanho 16×16 *pixels*, em $\mathbf{I}_O^B$ e em $\mathbf{I}_O^A$; o objetivo do algoritmo é obter a melhor aproximação a este último bloco. A Fig.

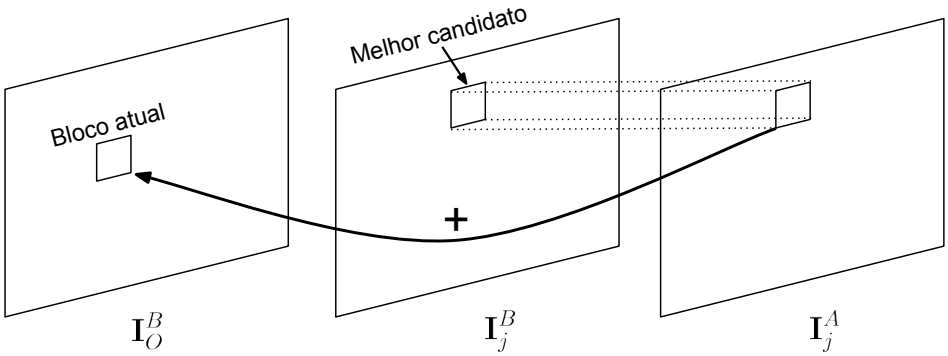


Figura 2.7: Super-resolução baseada em exemplos: uma imagem interpolada $\mathbf{I}_0^B$ recebe informações de alta frequência a partir de uma imagem em alta resolução $\mathbf{I}_j$ separada em versões com componentes de baixa e alta frequência, $\mathbf{I}_j^B$ e $\mathbf{I}_j^A$.

2.8(c) apresenta uma janela de busca de tamanho 176×176 em $\mathbf{I}_j^B$, e a Fig. 2.8(d) apresenta a janela correspondente em $\mathbf{I}_j^A$. De toda a janela, o bloco de tamanho 16×16 da Fig. 2.8(e) é escolhido como o mais próximo à Fig. 2.8(a), que é super-resolvida pela alta frequência correspondente, apresentada na Fig. 2.8(f).

Melhorias ao método podem ser obtidas alterando a forma de escolha do melhor candidato de alta frequência do banco de dados. Da mesma forma que os blocos de $\mathbf{I}_O^B$ e $\mathbf{I}_j^B$ devem estar bem correlacionados, os blocos escolhidos para comporem $\hat{\mathbf{I}}_O^A$, a alta frequência de $\mathbf{I}_O^B$, devem estar bem correlacionados entre si. Ao invés de comparar somente blocos de $\mathbf{I}_O^B$ e $\mathbf{I}_j^B$, acrescenta-se à comparação a informação dos blocos de alta frequência vizinhos já escolhidos, tanto em $\hat{\mathbf{I}}_O^A$ quanto em $\mathbf{I}_j^A$. Tendo a Fig. 2.9 como referência, ao invés de comparar somente os *pixels* nas posições A-P de $\mathbf{I}_O^B$ com as posições correspondentes em $\mathbf{I}_j^B$, acrescenta-se as posições Q-Y de $\hat{\mathbf{I}}_O^A$ e de $\mathbf{I}_j^A$ à comparação, de forma ponderada. Assim, os blocos de alta frequência que irão compor $\hat{\mathbf{I}}_O^A$ serão escolhidos não somente pela correlação entre $\mathbf{I}_O^B$ e $\mathbf{I}_j^B$, mas também pela própria correlação entre eles.

A melhora na qualidade de $\mathbf{I}_O$ é obviamente dependente do tipo de imagem que se usa como referência. O interessante desta técnica de super-resolução é que melhorias podem ser obtidas para a imagem $\mathbf{I}_O$ utilizando imagens aparentemente não correlacionadas. Por exemplo, Freeman *et. al.* [26] super-resolvem a imagem colorida de uma flor a partir de um banco de dados constituído por imagens coloridas em alta resolução de pessoas. Em compensação, a mesma imagem da flor é corrompida quando se aplica uma banco de dados constituído por texto monocromático. Ou seja, o banco de dados não pode ser aleatoriamente escolhido, sendo necessário algum conhecimento mínimo *a priori* da cena que se deseja super-resolver.

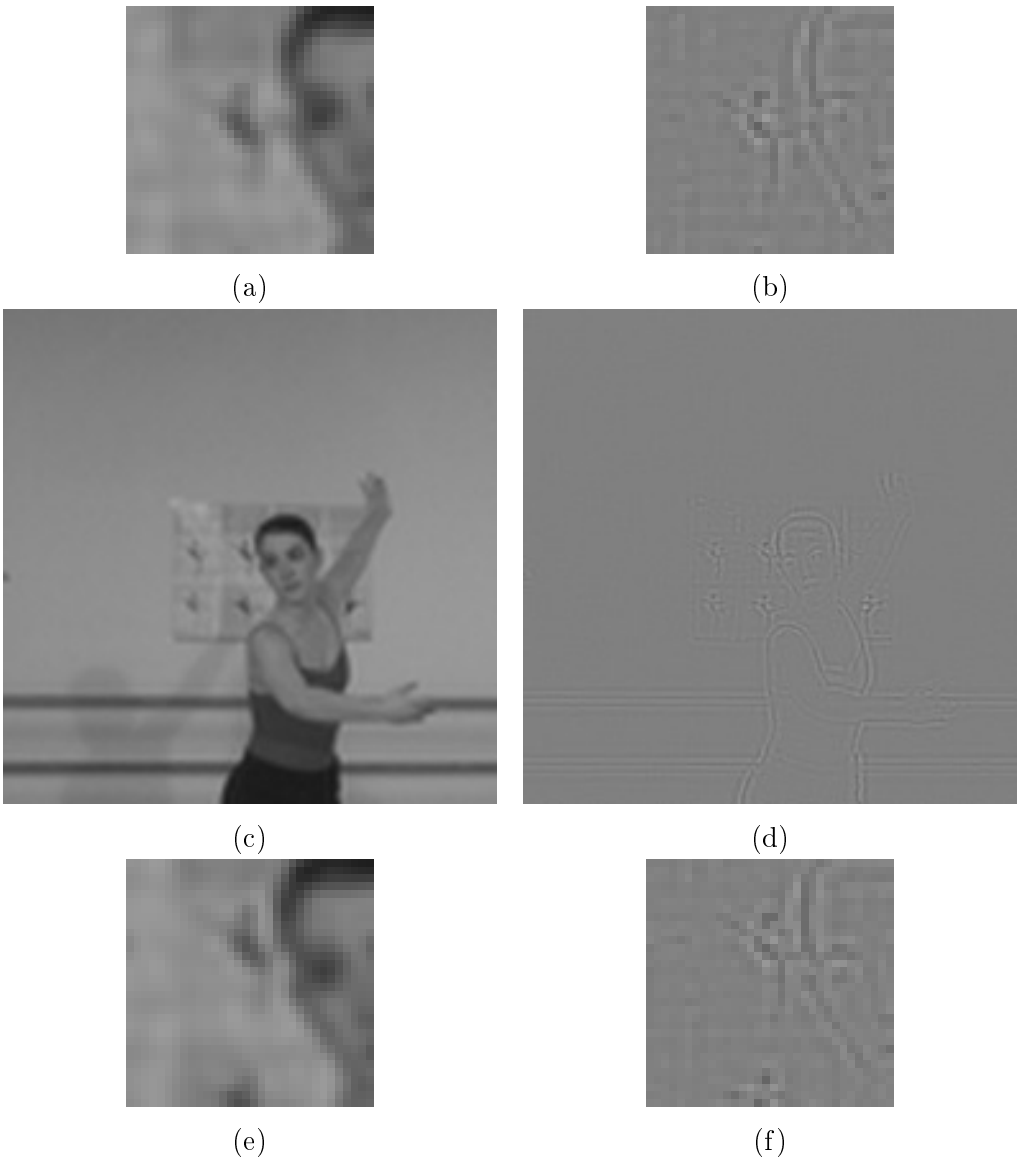


Figura 2.8: Super-resolução baseada em exemplos para uma sequência real: (a) Bloco de tamanho 16×16 em $\mathbf{I}_O^B$; (b) Bloco correspondente ao anterior em $\mathbf{I}_O^A$ (objetivo do algortimo); (c) Janela de busca de tamanho 176×176 em $\mathbf{I}_j^B$; (d) Altas frequências em $\mathbf{I}_j^A$, correspondentes à janela anterior; (e) Bloco de tamanho 16×16 em $\mathbf{I}_j^B$ mais similar a (a); (f) Alta frequência correspondente a (e), utilizada para super-resolver $\mathbf{I}_O^B$. Foi adicionado às Figs (b), (d) e (f) o valor 128 (metade da resolução em 8 *bits*), para que estas pudessem ser apresentadas corretamente, já que $\mathbf{I}_O^A$ e $\mathbf{I}_j^A$ possuem valores negativos.

2.5 Padrão H.264/AVC de compressão de video

Uma sequência de video normalmente acarreta uma grande quantidade de dados, devido à amostragem no tempo e no espaço. Por exemplo, uma sequência de video capturada e com taxa de amostragem de 30 quadros por segundo, em resolução 1280×720, no espaço de cores Y:Cb:Cr 4:2:0 com 8 *bits* de precisão, gera $30 \times 1280 \times 720 \times 3/2/2^{20} = 39,551$ *megabytes* por segundo de

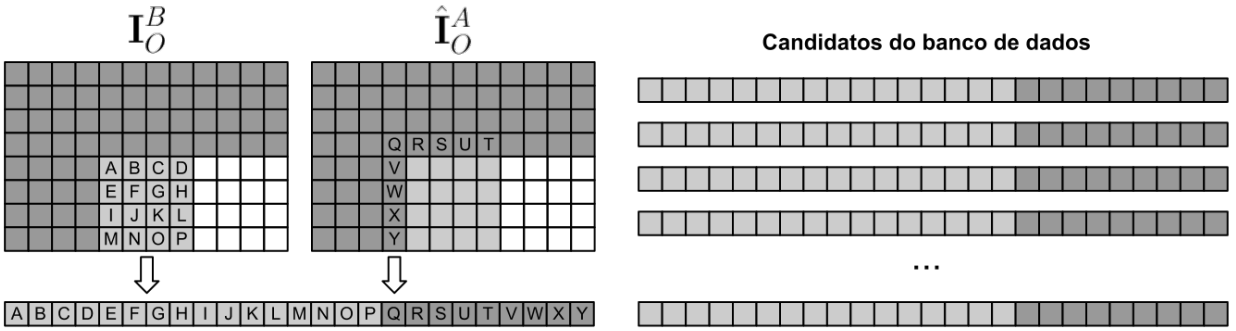


Figura 2.9: Extração do bloco atual de baixa frequência em $\mathbf{I}_O^B$ em conjunto com seus vizinhos espaciais de alta frequência em $\hat{\mathbf{I}}_O^A$, a serem utilizados na comparação com os candidatos correspondentes no banco de dados.

informação na forma de *pixels*. Dessa forma, a redução de dados de video faz-se necessária em uma grande gama de aplicações, tais como o *streaming* via *internet* e videoconferências [20].

O H.264/AVC é o padrão de codificação de video da indústria, desenvolvido pela ITU-T (*International Telecommunication Union*) e pelo ISO/IEC (*International Organisation for Standardisation/International Electrotechnical Commission*) [29]. Ele define uma sintaxe para video comprimido e um método para decodificar esta sintaxe, produzindo uma sequência de video. Desta maneira, o padrão H.264/AVC não define como codificar uma sequência de video, mas sim o formato final que deverá ser seguido pelo codificador e pelo decodificador.

Os esquemas típicos de um codificador e de um decodificador de video (denominados conjuntamente de *codec*) dentro do padrão H.264/AVC são apresentados nas Figs. 2.10(a) e (b). No codec, cada quadro da sequência de video é dividido em blocos de tamanho 16×16, denominados *macroblocos* (MB). No codificador, Fig. 2.10(a), cada macrobloco é processado basicamente pelas seguintes etapas: predição, transformação, escalonamento e codificação de entropia. Já no decodificador, Fig. 2.10(b), os macroblocos passam pelas etapas de decodificação de entropia, re-escalonamento, transformação inversa e reconstrução.

2.5.1 Predição

Utilizando informação já codificada, é gerada uma predição para cada macrobloco. Se a predição for gerada utilizando informação que pertence à versão reconstruída do quadro sendo codificado, a predição é denominada **Intra**, e se a informação utilizada vier de outro quadro reconstruído, a predição é denominada **Inter**. O codificador subtrai a predição do macrobloco original, gerando um resíduo, que poderá ser codificado mais adiante utilizando menos *bits* do que o macrobloco original.

Quanto melhor for a predição, mais semelhante ela será ao macrobloco original, e menor será a quantidade de *bits* necessários para representar o resíduo. Como existem diversas maneiras de prever um macrobloco, o codificador deverá indicar ao decodificador qual predição foi usada, o que

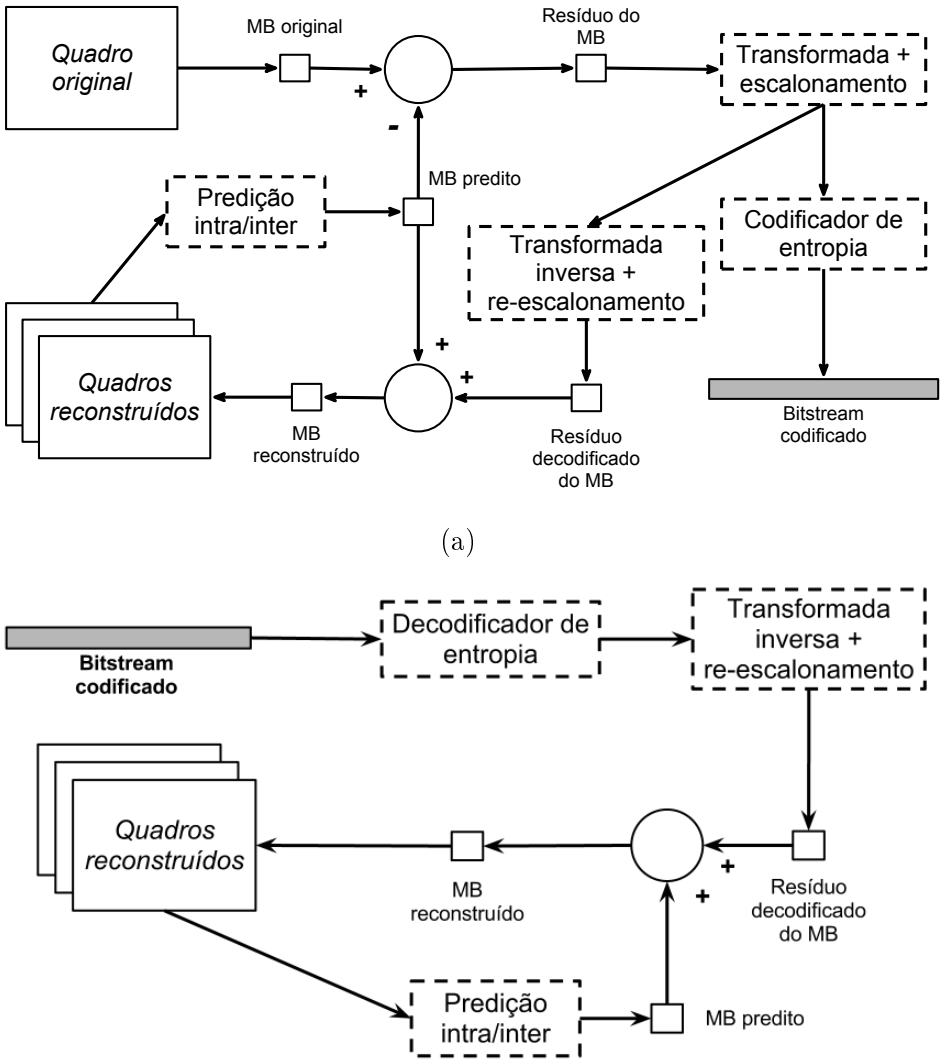


Figura 2.10: Esquema típico de um *codec* H.264/AVC: (a) codificador; (b) decodificador.

se traduz em mais informação a ser enviada. Assim, nem sempre a melhor predição será utilizada, caso a quantidade de *bits* necessários para indicar o tipo de predição aumente em demasia o número total de *bits* na representação do macrobloco.

A predição Intra é feita através da extrapolação de *pixels* vizinhos reconstruídos ao macrobloco sendo predito. Esta extrapolação pode ser feita a partir de várias direções, e para o macrobloco inteiro ou separadamente em blocos de tamanhos 4×4 ou 8×8. As direções de predição disponíveis para o bloco 4×4 são apresentadas na Fig. 2.11.

A predição Inter busca blocos semelhantes ao bloco a ser codificado nas versões reconstruídas de quadros já codificados. Os quadros já codificados podem ser temporalmente anteriores ou posteriores ao quadro atual. Macroblocos podem ser inter-preditos por blocos de tamanhos 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 e 4×4.

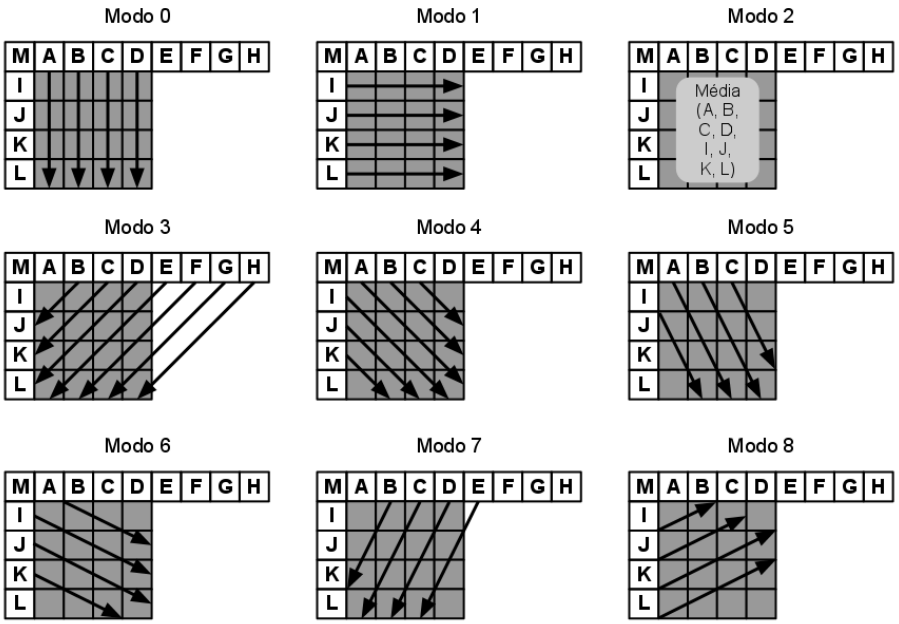


Figura 2.11: Direções de predição Intra para blocos de tamanho 4×4.

A predição Inter para um bloco de tamanho $M \times N$ é feita da seguinte maneira: primeiro, busca-se em um quadro já codificado (chamado quadro de referência) por um bloco semelhante ao bloco a ser predito. Esta busca é feita no quadro de referência em uma região em volta da posição do bloco atual, e a semelhança é avaliada por uma métrica como a SSD (do inglês *sum of squared differences*, ou soma do erro quadrático), por exemplo. A SSD é dada por:

$$\text{SSD} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \{I_A(i+u, j+v) - I_R(i+u+vm_u, j+v+vm_v)\}^2 , \tag{2.15}$$

onde $\mathbf{I}_A$ e $\mathbf{I}_R$ são os quadros atual e de referência, respectivamente, u e v são as coordenadas espaciais do primeiro *pixel* do bloco a ser predito, e $u+vm_u$ e $v+vm_v$ são as coordenadas espaciais do primeiro *pixel* do bloco de referência. vm_u e vm_v representam portanto o deslocamento espacial do bloco de referência em relação ao bloco atual, sendo denominados vetores de movimento.

Esta busca é também denominada **estimação de movimento**, e é ilustrada na Fig. 2.12. No padrão H.264/AVC, ela é feita com precisão de 1/4 de *pixel* (através da interpolação dos quadros de referência), o que melhora a qualidade da predição.

Terminada a estimação de movimento, tem-se o melhor candidato a predição (o bloco que minimiza a SSD) e os vetores de movimento correspondentes. Em seguida, realiza-se a **compensação de movimento**, em que o bloco atual é subtraído do bloco de referência, gerando o resíduo, da mesma forma que na predição Intra. Novamente, nem sempre o bloco que gera a melhor predição será escolhido, caso a quantidade de *bits* necessários para indicar os vetores de

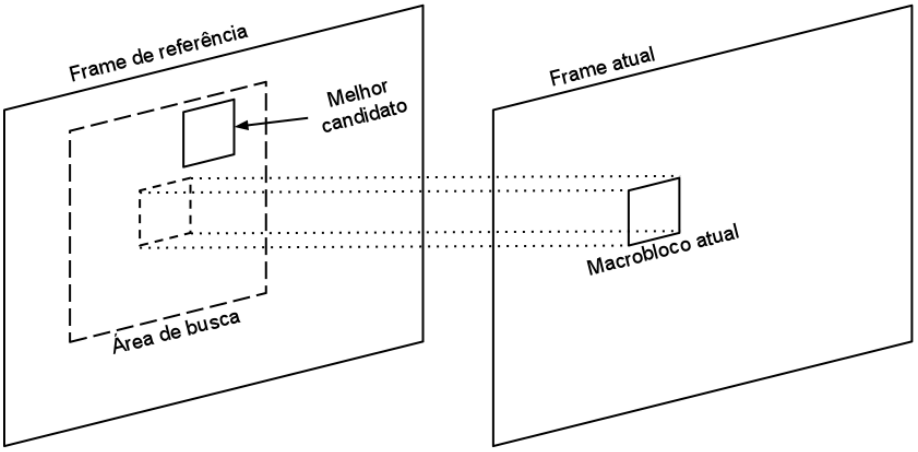


Figura 2.12: Ilustração dos conceitos básicos de estimação de movimento.

movimento aumente em demasia o número total de *bits* usados na representação do macrobloco atual.

2.5.2 Transformação e escalonamento

Após a etapa de predição, obtem-se o resíduo, que representa a diferença entre o macrobloco a ser predito e a predição escolhida. Este resíduo é transformado utilizando uma transformada de valores inteiros, de tamanho 4×4 ou 8×8, que aproxima a transformada DCT (do inglês *Discrete Cosine Transform*, ou transformada discreta de cossenos). Esta transformação procura diminuir a correlação entre as amostras e compactar a energia em um menor número de amostras. Dessa forma, o resíduo transformado pode ser representado mais adiante com menos *bits* do que o resíduo propriamente dito.

O resíduo transformado é também escalonado, o que corresponde ao valor inteiro da divisão de cada coeficiente por um valor especificado anteriormente, o **parâmetro de escalonamento** (QP, do inglês *quantization parameter*). Isso diminui a precisão do resíduo transformado, introduz distorção na imagem final depois da decodificação (pois o escalonamento é não-reversível), mas também reduz a quantidade de *bits* necessários para representação. O escalonamento introduz uma relação de compromisso entre a qualidade final do video decodificado e a quantidade de *bits* necessários na representação do video, regulada através do parâmetro de escalonamento.

No decodificador, são empregados os processos de re-escalonamento e transformação inversa, que são duais aos processos de escalonamento e transformação, respectivamente. O re-escalonamento corresponde à multiplicação dos coeficientes escalonados pelo parâmetro de escalonamento. Já a transformação inversa leva os coeficientes re-escalonados do domínio da transformada de volta para o domínio espacial.

2.5.3 Codificação e decodificação de entropia

As etapas anteriores de predição, transformação e escalonamento servem basicamente como uma preparação para a etapa de codificação de entropia, que é responsável por representar as informações das etapas anteriores de forma comprimida. Dentre estas informações, tem-se o resíduo transformado e escalonado, vetores de movimento, modos de predição Intra, dimensões de cada quadro, quantidade de quadros e espaço de cores utilizado.

O padrão H.264/AVC oferece dois tipos de codificação de entropia: codificação por comprimento variável (VLC, do inglês *Variable Length Coding*) e codificação aritmética (CABAC, do inglês *Context-Adaptive Binary Arithmetic Coding*). No primeiro caso, símbolos que ocorrem com maior frequência são representados por códigos com menor número de *bits*, o que leva a uma menor representação dos dados de forma geral. O VLC parte de uma tabela pré-definida, que mapeia os símbolos aos códigos.

No caso do CABAC, uma sequência de símbolos é convertida para um número fracional único, que é representado por uma série de códigos. Agrupando símbolos, o CABAC atinge melhores taxas de compressão do que o VLC.

Nas Figs. 2.10(a) e (b), os códigos criados pela codificação de entropia (ou pelo VLC ou pelo CABAC) foram denominados *bitstream* codificado.

A decodificação de entropia corresponde ao inverso do processo descrito nesta Seção, de forma que os códigos são interpretados como símbolos (VLC) ou sequências de símbolos (CABAC). Os processos de codificação e decodificação de entropia são reversíveis, não havendo perda de informação de um processo para outro. Sendo assim, o principal responsável no padrão H.264/AVC por introduzir distorção na sequência de video é a etapa de escalonamento.

2.5.4 Reconstrução e outros processos

No decodificador, o *bitstream* codificado passa pelas etapas de decodificação de entropia, reescalonamento e transformação inversa, gerando o resíduo decodificado do macrobloco (como mostra a Fig. 2.10(b)). Para obter a sequência de video ao final, é necessário somar a predição utilizada para o macrobloco atual ao resíduo decodificado do macrobloco, gerando a reconstrução do macrobloco.

Além dos processos descritos nas Seções 2.5.1 a 2.5.3, o padrão H.264/AVC oferece uma série de outras ferramentas que melhora ainda mais o desempenho do *codec* desenvolvido ou adequa o mesmo a uma série de aplicações. Entre eles, incluem-se o filtro redutor de blocagem, que reduz os artefatos de bloco introduzidos pela transformação baseada em bloco.

2.6 Métricas de qualidade de video

A compressão de uma sequência de video pode acarretar em diferentes níveis de degradação em relação à sequência original, de acordo com o parâmetro de escalonamento utilizado (Seção

2.5.2). Além disso, a mesma sequência de video convertida do espaço de cores RGB para Y:Cb:Cr e depois de volta para RGB pode não ser recuperada integralmente, visto que ambos os espaços são comumente representados em 8 *bits*, e a conversão da Eq. (2.3) requer o arredondamento de valores nos dois espaços. Estes diversos tipos de processamento de video geram perda de informação e erros em relação ao video original, de forma que é fundamental quantificar estes erros [20].

2.6.1 Métricas subjetivas de qualidade de video

Métricas subjetivas de qualidade de video [20] levam em consideração que o usuário final de uma sequência de video geralmente é um ser humano, apesar de nem sempre este ser o caso, como no uso de video para o reconhecimento automático de faces, por exemplo. Métricas subjetivas envolvem a avaliação de sequências por uma série de observadores humanos em um ambiente controlado.

A recomendação BT.500-11 da ITU-R [30] apresenta uma série de procedimentos para a avaliação subjetiva de video, sendo o mais comum o DSCQS (do inglês *Double Stimulus Continuous Quality Scale*, ou escala contínua de qualidade por estímulo duplo). Neste procedimento, o observador é apresentado aleatoriamente às versões originais ou corrompidas de sequências de video, devendo atribuir notas de 'ruim' a 'excelente'. As notas são convertidas a uma faixa normalizada, denominada MOS (do inglês *mean opinion score*, ou nota de opinião média), que representa a qualidade relativa entre as sequências originais e corrompidas.

2.6.2 Métricas objetivas de qualidade de video

Métricas como a DSCQS consomem muitos recursos de espaço, tempo e dinheiro, dado que elas requerem um ambiente controlado e um vasto número de observadores. Assim, tornam-se necessárias avaliações mais rápidas de qualidade de video, sendo a solução mais comum as métricas baseadas em algoritmos, ou métricas objetivas. Elas oferecem resultados rápidos e repetíveis, mas não refletem integralmente as nuances da percepção humana subjetiva.

A métrica mais utilizada é a PSNR (do inglês *peak signal-to-noise ratio*) [20]. Ela mede a razão entre o máximo sinal possível e o erro quadrático médio (MSE, do inglês *mean square error*):

$$\mathrm{PSNR}(\mathbf{I}_O, \mathbf{I}_C) = 10 * log_{10}\left(\frac{(2^n-1)^2}{\mathrm{MSE}(\mathbf{I}_O, \mathbf{I}_C)}\right), \tag{2.16}$$

onde $\mathbf{I}_O$ e $\mathbf{I}_C$ são as imagens original e corrompida, respectivamente, e n é o número de *bits* de resolução de cada *pixel* das imagens. O MSE é dado por:

$$\mathrm{MSE}(\mathbf{I}_O, \mathbf{I}_C) = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\left\{I_O(i,j) - I_C(i,j)\right\}^2, \tag{2.17}$$

onde W e H são a largura e o comprimento das imagens, respectivamente. Para sequências de video, a PSNR é calculada como sendo a média das PSNRs calculadas quadro a quadro [31].

A PSNR sofre de uma série de limitações. Em primeiro lugar, ela exige o conhecimento da sequência original, que pode não estar disponível. Em segundo lugar, ela não reflete testes subjetivos como o DSCQS com fidelidade. Desta maneira, outras métricas objetivas foram desenvolvidas, para melhor aproximar os resultados objetivos e subjetivos.

A métrica SSIM (do inglês *Structural Similarity Index*, ou índice de similaridade estrutural [32]) procura comparar duas imagens de forma similar a uma avaliação subjetiva das mesmas. Baseado no sistema visual humano, esta métrica aponta mudanças perceptíveis em informação estrutural na imagem, ao invés de apontar o erro médio da imagem, como na Eq. (2.16). O SSIM evita análises estruturais globais entre imagens, dividindo estas em blocos de tamanho $N \times N$. Para cada bloco $(\mathbf{u}_i, \mathbf{v}_i)$ (onde $\mathbf{u}_i$ e $\mathbf{v}_i$ são vetores de comprimento N, com as posições espaciais do bloco em questão), o SSIM entre as imagens original $\mathbf{I}_O(\mathbf{u}_i, \mathbf{v}_i)$ e corrompida $\mathbf{I}_C(\mathbf{u}_i, \mathbf{v}_i)$ é dado por:

$$\mathrm{SSIM}\left(\mathbf{I}_O(\mathbf{u}_i, \mathbf{v}_i), \mathbf{I}_C(\mathbf{u}_i, \mathbf{v}_i)\right) = \frac{\left(2\mu_O\mu_C + c_1\right)\left(2\sigma_{OC} + c_2\right)}{\left(\mu_O^2 + \mu_C^2 + c_1\right)\left(\sigma_O^2 + \sigma_C^2 + c_2\right)}, \tag{2.18}$$

onde μ_O e μ_C são as médias de $\mathbf{I}_O(\mathbf{u}_i, \mathbf{v}_i)$ e $\mathbf{I}_C(\mathbf{u}_i, \mathbf{v}_i)$, respectivamente, σ_O e σ_C são as variâncias respectivas, σ_{OC} é a covariância entre $\mathbf{I}_O(\mathbf{u}_i, \mathbf{v}_i)$ e $\mathbf{I}_C(\mathbf{u}_i, \mathbf{v}_i)$, e c_1 e c_2 são dados por:

$$\begin{aligned} c_1 &= 0.01(2^n - 1) \\ c_2 &= 0.03(2^n - 1) \end{aligned}. \tag{2.19}$$

Os valores c_1 e c_2 estão presentes na Eq. (2.18) para estabilizar a divisão por denominadores próximos a zero, como é o caso em blocos muito suaves, cuja variância é baixa.

Na prática, o índice SSIM é avaliado sobre toda a imagem, portanto costuma-se calcular a média dos valores de SSIM (MSSIM) em todos os blocos da imagem:

$$\mathrm{MSSIM}(\mathbf{I}_O, \mathbf{I}_C) = \frac{N^2}{WH} \sum_{j=1}^{WH/N^2} \mathrm{SSIM}\left(\mathbf{I}_O(\mathbf{u}_j, \mathbf{v}_j), \mathbf{I}_C(\mathbf{u}_j, \mathbf{v}_j)\right). \tag{2.20}$$

$\mathbf{u}_j$ e $\mathbf{v}_j$ indicam os *pixels* de cada bloco, em um total de WH/N^2 blocos. O MSSIM varia entre 0 e 1, sendo este último valor encontrado quando $\mathbf{I}_O = \mathbf{I}_C$. Wang *et.al.* recomendam que o MSSIM seja calculado somente no componente de luminância, e utilizam blocos de tamanho 8×8. Além disso, para evitar artefatos de bloco, as estimativas de μ_O, μ_C, σ_O, σ_C e σ_{OC} utilizam uma função circular-simétrica de pesos gaussianos normalizados, de tamanho 11×11, com desvio-padrão de $1,5$ amostras.

Capítulo 3

Representação de video em múltiplas vistas

3.1 Introdução

Neste Capítulo, estendem-se os conceitos de video digital para mais de uma vista, através da captura simultânea da cena por várias câmeras. Apresenta-se a relação geométrica entre estas câmeras e a representação da informação tridimensional da cena através de mapas de profundidade. Técnicas de compressão de múltiplas vistas são apresentadas em seguida, incluindo a compressão de mapas de profundidade e a codificação em resolução mista, que é objeto de estudo neste trabalho. O Capítulo é concluído com aplicações práticas para o uso de câmeras simultâneas.

3.2 Representação de múltiplas vistas

A representação de cenas naturais e sintéticas em múltiplas vistas refere-se ao uso de mais de uma câmera para capturar as imagens, simultaneamente. A princípio, a replicação da representação de vistas da Seção 2.2 é suficiente para descrever várias vistas, mas existem diversas relações geométricas a serem exploradas entre pares de câmeras, como será visto a seguir.

A representação de cenas naturais e sintéticas parte do modelo da câmera de câmara escura [33], ou modelo *pinhole*, que é composto por um plano de imagem e um centro ótico $\mathbf{C}$ a uma distância f do plano de imagem. f é a chamada distância focal, que é medida através do eixo principal da câmera, ou eixo ótico. A interseção $\mathbf{P}$ deste eixo com o plano de imagem denomina-se ponto principal. O plano que contém $\mathbf{C}$ e é paralelo ao plano de imagem é denominado plano focal, ou plano principal. A Fig. 3.1 ilustra estes conceitos.

Qualquer ponto tridimensional $\mathbf{M}$ em frente à câmera *pinhole* é ligado a $\mathbf{C}$ através de uma única linha, cuja interseção com o plano de imagem compõe a projeção $\mathbf{m}$ do ponto $\mathbf{M}$ sobre este plano. Como mostra a Fig. 3.2, pode-se obter a relação entre $\mathbf{m}$ e $\mathbf{M}$ por similaridade de triângulos, respeitando a razão entre y e z e entre f e v, que é a coordenada no plano de imagem.

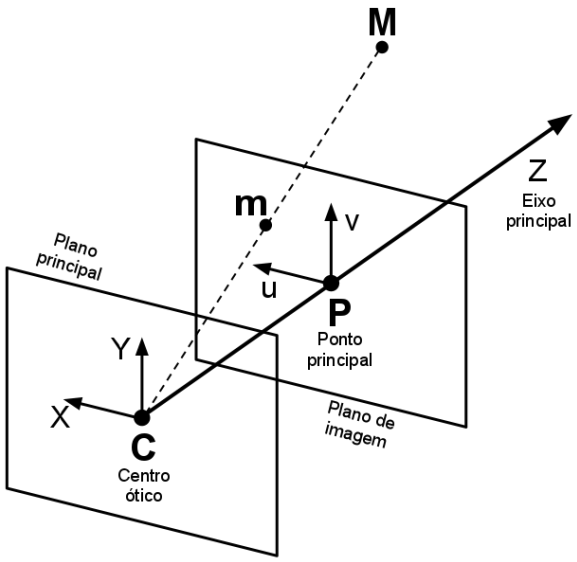


Figura 3.1: Geometria da câmera de modelo *pinhole*.

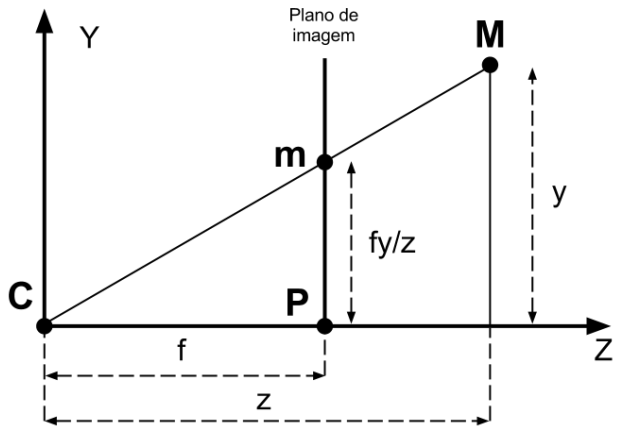


Figura 3.2: Relação entre os pontos $\mathbf{m}$ e $\mathbf{M}$ por similaridade de triângulos.

Se considerarmos pontos com coordenadas homogêneas $\mathbf{m} = (u, v, 1)^T$ e $\mathbf{M} = (x, y, z, 1)^T$, tem-se a seguinte relação:

$$z\mathbf{m} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{M}. \tag{3.1}$$

A Eq. 3.1 é uma versão simplificada da câmera *pinhole*, levando em conta somente a distância focal. A fim de incluir informações como a posição relativa da câmera ao plano euclidiano e a resolução de amostragem digital, tem-se a seguinte projeção perspectiva:

$$\zeta \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$
$$\mathbf{P} = \mathbf{K} \begin{bmatrix} \mathbf{R} \mid \mathbf{t} \end{bmatrix} \quad , \tag{3.2}$$
$$\mathbf{K} = \begin{bmatrix} f/s_x & 0 & o_x \\ 0 & f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

onde ζ é a distância do ponto $\mathbf{M}$ ao plano principal da câmera, $\mathbf{P}$ é a matriz 3×4 de projeção da câmera, $\mathbf{K}$ é a matriz de calibração da câmera, $\mathbf{R}$ é a matriz 3×3 de rotação da câmera e $\mathbf{t}$ é o vetor 3×1 de translação da câmera. $\mathbf{t}$ e $\mathbf{R}$ descrevem a posição e a orientação da câmera em relação à origem euclidiana, respectivamente, descrevendo portanto parâmetros extrínsecos à câmera. $\mathbf{K}$ realiza a transformação de coordenadas da câmera para coordenadas dos *pixels*, e depende de parâmetros intrínsecos da câmera, a saber:

- f, distância focal;
- s_x e s_y, largura e comprimento dos *pixels* da câmera, em milímetros;
- e o_x e o_y, coordenadas do centro da imagem, em *pixels*.

3.2.1 Geometria epipolar

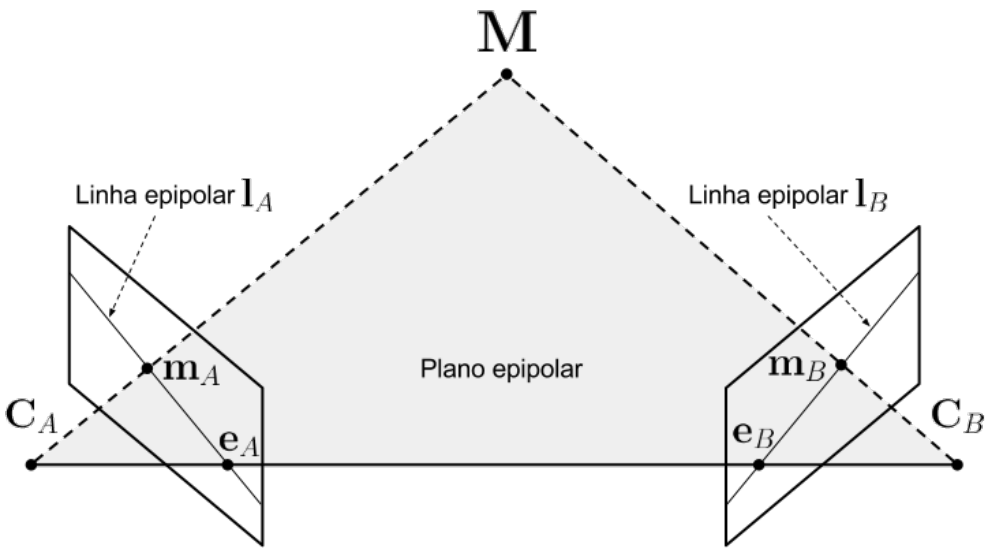


Figura 3.3: Geometria epipolar: duas câmeras, A e B, capturam uma cena a partir de dois pontos de vista, e o ponto $\mathbf{M}$ é mapeado para os pontos $\mathbf{m}_A$ e $\mathbf{m}_B$ respectivos às câmeras.

A projeção em (3.2) liga o ponto $\mathbf{M}$ ao centro ótico $\mathbf{C}$ através de uma única linha, o que é válido também para todos os pontos que pertencem a esta linha. Assim, todos estes pontos serão

encobertos pelo ponto na reta mais próximo a $\mathbf{C}$, perfazendo oclusões de todos os outros pontos. Uma segunda câmera poderá ser capaz de apresentar estes pontos encobertos, dependendo da geometria da cena.

Com base na Fig. 3.3, se considerarmos duas câmeras *pinhole*, A e B, e suas respectivas matrizes de projeção $\mathbf{P}_A$ e $\mathbf{P}_B$, o ponto $\mathbf{M}$ é projetado sobre os pontos $\mathbf{m}_A$ e $\mathbf{m}_B$ de cada uma dessas câmeras da seguinte maneira:

$$\begin{array}{l} \zeta_A \mathbf{m}_A = \mathbf{P}_A \mathbf{M} = \mathbf{K}_A(\mathbf{R}_A.[x\ y\ z]^T + \mathbf{t}_A) \\ \zeta_B \mathbf{m}_B = \mathbf{P}_B \mathbf{M} = \mathbf{K}_B(\mathbf{R}_B.[x\ y\ z]^T + \mathbf{t}_B) \end{array}, \tag{3.3}$$

onde ζ_A e ζ_B são as distâncias do ponto $\mathbf{M}$ aos planos principais das câmeras A e B, respectivamente. A Eq. 3.3 permite assim encontrar correspondências entre *pixels* de duas câmeras, dadas suas matrizes de projeção e as profundidades de cada *pixel*. Partindo da câmera A, projeta-se o ponto $\mathbf{m}_A$ ao ponto $\mathbf{M}$:

$$[x\ y\ z]^T = \mathbf{R}_A^{-1}(\mathbf{K}_A^{-1}\zeta_A \mathbf{m}_A - \mathbf{t}_A), \tag{3.4}$$

que é projetado diretamente ao ponto $\mathbf{m}_b$ através de:

$$\mathbf{m}_B = \mathbf{P}_B \mathbf{M}/\zeta_B. \tag{3.5}$$

Ainda com base na Fig. 3.3, percebe-se que o ponto $\mathbf{M}$ e os centros óticos das câmeras A e B, $\mathbf{C}_A$ e $\mathbf{C}_B$, definem o chamado plano epipolar, ao qual pertencem $\mathbf{m}_A$ e $\mathbf{m}_B$. Os pontos $\mathbf{e}_A$ e $\mathbf{e}_B$ representam as projeções de $\mathbf{C}_B$ na câmera A e de $\mathbf{C}_A$ na câmera B, respectivamente, e as linhas epipolares $\mathbf{l}_A$ e $\mathbf{l}_B$ constituem interseções entre o plano epipolar e os planos de imagens das câmeras A e B, respectivamente. É importante notar que se o ponto $\mathbf{m}_A$ é conhecido, e deseja-se encontrar o ponto $\mathbf{m}_B$ correspondente em B, o plano epipolar restringe esta busca à linha epipolar $\mathbf{l}_B$, o que reduz a área de busca significativamente.

3.2.2 Mapas de profundidade

O mapa de profundidades $\mathbf{D}$ apresenta as profundidades ζ de cada *pixel* [34] [35], e é representado com uma imagem de mesma resolução da imagem colorida da câmera, e que contém somente o canal de luminância. A fim de representar para cada posição (u, v) a profundidade ζ correspondente em 8 *bits*, é necessário conhecer as profundidades máxima, ζ_{max}, e mínima, ζ_{min}, e realizar um escalonamento:

$$D(u,v) = 255\left(\frac{1}{\zeta(u,v)} - \frac{1}{\zeta_{max}}\right) \Big/ \left(\frac{1}{\zeta_{min}} - \frac{1}{\zeta_{max}}\right). \tag{3.6}$$

Desta forma, $\zeta(u,v) = \zeta_{max}$ corresponde a $D(u,v) = 0$, e $\zeta(u,v) = \zeta_{min}$ corresponde a $D(u,v) = 255$. O processo pode ser revertido da seguinte maneira:

$$\zeta(u,v) = \left\{ \frac{D(u,v)}{255} \left(\frac{1}{\zeta_{min}} - \frac{1}{\zeta_{max}} \right) + \frac{1}{\zeta_{max}} \right\}^{-1}. \tag{3.7}$$

A conversão de profundidades ζ para representação em 8 *bits* possui duas vantagens [35]. Como os valores de ζ são invertidos, uma maior resolução em $D(u,v)$ é conferida a objetos mais próximos, como ilustra a Fig. 3.4. Alem disso, os valores de $D(u,v)$ são adimensionais, e não dependem de informações de câmeras adjacentes, como suas posições em relação à origem euclidiana, tamanho do sensor ou resolução da imagem. Quer dizer, independente da câmera que se utilize, $D(u,v)$ estará sempre representado por valores entre 0 e 255, para uma precisão de 8 *bits*.

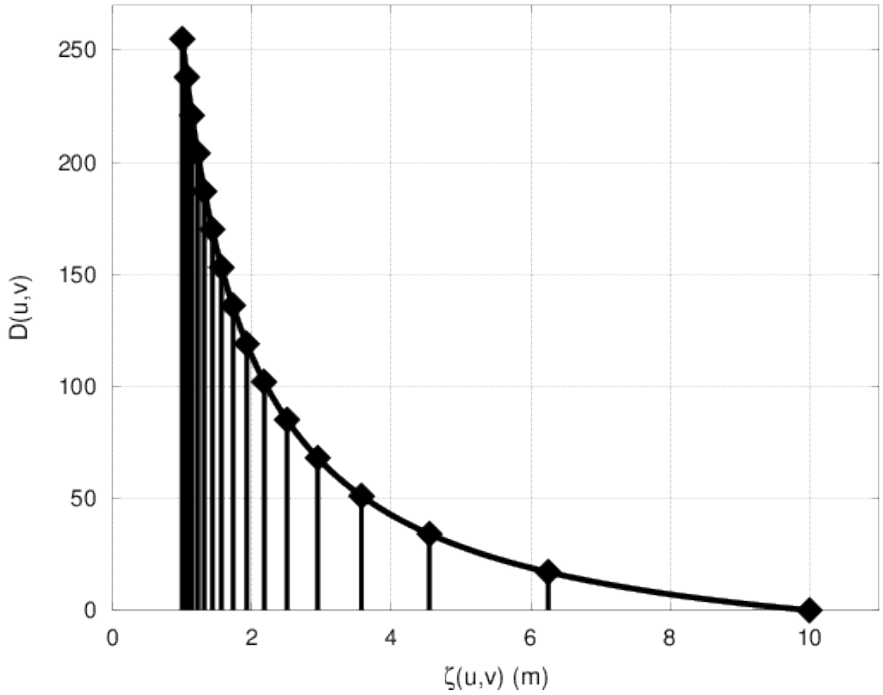
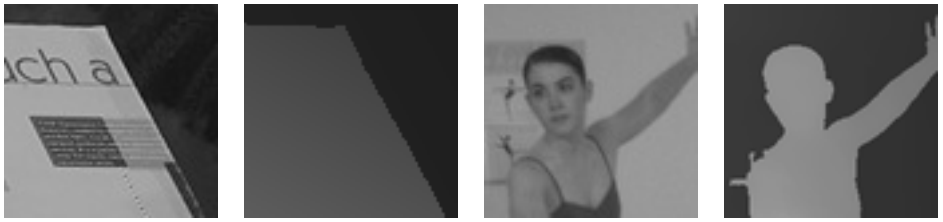


Figura 3.4: Conversão de profundidades $\zeta(u,v)$ entre 1 e 10 metros para representação em 8 *bits*: valores de $\zeta(u,v)$ mais próximos à câmera são amostrados com maior precisão.

Dada uma sequência de video, o mapa de profundidades correspondente e as matrizes de calibração, rotação e translação da câmera, é possível gerar vistas virtuais próximas à posição da câmera original, baseado na geometria epipolar (Eqs. 3.4 e 3.5). Esta técnica não evita problemas de oclusão, que devem ser tratados de maneira adequada, e permite não somente gerar múltiplas vistas de uma cena como também ajustar a sensação de profundidade para telas esteroscópicas.

Mapas de profundidade geralmente apresentam grandes áreas suaves, com pequenas variações de um *pixel* para outro, e contornos bem definidos de objetos presentes na cena [36]. Esta é uma característica não só de mapas utilizados para gerar cenas sintéticas, como também de mapas calculados para cenas naturais, como mostra a Fig. 3.5. É importante notar que mapas de profundidade estão sujeitos a erros, tais como representação em precisão reduzida, oclusões e escalonamento por codificação. Por exemplo, é possível visualizar na Fig. 3.5(d) uma imperfeição

no canto inferior esquerdo do mapa de profundidade, aonde é considerada a mesma profundidade para a bailarina e para uma pequena parte do poster atrás da bailarina.



(a) (b) (c) (d)

Figura 3.5: Detalhes de cenas com mapas de profundidade típicos: (a) componente de luminância da sequência sintética *Venus*, vista 6, instante 0; (b) mapa de profundidade correspondente; (c) componente de luminância da sequência real *Ballet*, vista 1, instante 1; (d) mapa de profundidade correspondente.

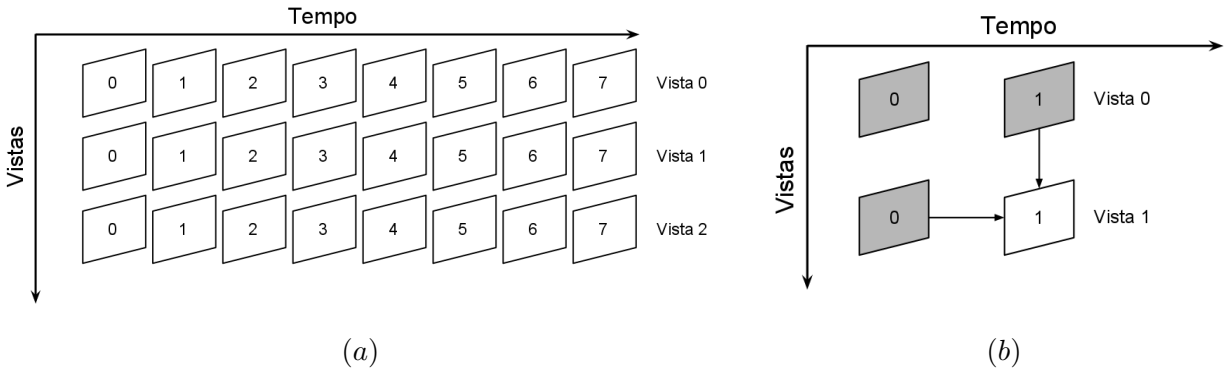
Existem três maneiras básicas para se obter o mapa de profundidades de uma cena. Para o caso de sequências sintéticas, ele já é parte inerente ao conteúdo, já que ele é utilizado diretamente na geração da cena sintética. Para cenas reais, o mapa de profundidades pode ser obtido prontamente através de câmeras sensoras de profundidade, ou pode ser extraído de um par de câmeras calibradas, através do cálculo de correspondências entre elas [2]. Neste último caso, existe uma grande gama de algoritmos disponíveis.

O cálculo de correspondências entre câmeras possui quatro etapas básicas: cálculo de custo das correspondências, agregação dos custos, otimização das disparidades e refinamento das disparidades. Por exemplo, se o cálculo de correspondências for feito com base na minimização da soma do erro quadrático entre as câmeras (Eq. 2.15), o custo das correspondências é medido através da diferença quadrática das imagens em uma dada disparidade, a agregação dos custos é feita pela soma da diferença quadrática dentro de um bloco, e a disparidade ótima é selecionada pela escolha do mínimo valor agregado *pixel* a *pixel*. Além da minimização da soma do erro quadrático, outros métodos incluem o uso da correlação cruzada normalizada [37] [38] e a utilização de métodos globais, tais como o arrefecimento cruzado [39] [40] (ou *simulated annealing*, em inglês), a difusão probabilística [41] e o corte de grafos [42] (ou *graph cuts*, em inglês).

3.3 Compressão de múltiplas vistas

A necessidade de compressão para sequências de video de uma vista foi apresentada na Seção 2.5. Sequências de video em múltiplas vistas acentuam esta necessidade, visto que a taxa de dados aumenta de forma proporcional ao número de vistas acrescentadas. Por exemplo, uma cena capturada em 8 vistas, com resolução 1280×720, no espaço de cores Y:Cb:Cr 4:2:0 com 8 *bits* de precisão, e com taxa de amostragem de 30 quadros por segundo gera $8 \times 1280 \times 720 \times 3/2 \times 30/2^{20} = 316,41$ *megabytes* por segundo de informação na forma de *pixels*, o que equivale a $8\times$ o valor apresentado na Seção 2.5.

A Fig. 3.6(a) ilustra os 8 primeiros quadros de uma sequência de 3 vistas. Utilizando o padrão H.264/AVC, é possível comprimir cada uma destas vistas separadamente, o que é conhecido como codificação H.264/AVC *simulcast* [1]. Nesse caso, não se leva em conta nenhuma redundância existente entre as vistas. O padrão H.264/AVC oferece uma extensão que oferece mais funcionalidades, de forma a melhor aproveitar estas redundâncias, sendo denominada *H.264/AVC Multiview Video Coding* (MVC) [1] [29].



(*a*) (*b*)

Figura 3.6: Conceitos de codificação em múltiplas vistas: (a) Quadros de uma sequência de video em múltiplas vistas; (b) Predição do quadro 1 da vista 1 usando os quadros 0 da vista 1 e 1 da vista 0 como referência.

3.3.1 Compressão de video em múltiplas vistas

Em termos da relação entre taxa e distorção, o que faz o MVC ser mais eficiente que a codificação H.264/AVC *simulcast* é a predição entre vistas [1], aonde os processos de estimação e compensação de movimento (Seção 2.5.1) são feitos tendo um quadro de outra vista já codificado como referência. Por exemplo, considerando a Fig. 3.6(b), o quadro 1 da vista 1 pode ser predito no MVC tanto pelo quadro 0 da vista 1 como pelo quadro 1 da vista 0. Como as câmeras são sincronizadas, não há exatamente um movimento de objetos do quadro 1 da vista 0 para o quadro 1 da vista 1. Nesse caso, a predição entre vistas é denominada **estimação e compensação de disparidade**.

O MVC oferece uma vista de base, codificada independentemente das outras vistas, de forma que o *bitstream* codificado atenda a decodificadores H.264/AVC. Assim, este *bitstream* pode atender tanto a televisores comuns quanto a receptores 3D, por exemplo. A fim de transmitir informações de outras vistas, tais como identificação e dependência entre elas, o MVC estende a sintaxe de alto nível, através do conjunto de parâmetros da sequência (SPS, do inglês *sequence parameter set*).

Existem outras técnicas que oferecem ganhos significativos ao MVC, mas que não foram incluídos neste padrão por representarem mudanças estruturais e de sintaxe significativas em relação ao H.264/AVC. Isso tornaria estes padrões incompatíveis, inviabilizando o aproveitamento de sistemas já existentes para a codificação de múltiplas vistas. Dentre estas técnicas, incluem-se a pré-compensação de iluminação para a estimação de disparidade [43] [44], a filtragem adaptativa das referências, que compensa grandes diferenças focais entre estas e o quadro a ser predito [45]

[46], e a predição através da síntese de vistas [47] [48] [49]. Também foi proposto um modo de predição de vetores de movimento a partir da correlação entre vetores já escolhidos em outras vistas [50] [51].

Na média, o MVC oferece 20% de redução na taxa de *bits* com relação ao H.264/AVC *simulcast*, para sequências de até 8 vistas. Em compensação, a quantidade de operações realizada no processo de codificação MVC é drasticamente superior a esta quantidade na codificação H.264/AVC *simulcast*, primariamente devido ao aumento de referências utilizadas no processo de predição.

3.3.2 Compressão de mapas de profundidade

Assim como as sequências de video em múltiplas vistas, mapas de profundidade também necessitam de compressão, para fins de transmissão e armazenamento. A forma mais direta é aplicar o MVC, exatamente como apresentado na Seção anterior. Porém, o H.264/AVC não leva em conta as características particulares aos mapas de profundidade [36].

Por se tratarem de sequências de video com grandes áreas suaves e transições bruscas entre objetos, os mapas de profundidade possuem coeficientes grandes tanto para altas frequências quanto as baixas. A compressão de video pode borrar as bordas dos objetos através do processo de escalonamento, causando artefatos claramente vísiveis ao observador humano. Além disso, os mapas não são utilizados em si, mas sim para auxiliar na síntese de vistas adjacentes.

A Fig. 3.7 ilustra estas questões, aonde a vista 6 da sequência *Lovebird1* [52], quadro 0, foi sintetizada utilizando a vistas 4 e 8 com e sem compressão H.264/AVC em modo Intra. O mapa de profundidades da vista 4 decodificado, Fig. 3.7(c), possui bordas extremamente borradas quando comparado ao mapa original, Fig. 3.7(a), e na vista 6 sintetizada a partir de quadros decodificados, 3.7(d), os contornos da pessoa na cena são claramente afetados, enquanto o rosto guarda grande semelhança, quando comparado com a vista 6 sintetizada com os quadros originais, sem compressão, Fig. 3.7(b).

Sendo assim, foi desenvolvida uma série de técnicas para preservar as bordas de objetos com mais fidelidade e garantir uma melhor qualidade nas vistas sintetizadas com o uso dos mapas. Uma das abordagens mais comuns para a compressão de mapas de profundidade é considerar modos de predição que minimizam o erro da vista sintetizada, e não o erro da representação do mapa [53]; alternativamente, pode-se empregar métodos de síntese de vistas que façam processamento diferenciado em bordas de objetos, compensando os erros causados nestas regiões pela compressão dos mapas de profundidade [19] [54]. Outras abordagens incluem: o uso de modos de predição que privilegiam a qualidade nas bordas de objetos, tais como a codificação por *wavelets* [55] [56] ou *platelets* [57]; a codificação dos mapas em baixa resolução, utilizando métodos não-lineares de interpolação no decodificador para recuperar as informações das bordas dos objetos [58].

Figura 3.7: Detalhes da síntese da vista 6 da sequência *Lovebird1*, quadro 0, utilizando diferentes mapas de profundidade: (a) mapa de profundidade original da vista 4; (b) vista 6 sintetizada com as vistas 4 e 8 originais (PSNR de $25,39$ dB e MSSIM de $80,57\%$); (c) mapa de profundidade da vista 4 decodificado após compressão com o padrão H.264/AVC, modo Intra, $QP = 40$; (d) após compressão das vistas 4 e 8 com o padrão H.264/AVC, modo Intra, $QP = 22$ para cor e $QP = 40$ para os mapas, obtém-se a vista 6 sintetizada (PSNR de 25.40 dB e MSSIM de $80,95\%$).

3.3.3 Teoria da supressão binocular

Uma série de estudos aponta outras soluções viáveis na redução de taxa em sistemas de múltiplas vistas, baseado em certas características do sistema visual humano. Diversos estudos psicológicos e fisiológicos mostram que o sistema visual humano obtém adequadamente a impressão

de profundidade em visão estéreo mesmo em situações de qualidade assimétrica, no que é conhecido por teoria da supressão binocular [11]. Por exemplo, se uma das vistas for borrada por um filtro passa-baixas, a imagem binocular e a informação de profundidade permanecem subjetivamente inalteradas, e os detalhes de alta frequência perdidos na vista borrada são compensados pela outra vista. Em compensação, caso uma das vistas seja degradada por erros de codificação por escalonamento, a impressão binocular pode ser piorada.

A Fig. 3.8 ilustra estes cenários. Detalhes das vistas originais 6 e 2 da sequência *Teddy* [2] podem ser vistas nas Figs. 3.8(a) e (b), respectivamente, e o leitor pode obter uma impressão tridimensional da cena cruzando os olhos até uma imagem coincidir com a outra (ao custo de acentuada fadiga visual). As Figs. 3.8(c) e (e) repetem a vista original 6 da cena. A Fig. 3.8(d) apresenta a vista 2 após ter sido decimada e interpolada por um fator $M = 2$, utilizando o filtro de Lanczos (Eqs. 2.5 e 2.6), o que resulta em uma PSNR de $30,19$ dB (Eqs. 2.16 e 2.17) e uma MSSIM de $88,98\%$ (Eqs. 2.18 a 2.20) em relação ao original. A Fig. 3.8(f) apresenta a versão decodificada da compressão da vista 2 com o padrão H.264/AVC em modo Intra (Seção 2.5.1), o que resulta em uma PSNR de 30.01 dB e uma MSSIM de $82,95\%$ em relação ao original. Repetindo o teste de vistas cruzadas, o leitor percebe pouca diferença entre as vistas tridimensionais resultantes.

Diversos estudos foram realizados para avaliar estes diferentes métodos de qualidade assimétrica. Tam *et. al.* [12] observaram que no caso de pares estéreo borrados assimetricamente, a qualidade da imagem binocular é dominada pela imagem menos borrada, e no caso de pares estéreo assimetricamente escalonados, a qualidade da imagem binocular corresponde aproximadamente à média do par estéreo. A impressão de profundidade é aproximadamente igual nos dois casos. Aflaki *et. al.* [13] apresentaram resultados subjetivos semelhantes para pares estéreo codificados em resolução mista, em resolução completa e em qualidade mista, considerando que a decimação de uma das vistas e sua subsequente interpolação para a resolução original funciona de forma semelhante a borrar a imagem com o filtro passa-baixas. Saygili *et. al.* [14] indicaram que para uma qualidade suficientemente alta em uma das vistas, a degradação só é percebida quando a outra vista é codificada em uma qualidade abaixo de um limiar dependente da tela 3D. Acima deste limiar, usuários preferem qualidade mista, e abaixo do limiar, preferem resolução mista.

3.3.4 Codificação de múltiplas vistas em resolução mista

Apoiando-se na teoria da supressão binocular, o formato de resolução mista (MRC, do inglês *Mixed-Resolution Coding*) torna-se particularmente atrativo para cenários estéreo e de múltiplas vistas. Assim, reduz-se a quantidade de informação a ser transmitida, mantendo-se a qualidade subjetiva da cena. A Fig. 3.9 ilustra uma maneira de aplicar este formato a um sistema com 4 vistas disponíveis, onde somente as vistas ímpares são codificadas em baixa resolução.

A codificação de sequências estéreo em resolução mista é particularmente adequada para a TV 3D em dispositivos móveis, oferecendo a impressão de profundidade sem realizar processamento complexo de síntese de vistas [5] [6]. Ekmekcioglu *et. al.* [7] reportaram ganhos objetivos de qualidade para baixas taxas, dentro do formato MRC estéreo e utilizando diferentes razões de decimação e interpolação. Aksay *et. al.* [8] investigaram o uso da escalabilidade temporal para

MRC estéreo, reportando redução de taxa com a mesma qualidade subjetiva. Chen *et. al.* [9] apresentaram formas de realizar a predição de uma vista em baixa resolução a partir de outra vista em resolução normal, evitando a realização da decimação dos quadros de referência, o que acelera o processo de codificação e decodificação estéreo em resolução mista. Técnicas de síntese de vistas também foram empregadas para gerar versões de resolução normal para as vistas codificadas em baixa resolução [10].

A maioria destes trabalhos não procura estimar as componentes de alta frequência para as vistas codificadas em baixa resolução, o que torna o MRC inadequado para a televisão de ponto-de-vista livre [15], pois o usuário verá uma sequência de video borrada caso ele escolha um ponto de vista próximo ou idêntico à vista codificada em baixa resolução. Alternativamente, a vista em baixa resolução poderia não ser enviada, e sim sintetizada a partir de vistas adjacentes em resolução normal [15]. Ainda assim, o usuário perceberia artefatos nas imagens sintetizadas, que aumentam em proporção à distância das vistas adjacentes.

3.4 Aplicações de videos em múltiplas vistas

A disponibilidade de mais de uma câmera captando a mesma cena presta-se a uma série de aplicações [1] [36]. Telas estereoscópicas com óculos especiais necessitam de duas câmeras devidamente distanciadas, que são apresentadas a cada um dos olhos do espectador separadamente, oferecendo a impressão de profundidade. Telas autoestereoscópicas possuem tecnologia mais sofisticada, oferecendo a impressão de profundidade sem a necessidade de óculos especiais. Em compensação, estas telas requerem um número muito maior de vistas disponíveis (geralmente acima de oito) para apresentarem uma cena tridimensional dentro de um amplo campo de visão.

De posse de duas ou mais câmeras e seus parâmetros intrínsecos e extrínsecos, é possível calcular as correspondências entre *pixels* [2], e a partir destas estimar a profundidade dos pontos tridimensionais na cena. Assim, cria-se um modelo tridimensional da cena, o que permite análises mais completas da estrutura da cena, tais como o reconhecimento de objetos e de gestos corporais, por exemplo.

Com os mapas de profundidades, também é possível projetar *pixels* para câmeras em posições virtuais [15], o que se aplica a uma série de cenários. De posse de uma câmera colorida e um mapa de profundidade correspondente, é possível gerar um par estereoscópico. A vantagem deste cenário é que a quantidade de informação a ser transmitida é bem menor, comparado ao envio de duas câmeras coloridas exclusivamente, visto que um mapa de profundidade não contém componentes de crominância e apresenta grandes áreas suaves, mais suscetíveis para a compressão.

O cenário anterior oferece um campo de visão limitado, dado que uma única câmera não resolve o problema das oclusões, mencionado anteriormente. De posse de duas ou três câmeras e de mapas de profundidade correspondentes, é possível aumentar o campo de visão através da síntese de vistas, atendendo inclusive a telas autoestereoscópicas e reduzindo drasticamente a quantidade de informação a ser transmitida.

A projeção de *pixels* de uma câmera a outra é especialmente importante para o cenário da televisão de ponto-de-vista livre (do inglês *free-viewpoint television*) [15], aonde o usuário decide sob qual ponto de vista ver a cena (paralaxe de movimento). Como não é possível captar e nem transmitir todos os pontos de vistas possíveis, os mapas de profundidades tornam-se fundamentais neste cenário, que é particularmente apropriado em teleconferências imersivas.

Figura 3.8: Ilustração da teoria da supressão binocular com detalhes da sequência *Teddy*, vistas 2 e 6, quadro 0. Cruzando os olhos até uma imagem coincidir com a outra, o leitor pode obter uma impressão tridimensional da cena, percebendo pouca diferença subjetiva entre as cenas tridimensionais geradas nas três linhas, apesar de elas terem sido geradas por diferentes pares estereoscópicos. As Figs. (a), (c) e (e) apresentam a versão original da vista 6 da sequência, e as Figs. (b), (d) e (f) apresentam diferentes versões da vista 2. A Fig. (b) apresenta a versão original da vista 2. A Fig. (d) apresenta a versão da vista 2 após a decimação e a interpolação por 2, utilizando o filtro de Lanczos, com PSNR de $30,19$ dB e MSSIM de $88,98\%$ em relação ao original. A Fig. (f) apresenta a versão da vista 2 após a compressão Intra H.264/AVC, com PSNR de 30.01 dB e MSSIM de $82,95\%$ em relação ao original.

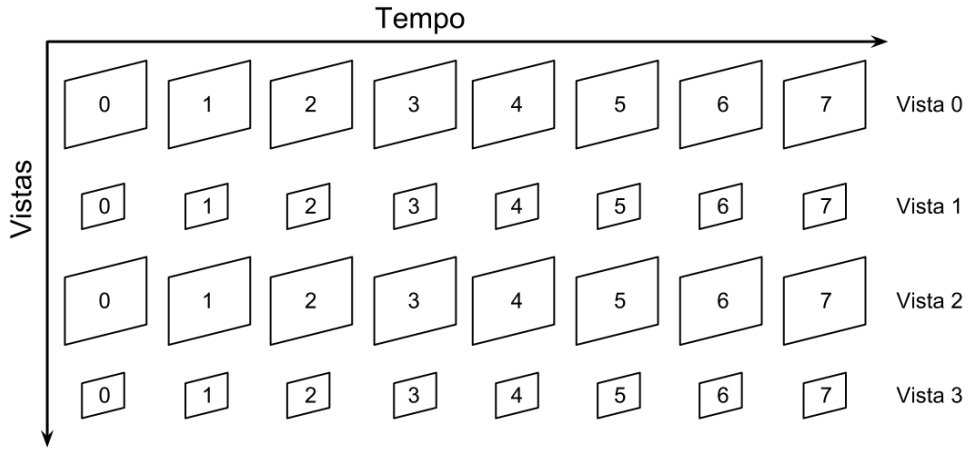


Figura 3.9: Arquitetura de codificação de múltiplas vistas em resolução mista.

Capítulo 4

Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade

4.1 Introdução

Este Capítulo apresenta a primeira das técnicas propostas para sequências de múltiplas vistas em resolução mista. Em primeiro lugar, introduz-se a arquitetura para a qual esta técnica se aplica. A partir deste contexto, apresenta-se a técnica de super-resolução na Seção seguinte. Resultados experimentais para uma série de sequências reais e sintéticas são então apresentados, levando em conta imagens sem perdas e imagens codificadas pelo padrão H.264/AVC.

4.2 Arquitetura em consideração

Nesta arquitetura de codificação, considera-se que cada vista é composta por uma sequência de video com quadros em resolução mista. Além disso, em um dado instante, os quadros em alta e em baixa resolução se alternam ao longo das vistas. A Fig. 4.1 ilustra um exemplo desta arquitetura com quatro vistas disponíveis. Se esta arquitetura for visualizada em uma tela autoestereoscópica, por exemplo, as vistas em baixa resolução deverão ser interpoladas de volta às suas resoluções originais, e o usuário verá um par de vistas em qualidade mista, já que a vista interpolada parecerá mais borrada do que a oura vista. Além disso, a vista borrada será sempre alternada entre os olhos direito e esquerdo do usuário ao longo do tempo. Esta arquitetura também pode ser aplicada a sequências estereoscópicas, onde seriam consideradas somente as vistas 0 e 1 para codificação, usando a Fig. 4.1 como referência.

A presente arquitetura aplica-se a sistemas que trabalham com sequências em resolução mista, mas não possuem poder computacional suficiente para calcular mapas de disparidade, seja do lado do codificador ou do lado do decodificador. No primeiro caso, considera-se aparelhos portáteis

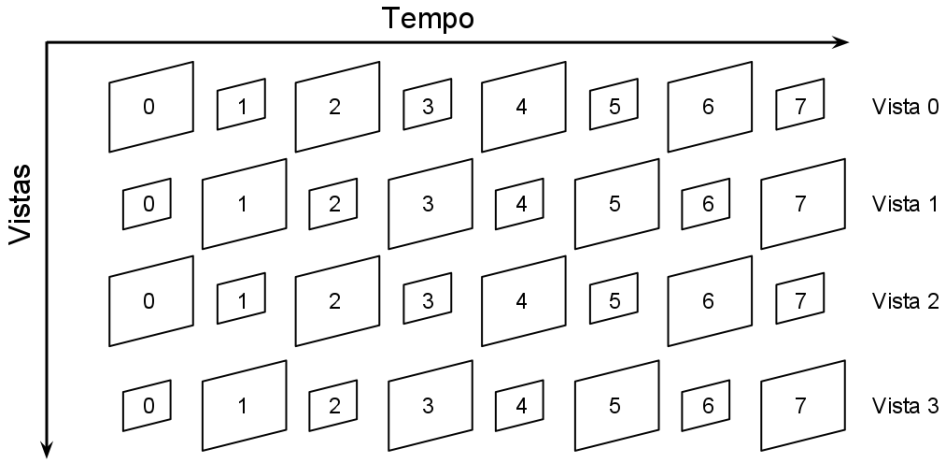


Figura 4.1: Arquitetura de codificação de múltiplas vistas em resolução mista alternada sem mapas de profundidade.

que necessitem de codificação estéreo, tais como câmeras digitais e celulares com câmeras estéreo. No segundo caso, incluem-se sistemas que necessitam prover visão estéreo em tempo real, tais como computadores pessoais sem placas de video dedicadas, e os mesmos aparelhos portáteis mencionados anteriormente.

4.3 Solução proposta

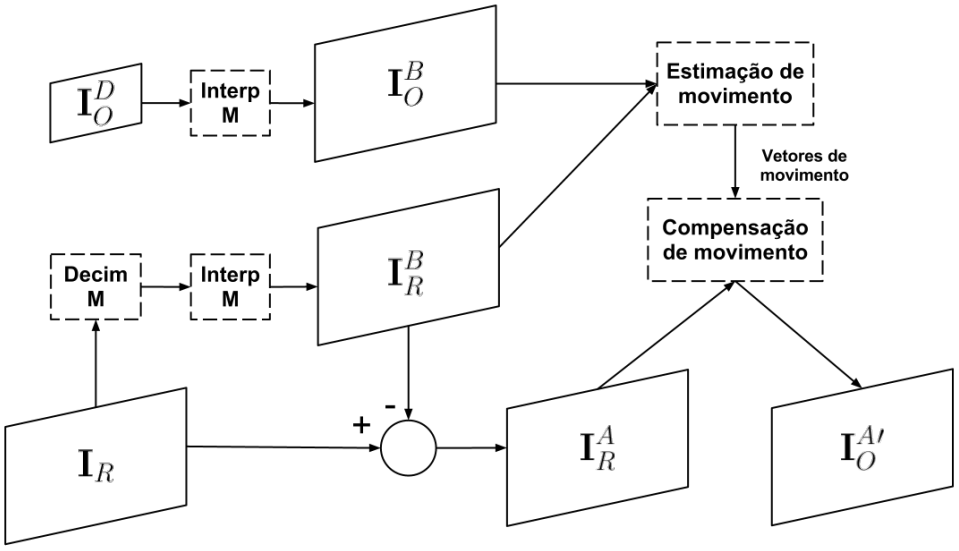
A fim de recuperar informações de alta frequência nas vistas em baixa resolução, apresenta-se uma técnica de super-resolução inspirada na super-resolução baseada em exemplos (Seção 2.4.2) [16] [26]. Ao invés de usar uma base de dados como referência, extrai-se informações de alta frequência a partir dos quadros adjacentes em alta resolução, que possuem alta correlação com o quadro a ser super-resolvido.

Dependendo do quadro, poderá haver até quatro quadros de referência, podendo pertencer tanto à outra vista quanto à mesma vista que o quadro a ser super-resolvido. Usando a Fig. 4.1 como referência, o quadro 2 da vista 1 pode ser super-resolvido pelos quadros 1 e 3 da vista 1, além dos quadros 2 das vistas 0 e 2.

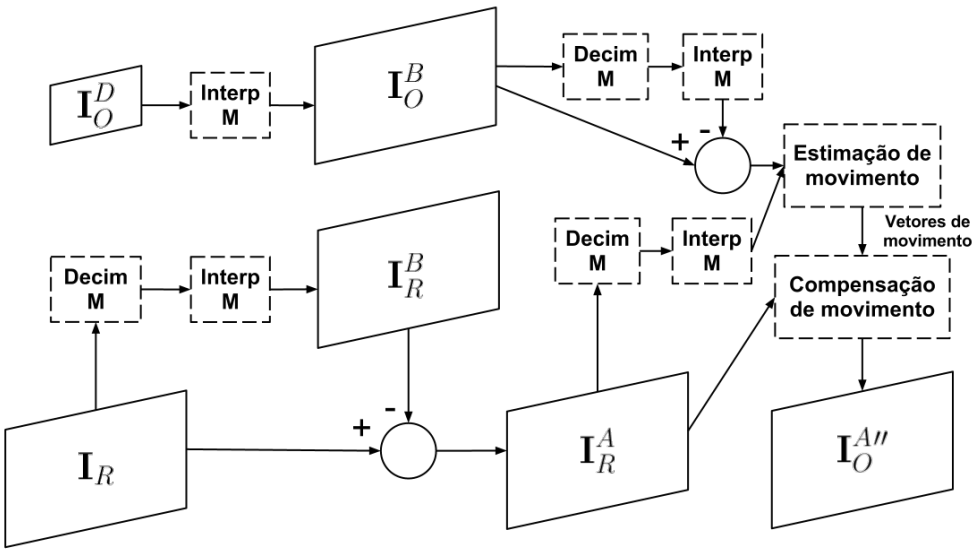
O método proposto é dividido em etapas de extração e combinação de altas frequências. A seguir, define-se cada uma destas etapas.

4.3.1 Extração de altas frequências

Dado um quadro decimado na j-ésima vista e k-ésimo instante, $\mathbf{I}^D_{j,k}$, ele é interpolado à sua resolução original, gerando $\mathbf{I}^B_{j,k}$. A fim de simplificar a notação, $\mathbf{I}^D_{j,k} = \mathbf{I}^D_O$ e $\mathbf{I}^B_{j,k} = \mathbf{I}^B_O$. Este último quadro constitui a versão de baixa frequência do quadro original $\mathbf{I}_O$ ($\mathbf{I}_{j,k}$). O objetivo



(a)



(b)

Figura 4.2: Extração de altas frequências do quadro de referência $\mathbf{I}_R$: (a) com base na correlação com a versão de baixa frequência de $\mathbf{I}_R$; (b) com base na correlação com a versão de alta frequência de $\mathbf{I}_R$.

final do algoritmo proposto é gerar a versão super-resolvida $\hat{\mathbf{I}}_O$ mais semelhante a $\mathbf{I}_O$, a partir da estimativa $\hat{\mathbf{I}}_O^A$ da alta frequência $\mathbf{I}_O^A$. Aqui, assume-se que:

$$\hat{\mathbf{I}}_O = \mathbf{I}_O^B + \hat{\mathbf{I}}_O^A. \tag{4.1}$$

Definindo $\mathbf{I}_R$ como o quadro adjacente na m-ésima vista e n-ésimo instante (onde $(m,n) \in \{(j-1,k),(j+1,k),(j,k-1),(j,k+1)\}$, dependendo da disponibilidade), é gerada sua versão

de baixa frequência, $\mathbf{I}_R^B$. Definindo o processo de decimação seguido de interpolação de $\mathbf{I}_R$ como $\mathrm{ID}(\mathbf{I}_R)$ (utilizando o mesmo fator M aplicado a $\mathbf{I}_O$), tem-se:

$$\mathbf{I}_R^B = \mathrm{ID}(\mathbf{I}_R). \tag{4.2}$$

Em seguida, aplica-se o processo de estimação de movimento (Seção 2.5.1) entre $\mathbf{I}_O^B$ e $\mathbf{I}_R^B$, onde este último é considerado a referência. Se estes dois quadros pertencerem a vistas diferentes, o processo é nomeado estimação de disparidade, sendo contudo o mesmo algoritmo. Para cada bloco $(\mathbf{u}_i, \mathbf{v}_i)$, obtém-se vetores de movimento $(vm_{u,i}, vm_{v,i})$ (ou de disparidade, se $(m,n) \in (j-1,k),(j+1,k)$).

De posse dos vetores de movimento e/ou disparidade, gera-se a primeira estimativa de alta frequência, $\mathbf{I}_O^{A\prime}$, realizando a compensação de movimento e/ou disparidade dos blocos do quadro $\mathbf{I}_R^A$ de alta frequência da referência. Este último é criado de forma análoga à Eq. 4.1 por $\mathbf{I}_R^A = \mathbf{I}_R - \mathbf{I}_R^B$. Para cada bloco $(\mathbf{u}_i, \mathbf{v}_i)$, tem-se:

$$\mathbf{I}_O^{A\prime}(\mathbf{u}_i, \mathbf{v}_i) = \mathbf{I}_R^A(\mathbf{u}_i + vm_{u,i}, \mathbf{v}_i + vm_{v,i}). \tag{4.3}$$

O quadro $\mathbf{I}_O^{A\prime}$ já pode ser considerado uma estimativa final da alta frequência de $\mathbf{I}_O$. Ele busca correlação entre os blocos de baixa frequência em $\mathbf{I}_O^B$ e $\mathbf{I}_R^B$, assumindo que as altas frequências correspondentes de $\mathbf{I}_R$ são uma boa estimativa para $\mathbf{I}_O^A$. Entretanto, nem sempre este é o caso, pois o algoritmo definido acima sempre acrescenta altas frequências, independentemente do valor de SSD entre $\mathbf{I}_O^B(\mathbf{u}_i, \mathbf{v}_i)$ e $\mathbf{I}_R^B(\mathbf{u}_i + vm_{u,i}, \mathbf{v}_i + vm_{v,i})$. Assim, é interessante obter outras estimativas para $\mathbf{I}_O^A$.

O quadro $\hat{\mathbf{I}}_O$ depende de $\hat{\mathbf{I}}_O^A$, como mostra a Eq. (4.1). Em termos de erro absoluto, tais como o MSE e a PSNR, quanto mais próximo $\hat{\mathbf{I}}_O$ for de $\mathbf{I}_O$, mais próximo $\mathrm{ID}(\hat{\mathbf{I}}_O)$ deverá ser de $\mathbf{I}_O^B$. Portanto, uma nova estimativa $\mathbf{I}_{O|R}^{A\prime\prime}$ pode ser gerada utilizando vetores de movimento/disparidade a partir da estimação de movimento/disparidade entre $\mathrm{ID}(\hat{\mathbf{I}}_O)$ e $\mathbf{I}_O^B$.

Para se efetuar a operação $\mathrm{ID}(\hat{\mathbf{I}}_O)$, é necessário conhecer a alta frequência que será somada ao quadro $\mathbf{I}_O^B$, tornando recursivo este processo de estimação de movimento/disparidade. A fim de simplificar este processo, utiliza-se uma filtragem linear no processo ID(), tal como o filtro Lanczos indicado na Eq. (2.7), de forma que:

$$\mathrm{ID}(\hat{\mathbf{I}}_O) = \mathrm{ID}(\mathbf{I}_O^B + \hat{\mathbf{I}}_O^A) = \mathrm{ID}(\mathbf{I}_O^B) + \mathrm{ID}(\hat{\mathbf{I}}_O^A). \tag{4.4}$$

O processo de estimação de movimento/disparidade minimiza a SSD, que é uma métrica dependente do erro quadrático. Para quaisquer três números α, β e γ, tem-se que $\{\alpha - (\beta+\gamma)\}^2 = \{(\alpha-\beta)-\gamma\}^2$. Portanto, a SSD entre quaisquer blocos de $\mathbf{I}_O^B$ e $\mathrm{ID}(\hat{\mathbf{I}}_O)$ é igual à SSD entre as mesmas posições destes blocos em $\{\mathbf{I}_O^B - \mathrm{ID}(\mathbf{I}_O^B)\}$ e $\mathrm{ID}(\hat{\mathbf{I}}_O^A)$. Assim, cada bloco da nova estimativa de alta frequência $\mathbf{I}_O^{A\prime\prime}$ é dado por:

$$\mathbf{I}_O^{A\prime\prime}(\mathbf{u}_i, \mathbf{v}_i) = \mathbf{I}_R(\mathbf{u}_i + vm'_{u,i}, \mathbf{v}_i + vm'_{v,i}), \tag{4.5}$$

onde o par $(vm'_{u,i}, vm'_{v,i})$ advém da estimação de movimento/disparidade entre ID($\mathbf{I}_R^A$) e $\{\mathbf{I}_O^B - \text{ID}(\mathbf{I}_O^B)\}$.

A estimativa $\hat{\mathbf{I}}_O = \mathbf{I}_O^B + \mathbf{I}_O^{A\prime\prime}$ será portanto aquela cuja versão de baixas frequências ID($\hat{\mathbf{I}}_O$) é a mais semelhante ao quadro de baixa frequência $\mathbf{I}_O^B$, que é conhecido. Desta forma, $\mathbf{I}_O^{A\prime}$ baseia-se em informação de $\mathbf{I}_R^B$, e $\mathbf{I}_O^{A\prime\prime}$ baseia-se em informação de $\mathbf{I}_R^A$. As Figs. 4.2(a) e (b) ilustram as duas técnicas de extração das altas frequências do quadro de referência.

4.3.2 Combinação de altas frequências

A Seção anterior apresentou duas estimativas para as altas frequências de $\mathbf{I}_O$ (Eqs. 4.3 e 4.5) a partir de um único quadro de referência $\mathbf{I}_R$. A arquitetura apresentada na Seção 4.2 prevê o uso de até quatro referências, sendo portanto necessário combinar estas informações de alta frequência em um único quadro de altas frequências $\hat{\mathbf{I}}_O^A$, de acordo com a Eq. (4.1).

A solução proposta utiliza uma soma ponderada para combinar os blocos $(\mathbf{u}_i, \mathbf{v}_i)$ de cada quadro $\mathbf{I}_O^{A\prime}$ e $\mathbf{I}_O^{A\prime\prime}$. Definindo $\mathbf{I}_{O,r}^{A}{}'$ e $\mathbf{I}_{O,r}^{A}{}''$ como as estimativas criadas a partir da r-ésima referência ($r \in \{1,2,3,4\}$, dependendo da disponibilidade), cada bloco de $\hat{\mathbf{I}}_O^A$ é dado por

$$\hat{\mathbf{I}}_O^A(\mathbf{u}_i, \mathbf{v}_i) = \frac{\sum_{r=1}^{4} \left(w'_{r,i} \mathbf{I}_{O,r}^{A}{}'(\mathbf{u}_i, \mathbf{v}_i) + w''_{r,i} \mathbf{I}_{O,r}^{A}{}''(\mathbf{u}_i, \mathbf{v}_i) \right)}{\sum_{r=1}^{4} \left(w'_{r,i} + w''_{r,i} \right)}, \tag{4.6}$$

onde os pesos $w'_{r,i}$ e $w''_{r,i}$ são dados por

$$\begin{aligned} w'_{r,i} &= 1/\text{SSD}(\mathbf{I}_O^B(\mathbf{u}_i, \mathbf{v}_i),\, \mathbf{I}_{O|R}^{B}{}'(\mathbf{u}_i, \mathbf{v}_i)) \\ w''_{r,i} &= 1/\text{SSD}(\mathbf{I}_O^B(\mathbf{u}_i, \mathbf{v}_i),\, \mathbf{I}_{O|R}^{B}{}''(\mathbf{u}_i, \mathbf{v}_i)) \end{aligned}. \tag{4.7}$$

Os quadros $\mathbf{I}_{O|R}^{B}{}'$ e $\mathbf{I}_{O|R}^{B}{}''$ são obtidos fazendo a compensação de movimento e/ou disparidade do quadro $\mathbf{I}_r^B$ a partir dos vetores de movimento e/ou disparidade utilizados para criar $\mathbf{I}_{O,r}^{A}{}'$ e $\mathbf{I}_{O,r}^{A}{}''$. Quanto mais semelhante a $\mathbf{I}_O^B(\mathbf{u}_i, \mathbf{v}_i)$ for o bloco $\mathbf{I}_{O|R}^{B}{}'(\mathbf{u}_i, \mathbf{v}_i)$, menor será a SSD entre estes dois últimos blocos, e maior será o peso $w'_{r,i}$. O mesmo raciocínio vale para $\mathbf{I}_{O|R}^{B}{}''(\mathbf{u}_i, \mathbf{v}_i)$ e $w''_{r,i}$.

4.4 Resultados experimentais

A fim de avaliar o desempenho do algoritmo proposto na Seção anterior, foi feita uma série de testes com sequências reais e sintéticas de múltiplas vistas, seguindo a arquitetura proposta na Seção 4.2, com e sem codificação H.264/AVC. Os testes sem codificação constituem o limite superior de qualidade alcançado pelo algoritmo proposto, visto que neste caso as sequências não são corrompidas pelo processo de escalonamento (Seção 2.5.2), enquanto os testes com codificação representam o algoritmo em sua aplicação prática.

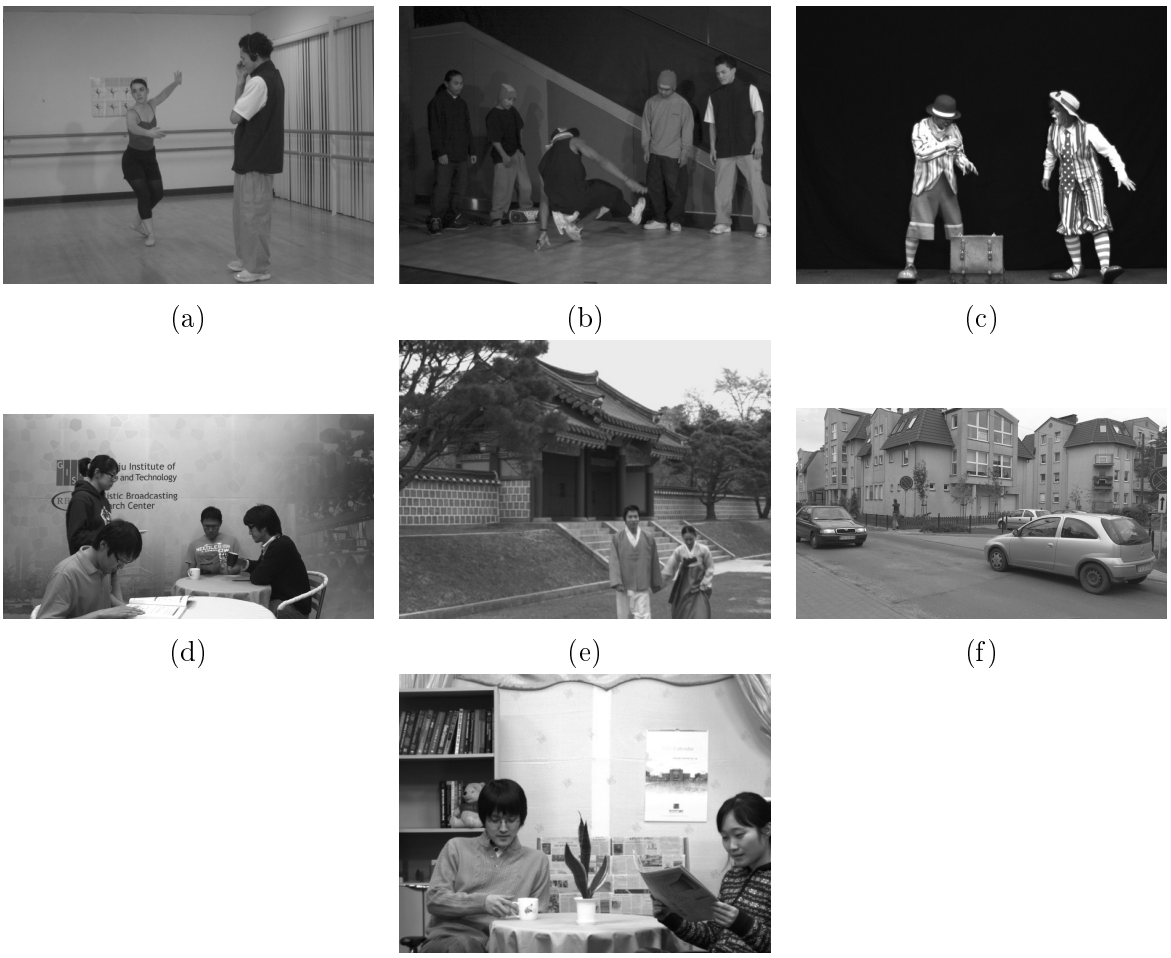


Figura 4.3: Quadros de exemplo das sequências reais testadas: (a) *Ballet*; (b) *Breakdancers*; (c) *Pantomime*; (d) *Cafe*; (e) *Lovebird1*; (f) *Poznan Street*; (g) *Newspaper*.

As Figs. 4.3 e 4.4 apresentam quadros de exemplo das sequências reais e sintéticas testadas, respectivamente, e a Tabela 4.1 detalha as informações destas sequências. A escolha das vistas foi feita de acordo com a disponibilidade de mapas de profundidade correspondentes, de forma a viabilizar os testes das técnicas propostas nos próximos Capítulos. A numeração das vistas e a quantidade de quadros disponíveis foram previamente determinadas pelas fontes das sequências [2] [19] [52] [59] [60].

É importante notar que algumas das sequências não foram testadas em suas resoluções originais, e sim em resoluções menores, por dois motivos. Algumas das sequências reais não apresentavam componentes de alta frequência suficientes para os testes, sendo portanto decimadas por 2. Por exemplo, a PSNR entre $\mathbf{I}_O$ e $\mathbf{I}_O^B$ para a sequência *Newspaper* [59] em sua resolução original, vista 4, quadro 1, é de $40,85$ dB. Repetindo este teste em metade da resolução original, a PSNR entre $\mathbf{I}_O$ e $\mathbf{I}_O^B$ é de $33,62$ dB para o mesmo quadro. A Fig. 4.5 apresenta um detalhe de cada um destes quadros, onde percebe-se que, na resolução original, a diferença subjetiva entre $\mathbf{I}_O$ e $\mathbf{I}_O^B$ é muito pequena, enquanto na metade da resolução, a diferença subjetiva entre $\mathbf{I}_O$ e $\mathbf{I}_O^B$ é maior.

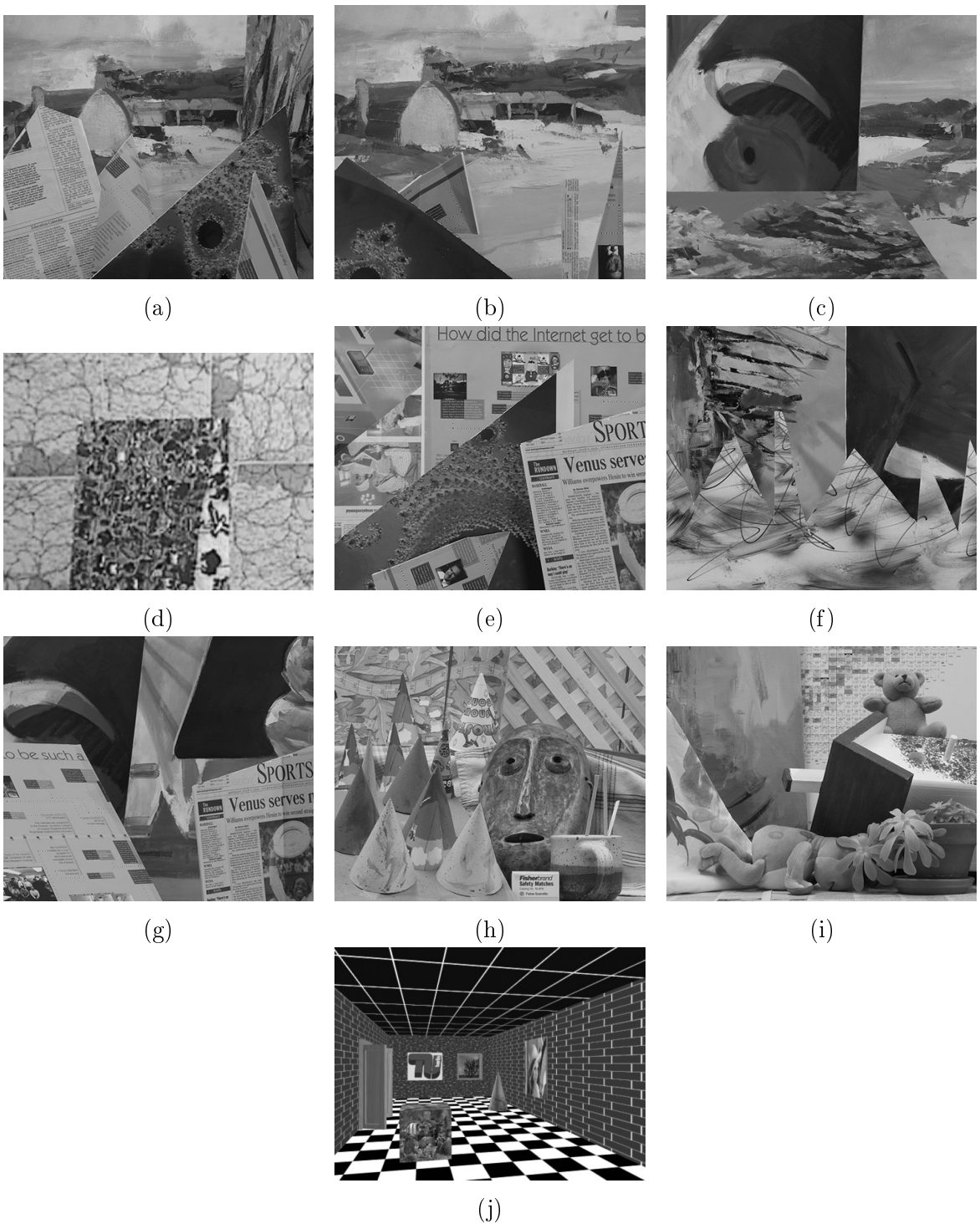


Figura 4.4: Quadros de exemplo das sequências sintéticas testadas: (a) *Barn1*; (b) *Barn2*; (c) *Bull*; (d) *Map*; (e) *Poster*; (f) *Sawtooth*; (g) *Venus*; (h) *Cones*; (i) *Teddy*; (j) *Room3D*.

Outras sequências não foram interpoladas, mas tiveram algumas de suas linhas e/ou colunas eliminadas, caso a largura e/ou comprimento não fossem múltiplos de 16, que é o tamanho do macrobloco utilizado na codificação H.264/AVC. Caso contrário, as sequências teriam linhas e/ou colunas acrescentadas artificialmente, aumentando a quantidade de dados a serem codificados.

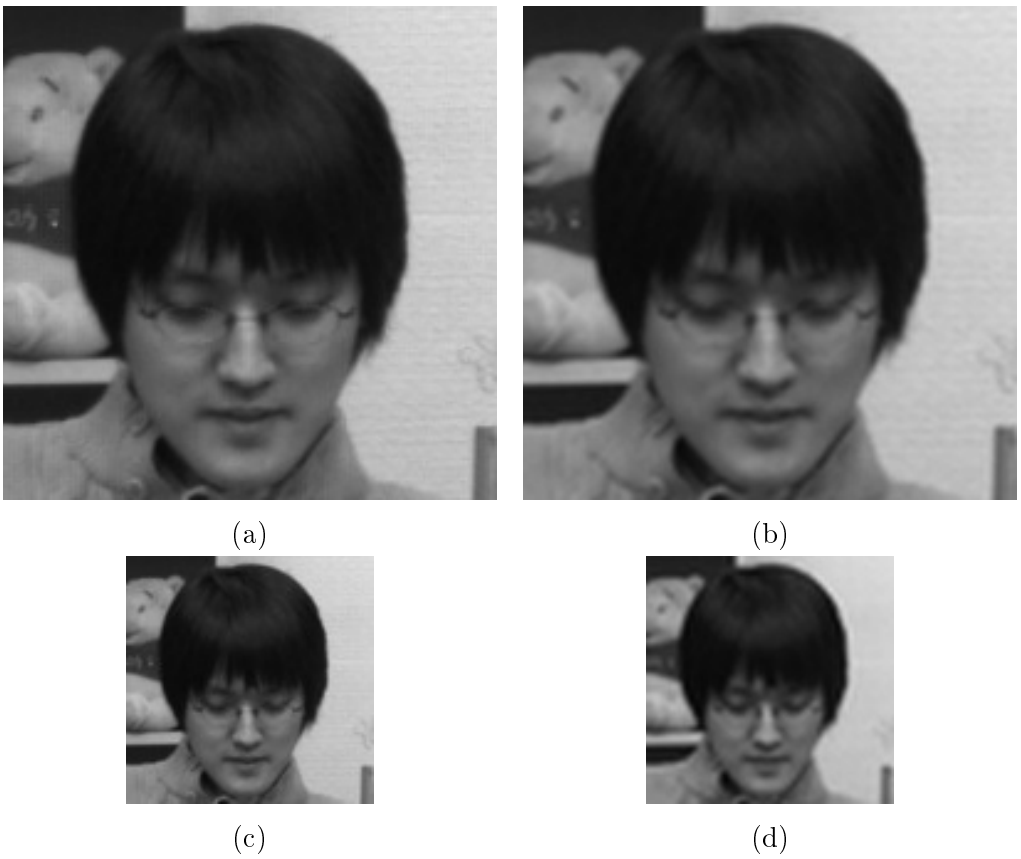


Figura 4.5: Detalhes da sequência *Newspaper*, vista 4, quadro 1, apresentando pouca quantidade de altas frequências: (a) $\mathbf{I}_O$ em resolução 1024×768; (b) $\mathbf{I}_O^B$ em resolução 1024×768 ($40,85$ dB); (c) $\mathbf{I}_O$ em resolução 512×384; (d) $\mathbf{I}_O^B$ em resolução 512×384 ($33,62$ dB).

4.4.1 Testes sem codificação H.264/AVC

A fim de comparar o desempenho do algoritmo proposto neste Capítulo em relação à interpolação de quadros em baixa resolução, foram realizados testes sob as seguintes condições:

- O método foi empregado sobre os quadros originais em resolução mista, tal como na Fig. 4.1, mas sem codificação, e foram medidos os ganhos de qualidade dos quadros super-resolvidos em relação à interpolação dos mesmos, com base nas médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros interpolados e super-resolvidos.

- Apesar de a vista considerada para os testes em cada sequência possuir quadros em resolução mista, os resultados objetivos foram calculados pela média dos quadros em baixa resolução. Se a PSNR dos quadros em alta resolução fosse considerada nos cálculos, cada um deles forneceria um erro quadrático médio nulo (Eq. 2.17), e um valor infinito de PSNR, consequentemente.

- A luminância foi escolhida para os testes, por se tratar da componente à qual o sistema visual humano possui maior resolução, como apresentado na Seção 2.2.

Tabela 4.1: Sequências utilizadas

Nome	**Resolução original**	**Resolução testada**	$\mathbf{I}_O$ **{vista}, {quadros}**	$\mathbf{I}_R$ **{vistas}, {quadros}**
Sequências reais				
Ballet [19]	1024×768	512×384	$\{1\}, \{0-99\}$	$\{0,1,2\}, \{0-99\}$
Breakdancers [19]	1024×768	512×384	$\{1\}, \{0-99\}$	$\{0,1,2\}, \{0-99\}$
Cafe [59]	1920×1080	960×528	$\{3\}, \{0-99\}$	$\{2,3,4\}, \{0-99\}$
Pantomime [60]	1280×960	640×480	$\{39\}, \{0-99\}$	$\{37,39,41\}, \{0-99\}$
Lovebird [59]	1024×768	512×384	$\{6\}, \{0-99\}$	$\{4,6,8\}, \{0-99\}$
Newspaper [59]	1024×768	512×384	$\{4\}, \{0-99\}$	$\{2,4,6\}, \{0-99\}$
Poznan Street [52]	1920×1088	960×544	$\{4\}, \{0-99\}$	$\{3,4,5\}, \{0-99\}$
Sequências sintéticas				
Barn1 [2]	432×381	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Barn2 [2]	430×381	416×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Bull [2]	433×381	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Map [2]	284×216	272×208	$\{1\}, \{0\}$	$\{0\}, \{0\}$
Poster [2]	435×383	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Sawtooth [2]	434×380	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Venus [2]	434×383	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Cones [2]	450×375	448×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Teddy [2]	450×375	448×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Room3D [2]	480×360	480×360	$\{2\}, \{0-99\}$	$\{1,2\}, \{0-99\}$

- Foram utilizados blocos de tamanho 16×16 para os processos de estimação/compensação de movimento e de combinação de altas frequências, a fim de acompanhar os macroblocos do padrão H.264/AVC, empregados nos testes com codificação. Aplicou-se também uma janela de busca de tamanho 80×80, que capta os deslocamentos presentes nas sequências de teste de forma fidedigna.

- Os ganhos obtidos pela super-resolução proposta em relação à interpolação são apresentados no Anexo I, Tabelas I.1 a I.4, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$).

- As Tabelas também apresentam os resultados obtidos empregando o método proposto com cada uma das referências individualmente (resultados 1 a 13 para as sequências reais, e 1 a 3 para as sinteticas), a fim de avaliar o efeito de se reduzir a quantidade de referências disponíveis.

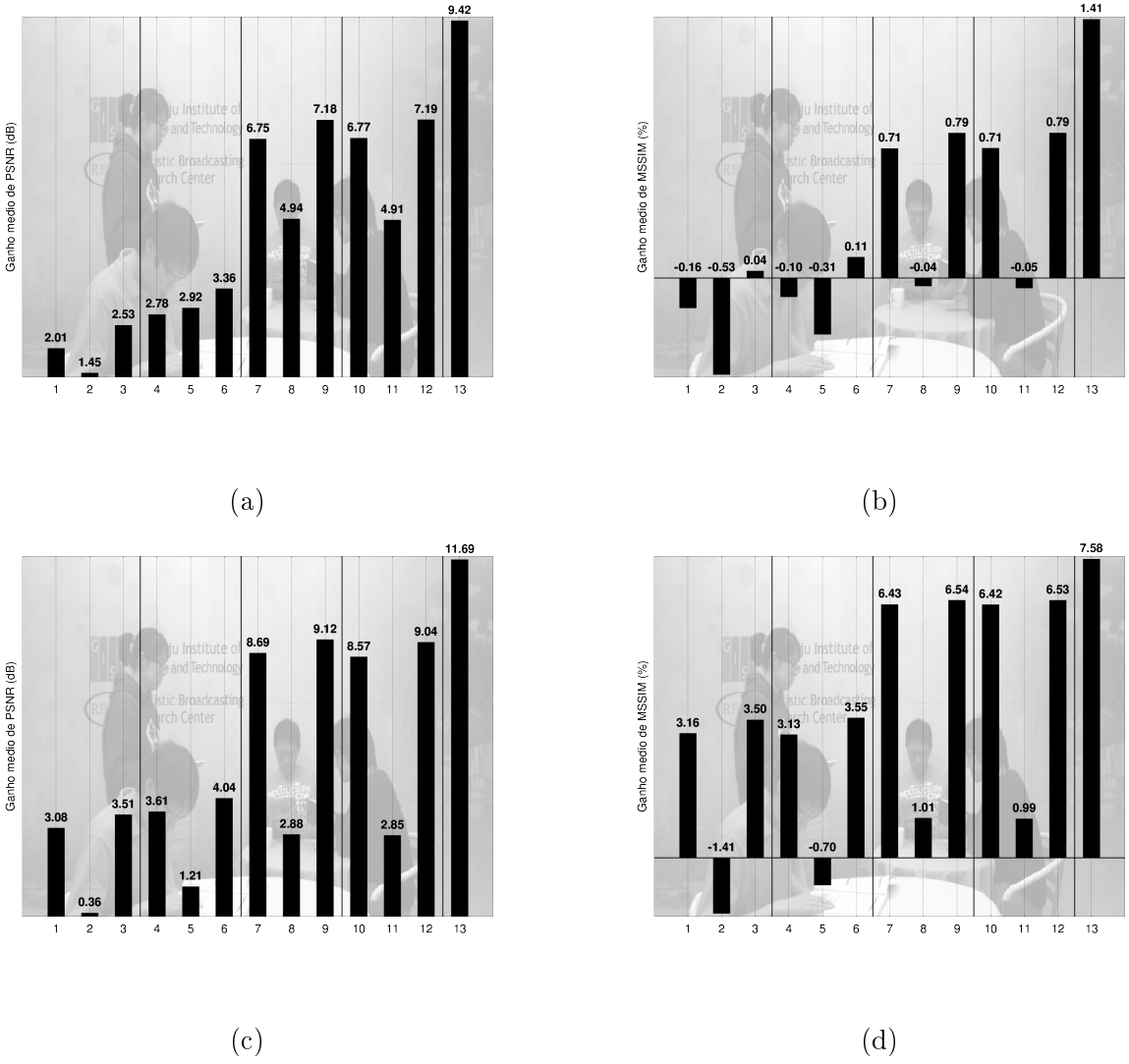


Figura 4.6: Resultados sem codificação para a componente de luminância da sequência *Cafe*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 13$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_{O}^{B}$; (2) $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_{O}^{B}$; (3) $\hat{\mathbf{I}}_{O1}$; (4) $\mathbf{I}_{O2}^{A}{}' + \mathbf{I}_{O}^{B}$; (5) $\mathbf{I}_{O2}^{A}{}'' + \mathbf{I}_{O}^{B}$; (6) $\hat{\mathbf{I}}_{O2}$; (7) $\mathbf{I}_{O3}^{A}{}' + \mathbf{I}_{O}^{B}$; (8) $\mathbf{I}_{O3}^{A}{}'' + \mathbf{I}_{O}^{B}$; (9) $\hat{\mathbf{I}}_{O3}$; (10) $\mathbf{I}_{O4}^{A}{}' + \mathbf{I}_{O}^{B}$; (11) $\mathbf{I}_{O4}^{A}{}'' + \mathbf{I}_{O}^{B}$; (12) $\hat{\mathbf{I}}_{O4}$; (13) $\hat{\mathbf{I}}_{O}$.

Os resultados obtidos indicam que a combinação de todas as referências (resultado 13 para as sequências reais, resultado 3 paras as sequências sintéticas) propicia os maiores ganhos em PSNR e MSSIM. Verificou-se apenas um caso em que a super-resolução proposta não ofereceu ganho de qualidade, a sequência *Cones*. Para $M = 2$, houve um pequeno ganho de PSNR, e uma pequena perda de MSSIM.

Para as sequências reais, constata-se que os quadros de referência pertencentes à mesma vista que a vista interpolada ($O3$ e $O4$ - resultados 7 a 12) fornecem as melhores vistas super-resolvidas

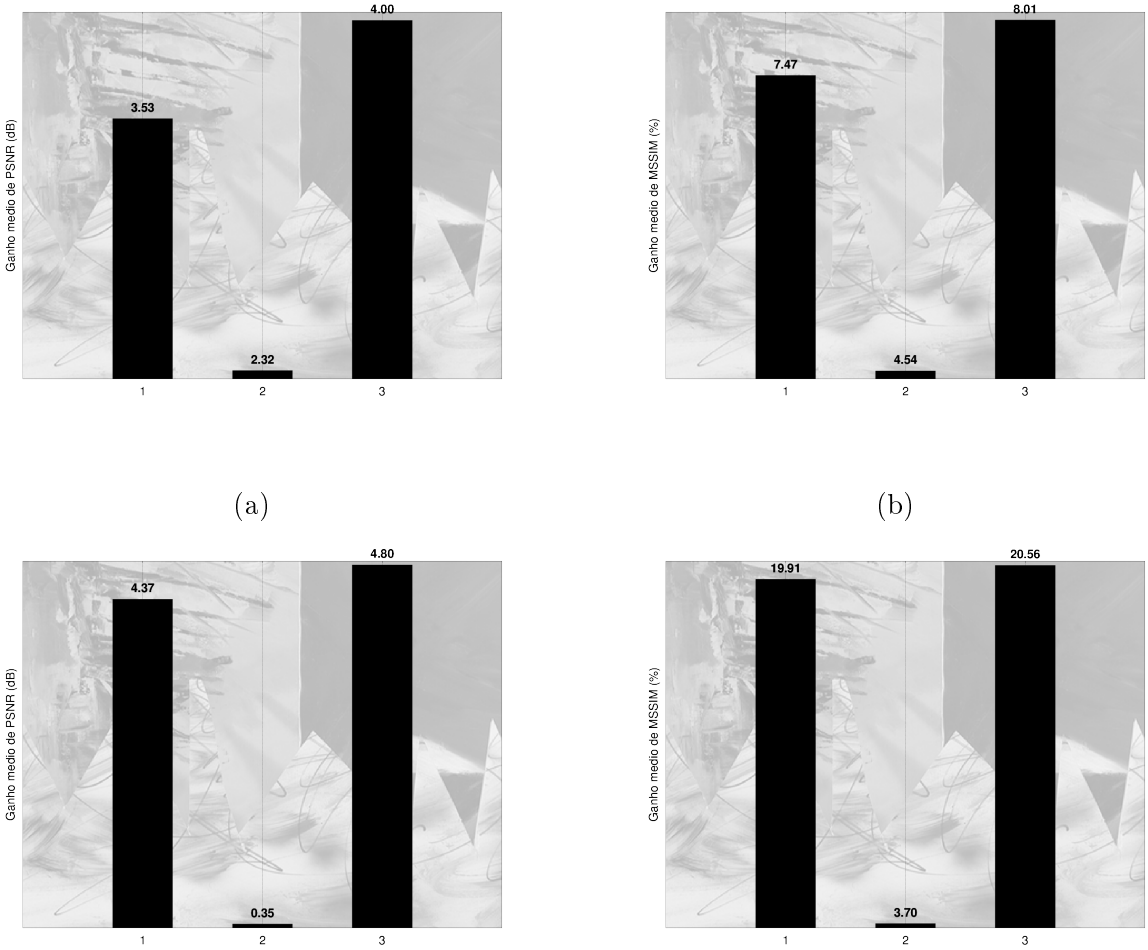


Figura 4.7: Resultados sem codificação para a componente de luminância da sequência *Sawtooth*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 3$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$; (2) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_O$.

na maior parte dos casos, quando se utiliza somente um quadro como referência. As exceções são as sequências *Pantomime*, para $M = 2$ e $M = 4$, e *Breakdancers*, para $M = 4$. Além disso, tanto para as sequências reais como as sintéticas, percebe-se que $\mathbf{I}_{Oi}^{A\,\prime} + \mathbf{I}_O^B$ (resultados 1, 4, 7 e 10) geralmente possui melhor qualidade do que $\mathbf{I}_{Oi}^{A\,\prime\prime} + \mathbf{I}_O^B$, (resultados 2, 5, 8 e 11). Ja á combinação de $\mathbf{I}_{Oi}^{A\,\prime}$ e $\mathbf{I}_{Oi}^{A\,\prime\prime}$ (resultados 3, 6, 9 e 12) oferece melhora de qualidade em relação a estes quadros individualmente, com exceção da sequência *Newspaper*, $M = 2$, referência $O1$. Ou seja, se houver disponibilidade de somente um quadro de referência, obtem-se quadros de melhor qualidade combinando $\mathbf{I}_{Oi}^{A\,\prime}$ e $\mathbf{I}_{Oi}^{A\,\prime\prime}$.

As Figs. 4.6 e 4.7 apresentam os ganhos médios obtidos pela super-resolução para as sequências *Cafe* e *Sawtooth*, respectivamente. Estas sequências ilustram o comportamento típico obtido para as demais sequências.

A Fig. 4.8 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Cafe*, $M = 2$. É possível perceber a diferença de qualidade entre as imagens decimada e interpolada e original, Figs. 4.8(a) e (f), respectivamente, pela definição das letras escritas na parede da cena e na camisa do rapaz. Na Fig. 4.8(b), $\mathbf{I}_{O3}^{A}{}' + \mathbf{I}_{O}^{B}$ apresenta melhora considerável, como nos detalhes das letras escritas na parede da cena, mas também possui erros, como o ponto acrescentado ao topo da letra 's' na parede. Na Fig. 4.8(c), $\mathbf{I}_{O3}^{A}{}'' + \mathbf{I}_{O}^{B}$ reduz estes erros consideravelmente, e na Fig. 4.8(d), a combinação de $\mathbf{I}_{O3}^{A}{}'$ e $\mathbf{I}_{O3}^{A}{}''$ reintroduz um pouco destes erros. Já na Fig. 4.8(e), o quadro $\hat{\mathbf{I}}_{O}$ reduz este erro, e elimina um erro surgido na letra 'r' na parede.

A Fig. 4.9 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Teddy*, $M = 2$. Novamente, a imagem $\mathbf{I}_{O}^{B}$ é mais borrada do que $\mathbf{I}_{O}$, Figs. 4.9(a) e (e) respectivamente. As Figs. 4.9(b), (c) e (d) apresentam os resultados respectivos a $\mathbf{I}_{O}^{A\prime} + \mathbf{I}_{O}^{B}$, $\mathbf{I}_{O}^{A\prime\prime} + \mathbf{I}_{O}^{B}$ e $\hat{\mathbf{I}}_{O}$. $\mathbf{I}_{O}^{A\prime} + \mathbf{I}_{O}^{B}$ e $\mathbf{I}_{O}^{A\prime\prime} + \mathbf{I}_{O}^{B}$ apresentam artefatos de bloco sobre o vaso à direita da cena, e $\hat{\mathbf{I}}_{O}$ reduz tais artefatos, o que é refletido em um aumento em termos de PSNR e MSSIM.

A Fig. 4.10 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Newspaper*, $M = 4$. $\mathbf{I}_{O}^{B}$ é extremamente borrada, Fig. 4.10(a), e o quadro super-resolvido $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_{O}^{B}$ recupera grande parte dos detalhes da imagem, Fig. 4.10(b). $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_{O}^{B}$ apresenta artefatos de bloco sobre a ponta da planta à direita da cena e sobre o urso de pelúcia, que levam a grande queda em qualidade, Fig. 4.10(c), e $\hat{\mathbf{I}}_{O1}$ e $\hat{\mathbf{I}}_{O}$ apresentam bastante detalhes da imagem original $\mathbf{I}_{O}$, Figs. 4.10(d) e (e), respectivamente.

A Fig. 4.11 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Cones*, $M = 2$. Esta foi a única sequência em que a super-resolução não melhorou a qualidade da imagem em termos de MSSIM, claramente devido à presença de artefatos de bloco próximos às bordas dos cones da cena, e devido à falta de definição no texto sobre a caixa na região inferior da cena.

4.4.2 Testes com codificação H.264/AVC

Os resultados apresentados na Seção anterior representam o limite superior de qualidade atingido pelo algoritmo de super-resolução apresentado neste Capítulo. A fim de avaliar seu desempenho em uma situação prática, o algoritmo foi aplicado às mesmas sequências codificadas em resolução mista, dentro das seguintes condições:

- Cada sequência foi codificada separadamente utilizando o padrão H.264/AVC em modo Intra, em resolução mista. O modo de predição Intra foi utilizado como prova de conceito, visto que o algoritmo proposto não faz nenhuma assunção sobre o modo de predição utilizado para as vistas disponíveis.

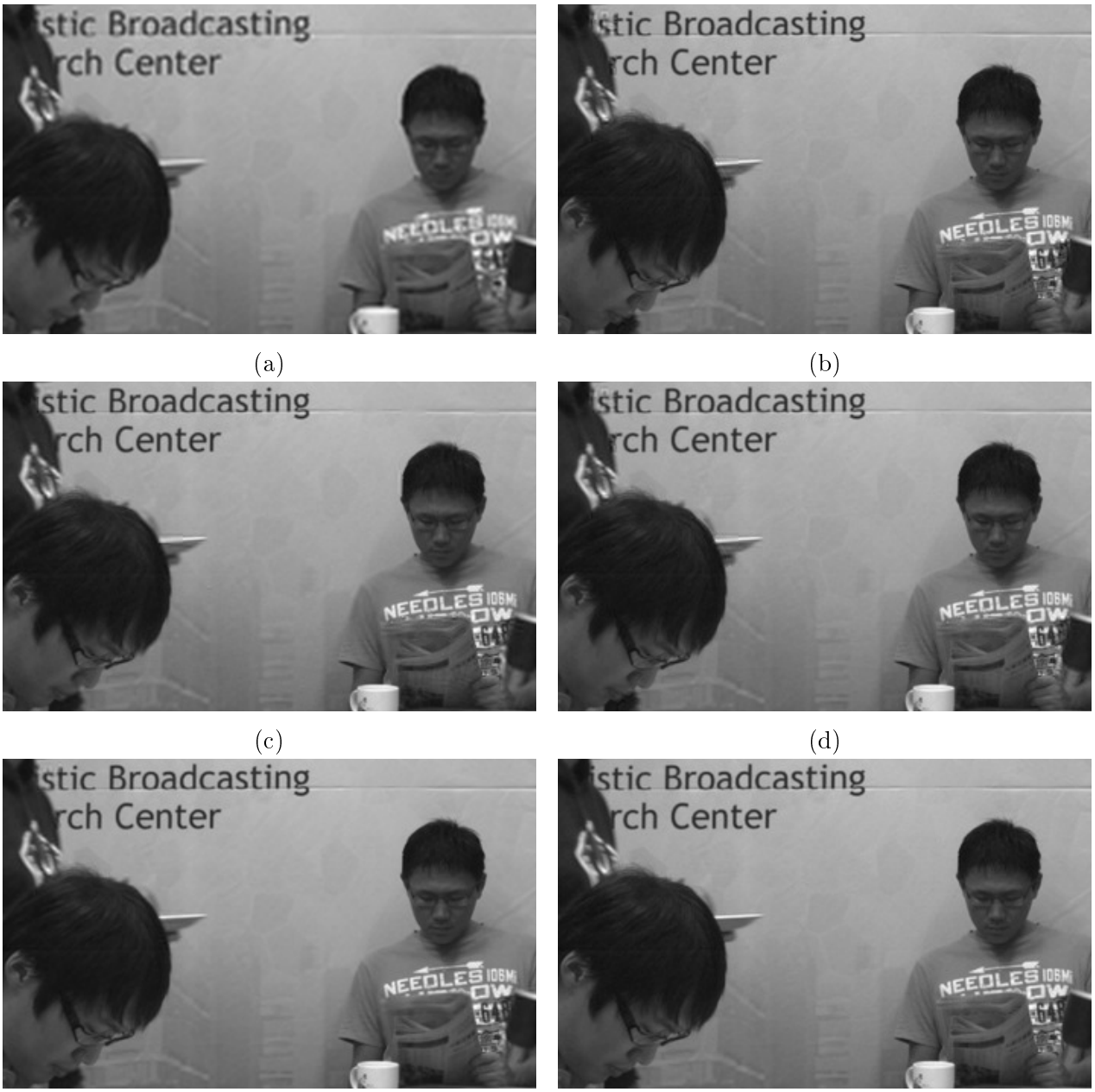


Figura 4.8: Detalhes da sequência *Cafe*, vista 3, quadro 2, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 36,53 dB / MSSIM×100 = 97,36%); (b) $\mathbf{I}_{O3}^A{}' + \mathbf{I}_O^B$ (43,12 dB / 98,02%); (c) $\mathbf{I}_{O3}^A{}'' + \mathbf{I}_O^B$ (42,13 dB / 97,30%); (d) $\hat{\mathbf{I}}_{O3}$ (43,67 dB / 98,11%); (e) $\hat{\mathbf{I}}_O$ (46,01 dB / 98,72%); (f) $\mathbf{I}_O$.

- A codificação foi feita de acordo com a Seção 4.2 e a Fig. 4.1, e seguindo a escolha de quadros e vistas da Tabela 4.1.

- Aplicou-se os valores $QP = \{22, 27, 32, 37\}$ aos parâmetros de escalonamento, pois estes correspondem a taxas típicas de transmissão e armazenamento de sequências de video [31].

- O *software* de referência JM 17.2 para o padrão H.264/AVC foi utilizado [61].

- A luminância foi escolhida para os testes, e mediu-se a média quadro-a-quadro da PSNR e da MSSIM da vista em consideração, incluindo os quadros em resolução baixa e completa, já que todos foram corrompidos pelo processo de escalonamento na etapa de codificação.

- A taxa considerada foi a taxa da arquitetura inteira, a fim de refletir um cenário de *free-viewpoint television* em que o usuário seleciona assistir à vista em consideração, Tabela 4.1. Como a super-resolução proposta utiliza somente os quadros em resolução completa para super-resolver os quadros em baixa resolução, ela não precisa de nenhuma informação extra, de forma que a mesma taxa considerada é tanto para os quadros interpolados quanto para os super-resolvidos.

- Foram utilizados blocos de tamanho 16×16 para os processos de estimação/compensação de movimento e de combinação de altas frequências, e uma janela de busca de tamanho 80×80, a fim de repetir as condições dos testes feitos sem codificação.

- Foram medidos os ganhos médios [62] de qualidade (PSNR e MSSIM) para o algoritmo proposto em relação a interpolar os quadros em baixa resolução, utilizando tanto cada referência individualmente quanto todas combinadas.

- Os ganhos médios do algoritmo proposto sobre à interpolação são apresentados no Anexo I, Tabelas I.5 a I.8, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$).

Assim como no caso sem codificação, a combinação de todas as referências propicia os maiores ganhos em termos de PSNR e em termos de MSSIM. Novamente, todas as sequências apresentaram ganho de qualidade, exceto para a sequência *Cones*, na qual verificou-se uma pequena perda média de PSNR para $M = 4$.

Em relação ao uso de somente um quadro como referência, os quadros $\hat{\mathbf{I}}_{O3}$ e $\hat{\mathbf{I}}_{O4}$ (resultados 9 e 12, respectivamente), pertencentes à vista sendo super-resolvida, forneceram os melhores na maior parte dos casos, exceto para *Breakdancers*, $M = 4$, e *Pantomime*, $M = 2$ e $M = 4$. Além disso, para todas as sequências reais e sintéticas, $\mathbf{I}_{Oi}^{A\,\prime} + \mathbf{I}_{O}^{B}$ sempre possui melhor qualidade do que $\mathbf{I}_{Oi}^{A\,\prime\prime} + \mathbf{I}_{O}^{B}$, e a combinação de ambos oferece melhora de qualidade em relação a estes quadros individualmente. A única exceção é a sequência *Newspaper*, $M = 2$, resultados 1 e 2,

A Fig. 4.12 apresenta o desempenho da super-resolução proposta em termos de taxa e distorção para a sequência *Poznan Street*, que ilustra o comportamento típico do desempenho para as sequências testadas. De acordo com o aumento da taxa, melhora a qualidade das imagens interpoladas e super-resolvidas, bem como o desempenho do algoritmo proposto. Isso acontece devido à grande perda de altas frequências à medida que o escalonamento é acentuado, afetando a qualidade das referências usadas pelo processo de super-resolução. Para altas taxas, as referências dispõem de altas frequências mais fidedignas, melhorando o resultado da super-resolução.

Caso fosse utilizado um outro modo de predição na codificação, tais como a predição Inter ou o MVC, haveria uma grande probabilidade de que as referências seriam corrompidas por menor

ruído de escalonamento, visto que estas formas de predição são mais precisas do que a predição Intra. Sendo assim, o algoritmo apresentaria desempenho superior ao apresentada nesta Seção, visto que as referências ofereceriam altas frequências mais fidedignas.

É importante ressaltar que não são iguais os comportamentos das curvas de taxa-distorção baseadas em PSNR e MSSIM. Por exemplo, a diferença entre as versões super-resolvida e interpolada não são iguais nas Figs. 4.12(a) e (b). Em alguns casos, é possível que a melhora objetiva proporcionada pela super-resolução não seja tão grande em termos subjetivos, e vice-versa. Por exemplo, a sequência *Ballet* apresenta o maior ganho médio em termos de PSNR, $2,55$ dB, e a *Poznan Street* apresenta o maior ganho médio em termos de MSSIM, $2,11\%$.

A Fig. 4.13 ilustra o desempenho da sequência *Cones*, que representa a única perda de qualidade dentre todas as sequências testadas. É possível ver que existe um pequeno ganho de PSNR para $M = 2$, com uma média de $0,32$ dB, e um ganho de MSSIM, com uma média de 0,24%. Para $M = 4$, o algoritmo proposto só apresenta ganho de MSSIM para altas taxas.

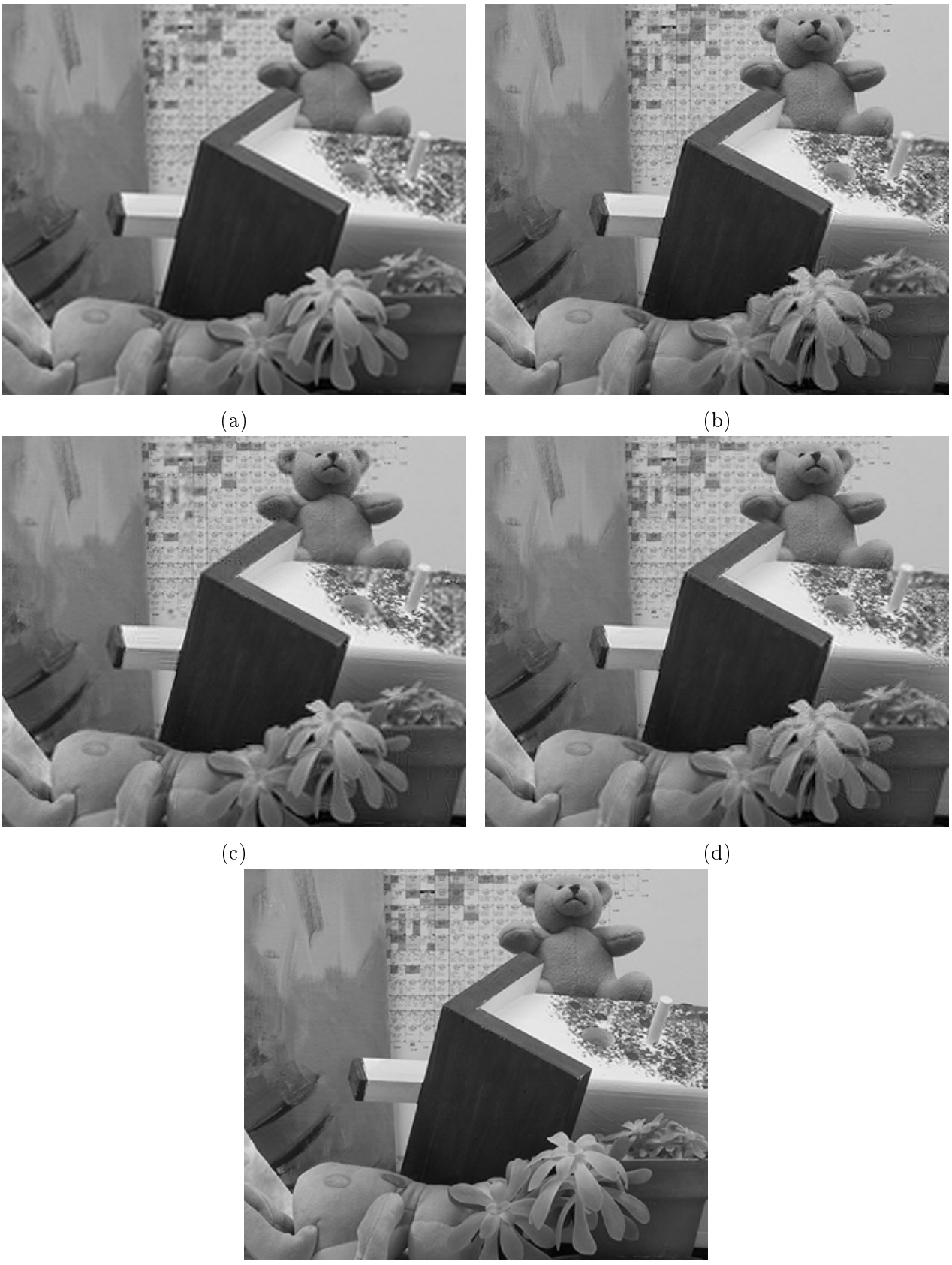


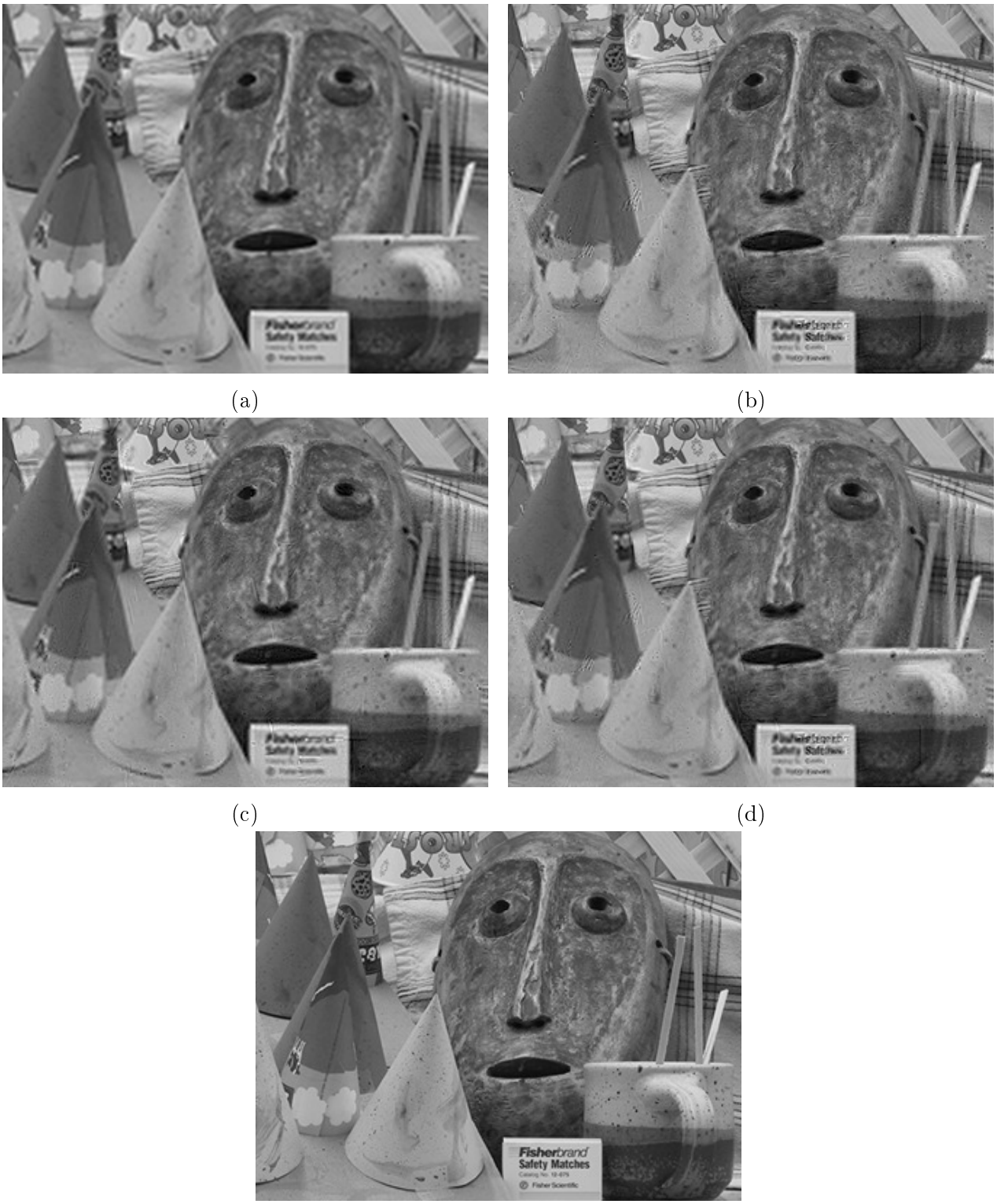
Figura 4.9: Detalhes da sequência *Teddy*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 29,69 dB / MSSIM×100 = 89,19%); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ (30,33 dB / 90,70%); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ (30,20 dB / 88,72%); (d) $\hat{\mathbf{I}}_O$ (31,00 dB / 91,69%); (e) $\mathbf{I}_O$.

(a) (b)

(c) (d)

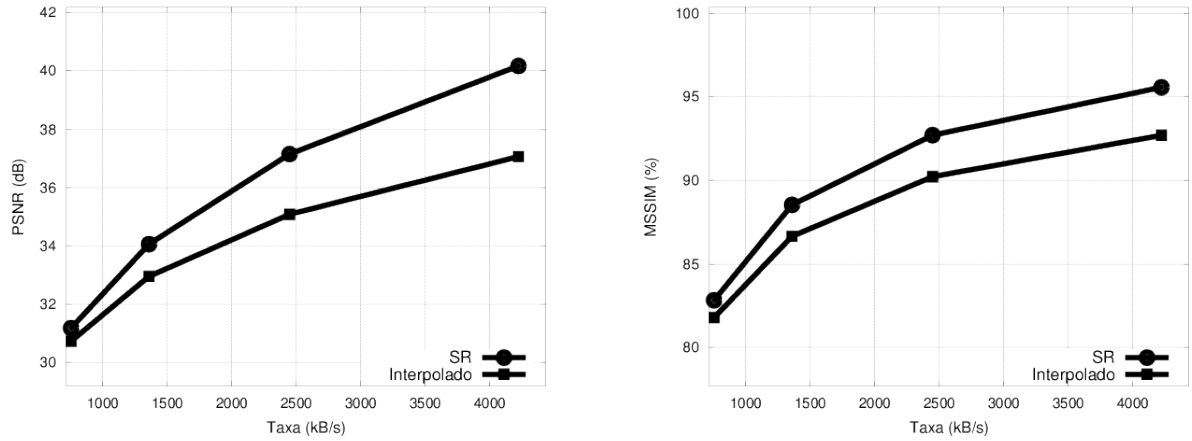
(e) (f)

Figura 4.10: Detalhes da sequência *Newspaper*, vista 4, quadro 2, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR = 27,19 dB / MSSIM×100 = 81,46%); (b) $\mathbf{I}_{O1}^A{}' + \mathbf{I}_O^B$ (37,47 dB / 97,09%); (c) $\mathbf{I}_{O1}^A{}' + \mathbf{I}_O^B$ (29,93 dB / 88,11%); (d) $\hat{\mathbf{I}}_{O1}$ (37,58 dB / 97,16%); (e) $\hat{\mathbf{I}}_O$ (40,48 dB / 98,60%); (f) $\mathbf{I}_O$.



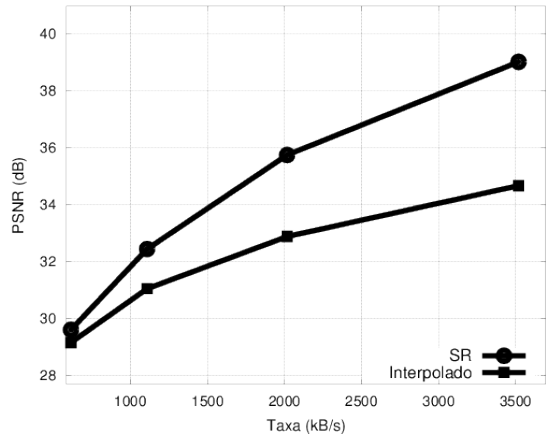
(a) (b) (c) (d) (e)

Figura 4.11: Detalhes da sequência *Cones*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 28,41 dB / MSSIM×100 = 84,76%); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ (28,23 dB / 82,76%); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ (28,67 dB / 84,27%); (d) $\hat{\mathbf{I}}_O$ (28,83 dB / 84,64%); (e) $\mathbf{I}_O$.

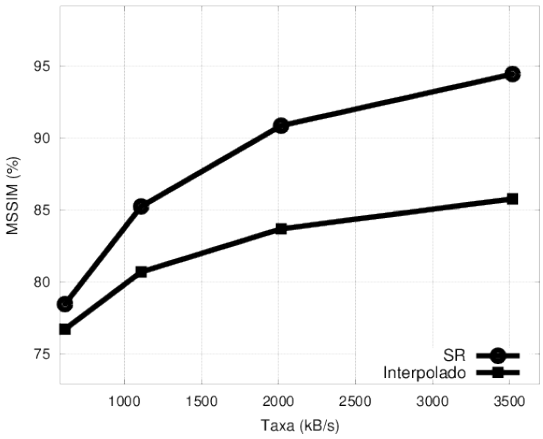


(a)

(b)

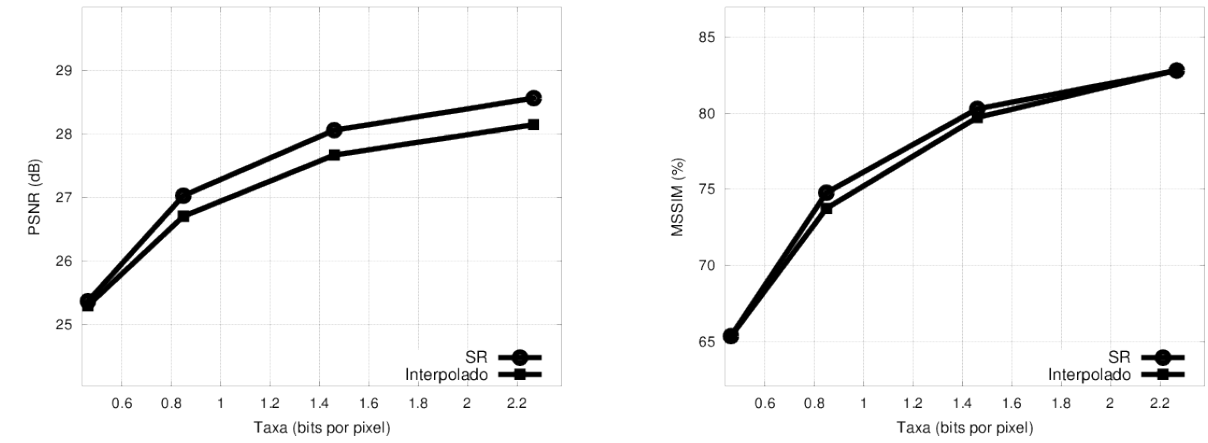


(c)



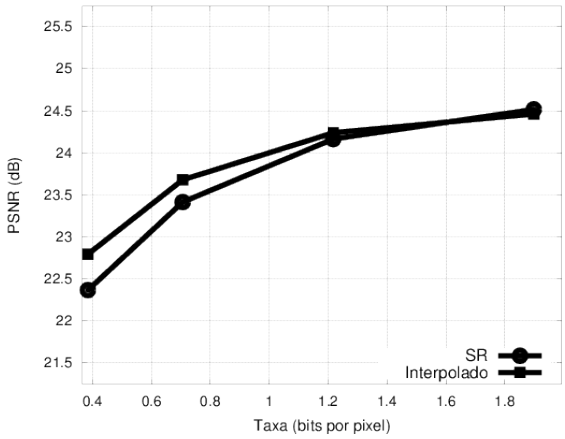
(d)

Figura 4.12: Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Poznan Street*, vista 4: (a) taxa e PSNR, $M = 2$ (ganho médio de $1,61$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $2,11\%$); (c) taxa e PSNR, $M = 4$ (ganho médio de $2,16$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $5,66\%$).

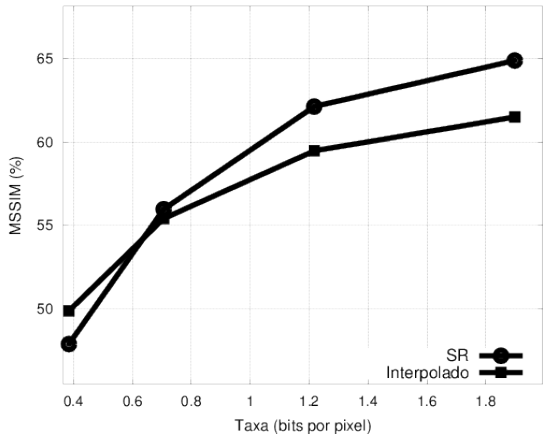


(a)

(b)



(c)



(d)

Figura 4.13: Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Cones*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$ (ganho médio de $0,32$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $0,64\%$); (c) taxa e PSNR, $M = 4$ (perda média de $0,19$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $1,15\%$).

Capítulo 5

Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade

5.1 Introdução

Este Capítulo apresenta a segunda das técnicas propostas para sequências de múltiplas vistas em resolução mista. Assim como no Capítulo anterior, apresenta-se a arquitetura em consideração, e define-se a técnica proposta na Seção seguinte. Resultados experimentais são então apresentados, levando em conta sequências sem perdas e sequências codificadas pelo padrão H.264/AVC.

5.2 Arquitetura em consideração

Nesta arquitetura de codificação, considera-se que algumas vistas são codificadas em resolução normal, e outras, em baixa resolução, como ilustrado na Fig. 5.1 para quatro vistas. Além disso, são transmitidos os mapas de profundidade correspondentes para todas as vistas, em suas resoluções originais. Assim como na arquitetura ilustrada na Fig. 4.1, as vistas em baixa resolução deverão ser interpoladas de volta às suas resoluções originais, e o usuário verá um par de vistas em qualidade mista, já que a vista interpolada parecerá mais borrada do que a outra vista. Todavia, a vista borrada será sempre a mesma, ao invés de alternada entre os olhos direito e esquerdo do usuário ao longo do tempo. Esta arquitetura também pode ser aplicada a sequências estereoscópicas, onde seriam consideradas somente as vistas 0 e 1 para codificação, usando a Fig. 5.1 como referência.

A arquitetura presente destina-se a sistemas de transmissão que devem atender a diversos tipos de receptores, incluindo telas autoestereoscópicas e *free-viewpoint television*. O formato em resolução mista seria aplicado para fins de economia de taxa de transmissão, apoiado na teoria da supressão binocular (Seção 3.3.3). A técnica de super-resolução apresentada neste Capítulo seria empregada a fim de adequar a sequência transmitida a aplicações em *free-viewpoint television*.

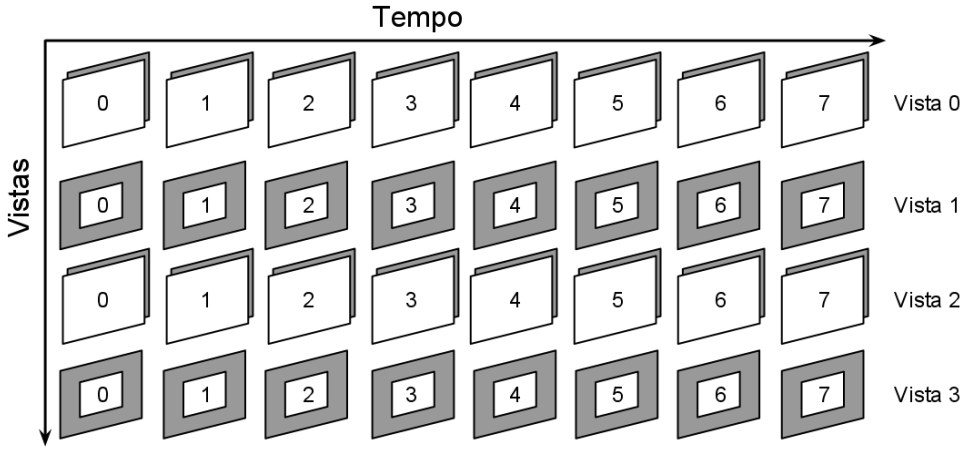


Figura 5.1: Arquitetura de codificação de múltiplas vistas em resolução mista alternada com mapas de profundidade.

5.3 Solução proposta

A recuperação de informações de alta frequência nas vistas em baixa resolução também é inspirada na super-resolução baseada em exemplos (Seção 2.4.2), extraindo informações de alta frequência a partir dos quadros adjacentes em alta resolução [17] [18].

Para cada quadro em baixa resolução, haverá um ou dois quadros adjacentes de referência, que pertencem a outra vista. Usando a Fig. 5.1 como referência, o quadro 0 da vista 1 pode ser super-resolvido pelos quadros 0 das vistas 0 e 2.

O método proposto é dividido em etapas de extração e combinação de altas frequências. A seguir, define-se cada uma destas etapas.

5.3.1 Extração de altas frequências

Assim como na Seção 4.3.1, tem-se um quadro decimado na j-ésima vista e k-ésimo instante, $\mathbf{I}^D_{j,k} = \mathbf{I}^D_O$, que é interpolado à resolução original, $\mathbf{I}^B_{j,k} = \mathbf{I}^B_O$, constituindo a versão de baixa frequência do quadro original $\mathbf{I}_{j,k} = \mathbf{I}_O$. O algoritmo proposto neste Capítulo gera uma versão super-resolvida $\hat{\mathbf{I}}_O$, a partir da estimativa $\hat{\mathbf{I}}^A_O$ da alta frequência $\mathbf{I}^A_O$, como na Eq. (4.1).

Estão disponíveis o quadro $\mathbf{I}_R$, no mesmo instante k e na vista adjacente $j-1$ e/ou $j+1$, e os mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$ dos quadros original e de referência, respectivamente, bem como suas respectivas matrizes de projeção $\mathbf{P}_O$ e $\mathbf{P}_R$. Seguindo as Eqs. (3.4) a (3.7), é possível obter correspondências *pixel* a *pixel* entre os quadros original e de referência. Sendo assim, a vista O é reconstruída a partir do quadro $\mathbf{I}_R$, gerando o quadro $\mathbf{I}'_{O|R}$.

Na projeção da vista R para O, existem duas considerações a serem feitas. Em primeiro lugar, os mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$ são suscetíveis a erros, devidos a oclusões, imprecisões de

representação e escalonamento, no caso de codificação com perdas, como visto na Seção 3.2.2. Além disso, as correspondências indicadas pelas Eqs. (3.4) e (3.5) indicam posições de *pixels* em ponto fixo, ao invés de valores correspondentes a posições inteiras. Estas duas questões devem ser tratadas a fim de gerar o quadro $\mathbf{I}'_{O|R}$.

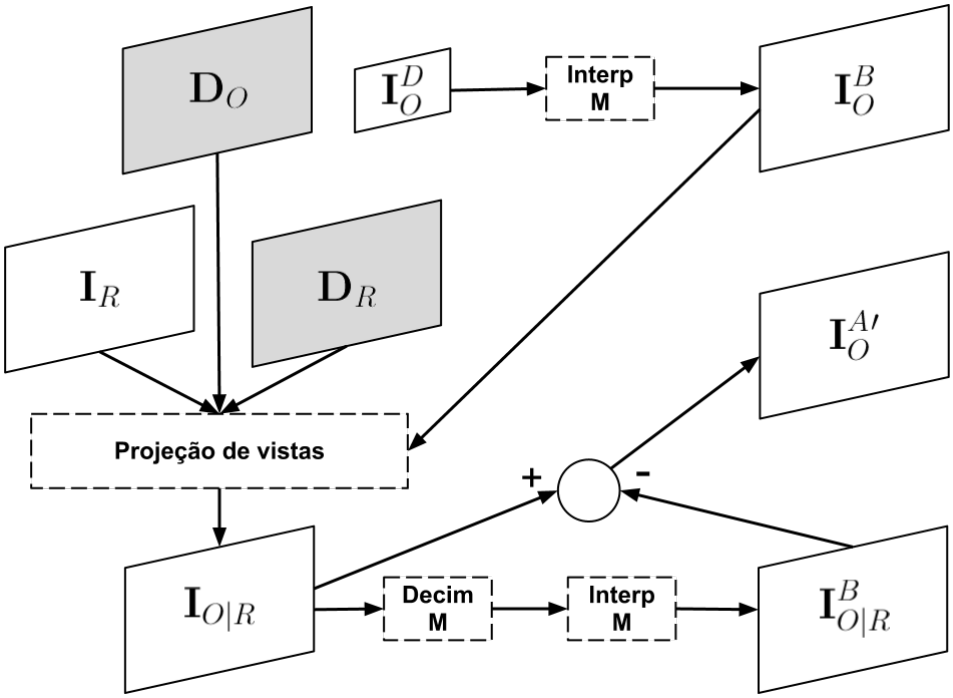


Figura 5.2: Extração de altas frequências para sequências em múltiplas vistas. Uma estimativa de alta frequência $\mathbf{I}_O^{A\prime}$ para a vista O é gerada a partir de uma vista adjacente em alta resolução $\mathbf{I}_R$, e dos mapas de profundidade correspondentes, $\mathbf{D}_O$ e $\mathbf{D}_R$.

A fim de tratar os erros dos mapas de profundidade, confere-se a consistência entre as informações de $\mathbf{D}_O$ e $\mathbf{D}_R$. Utilizando as Eqs. (3.4) e (3.5), o ponto $\mathbf{m}_O = [u, v, 1]^T$ em $\mathbf{D}_O$ é mapeado para o ponto $\mathbf{m}_R = [u', v', 1]^T$ em $\mathbf{D}_R$. Calcula-se então as posições inteiras $\mathbf{m}_{R1}$, $\mathbf{m}_{R2}$, $\mathbf{m}_{R3}$ e $\mathbf{m}_{R4}$ mais próximas a $\mathbf{m}_R$:

$$\begin{aligned}
\mathbf{m}_{R1} &= [u'_a, v'_a, 1]^T = [\lfloor u' \rfloor, \lfloor v' \rfloor, 1]^T \\
\mathbf{m}_{R2} &= [u'_b, v'_a, 1]^T = [\lceil u' \rceil, \lfloor v' \rfloor, 1]^T \\
\mathbf{m}_{R3} &= [u'_a, v'_b, 1]^T = [\lfloor u' \rfloor, \lceil v' \rceil, 1]^T \\
\mathbf{m}_{R4} &= [u'_b, v'_b, 1]^T = [\lceil u' \rceil, \lceil v' \rceil, 1]^T
\end{aligned}, \tag{5.1}$$

onde $\lfloor q \rfloor$ é a parte inteira de q, e $\lceil q \rceil$ é a parte inteira de $q + 1$. Dentre todos os pontos $\mathbf{m}_{Ri}$, seleciona-se o mais próximo a $\mathbf{m}_R$, $[\lfloor u' + 0,5 \rfloor, \lfloor v' + 0,5 \rfloor, 1]^T$, e a partir deste, segue-se para o ponto $\mathbf{m}_{O|R} = [u'', v'', 1]^T$ em $\mathbf{D}_O$, baseado em $\mathbf{D}_R$. Se $\mathbf{m}_{O|R}$ cair fora de um raio de 1 *pixel* em volta de $\mathbf{m}_O$, $I'_{O|R}(u, v) = I_O^B(u, v)$. Ou seja, não se utiliza nenhuma projeção da vista adjacente, pois a informação em $D_O(\mathbf{m}_O)$ não foi consistente com $D_R(\mathbf{m}_R)$.

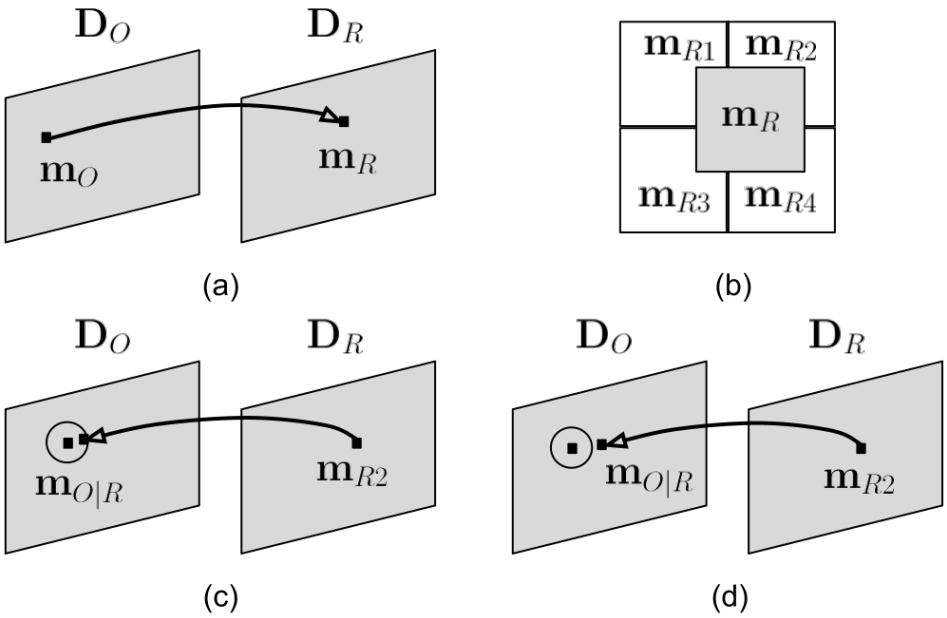


Figura 5.3: Teste de consistência entre os mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$: (a) projeção do ponto $\mathbf{m}_O$ para o ponto $\mathbf{m}_R$, utilizando as Eqs. (3.4) e (3.5); (b) posição em ponto fixo $\mathbf{m}_R$ no mapa $\mathbf{D}_R$, e posições inteiras mais próximas $\mathbf{m}_{R1}$, $\mathbf{m}_{R2}$, $\mathbf{m}_{R3}$ e $\mathbf{m}_{R4}$; (c) re-projeção de $\mathbf{m}_{R2}$ de $\mathbf{D}_R$ para $\mathbf{D}_O$, caindo dentro de um raio de 1 *pixel* e resultando em uma projeção válida; (d) re-projeção de $\mathbf{m}_{R2}$ de $\mathbf{D}_R$ para $\mathbf{D}_O$, caindo fora de um raio de 1 *pixel* e resultando em uma projeção inválida.

Se $\mathbf{m}_{O|R}$ cair dentro de um raio de 1 *pixel* em volta de $\mathbf{m}_O$, $\mathbf{m}_{R1}$, $\mathbf{m}_{R2}$, $\mathbf{m}_{R3}$ e $\mathbf{m}_{R4}$ são usados para calcular $I'_{O|R}(u,v)$ por interpolação bilinear:

$$I'_{O|R}(u,v) = \begin{bmatrix} u'_b - u' & u' - u'_a \end{bmatrix} \begin{bmatrix} I_R(u'_a, v'_a) & I_R(u'_a, v'_b) \\ I_R(u'_b, v'_a) & I_R(u'_b, v'_b) \end{bmatrix} \begin{bmatrix} v'_b - v' \\ v' - v'_a \end{bmatrix}. \tag{5.2}$$

A escolha do raio de 1 *pixel* foi feita de acordo com resultados empíricos.

A consistência entre os mapas $\mathbf{D}_O$ e $\mathbf{D}_R$ é também armazenada em uma imagem denominada $\mathbf{I}_{VAL}$. Se $\mathbf{m}_{O|R}$ cair fora de um raio de 1 *pixel* em volta de $\mathbf{m}_O$, $I'_{VAL}(u,v) = 0$; caso contrário, $I'_{VAL}(u,v) = 1$.

Seguindo a Eq. (4.2), o quadro $\mathbf{I}'_{O|R}$ é decomposto em versões de baixa e alta frequência, respectivamente $\mathbf{I}^B_{O|R}{}'$ e $\mathbf{I}^A_{O|R}{}'$. Esta última é considerada a primeira estimativa de alta frequência:

$$\mathbf{I}^{A\prime}_O = \mathbf{I}^A_{O|R}{}'. \tag{5.3}$$

A Fig. 5.2 ilustra o processo geral de extração de altas frequências, e a Fig. 5.3 ilustra a forma como a consistência entre os mapas $\mathbf{D}_O$ e $\mathbf{D}_R$ é conferida.

Dependendo da qualidade dos mapas $\mathbf{D}_O$ e $\mathbf{D}_R$, pode ser interessante erodir $\mathbf{I}'_{VAL}$. Isto pode ser útil para o caso em que a consistência entre os mapas reflita erros comuns aos dois mapas, o

que acontece com certa frequência na borda de objetos, por exemplo. Como ilustrado na Fig. 5.4, um novo mapa de correspondências válidas $\mathbf{I}''_{VAL}$ é criado da seguinte maneira:

$$I''_{VAL}(u,v) = \begin{cases} & I'_{VAL}(u,v-1) + I'_{VAL}(u-1,v) + \\ 1, & I'_{VAL}(u,v+1) + I'_{VAL}(u+1,v) + \\ & I'_{VAL}(u,v) = 5 \\ 0, & \text{caso contrário} \end{cases} . \tag{5.4}$$

A partir de $\mathbf{I}''_{VAL}$, cria-se uma nova projeção $\mathbf{I}''_{O|R}$:

$$I''_{O|R}(u,v) = \begin{cases} I'_{O|R}(u,v), & I''_{VAL}(u,v) = 1 \\ I^B_O(u,v), & I''_{VAL}(u,v) = 0 \end{cases} , \tag{5.5}$$

e a partir da Eq. (4.2), $\mathbf{I}''_{O|R}$ é decomposto em versões de baixa e alta frequência, respectivamente ${\mathbf{I}^B_{O|R}}''$ e ${\mathbf{I}^A_{O|R}}''$. Esta última é considerada a segunda estimativa de alta frequência, $\mathbf{I}^{A\prime\prime}_O$:

$$\mathbf{I}^{A\prime\prime}_O = {\mathbf{I}^A_{O|R}}''. \tag{5.6}$$

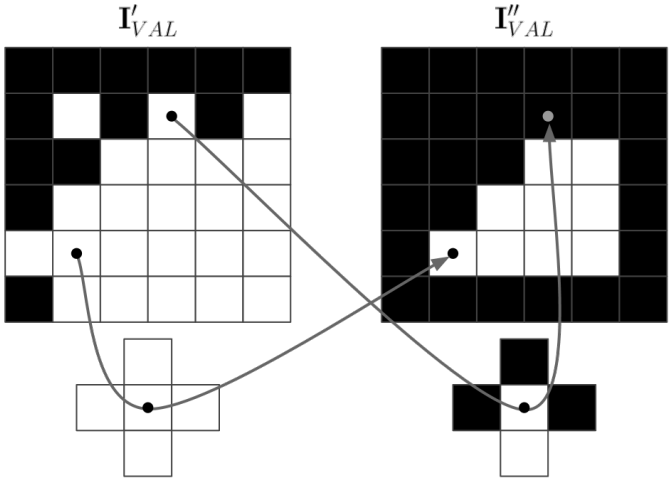


Figura 5.4: Erosão do mapa $\mathbf{I}'_{VAL}$ de consistência entre mapas de profundidade $\mathbf{D}_O$ e $\mathbf{D}_R$. Considerando *pixels* brancos como $I'_{VAL}(u,v) = 1$ e *pixels* pretos como $I'_{VAL}(u,v) = 0$, para cada *pixel* $I'_{VAL}(u,v)$, avalia-se os *pixels* $I'_{VAL}(u-1,v)$, $I'_{VAL}(u,v-1)$, $I'_{VAL}(u,v)$, $I'_{VAL}(u+1,v)$ e $I'_{VAL}(u,v+1)$. Se todos estes forem iguais a 1, $I''_{VAL}(u,v) = 1$; caso contrário, $I''_{VAL}(u,v) = 0$.

5.3.2 Combinação de altas frequências

Na Seção anterior, apresentou-se duas maneiras de estimar a alta frequência de $\mathbf{I}^A_O$. É necessário combinar esta estimativas, considerando também que pode haver até dois quadros de referência, advindos de vistas adjacentes.

Assim como na Seção 4.3.2, propõe-se uma soma ponderada para combinar os blocos $(\mathbf{u}_i, \mathbf{v}_i)$ de cada estimativa de alta frequência. Definindo $\mathbf{I}^A_{O,r}{}'$ e $\mathbf{I}^A_{O,r}{}''$ como as estimativas criadas a partir da r-ésima referência ($r \in \{1, 2\}$, dependendo da disponibilidade), cada bloco de $\hat{\mathbf{I}}^A_O$ é dado por

$$\hat{\mathbf{I}}^A_O(\mathbf{u}_i, \mathbf{v}_i) = \frac{\sum_{r=1}^{2} \left(w'_{r,i} \mathbf{I}^A_{O,r}{}'(\mathbf{u}_i, \mathbf{v}_i) + w''_{r,i} \mathbf{I}^A_{O,r}{}''(\mathbf{u}_i, \mathbf{v}_i) \right)}{\sum_{r=1}^{2} \left(w'_{r,i} + w''_{r,i} \right)}, \tag{5.7}$$

aonde os pesos $w'_{r,i}$ e $w''_{r,i}$ são dados por

$$\begin{aligned} w'_{r,i} &= 1/\mathrm{SSD}(\mathbf{I}^B_O(\mathbf{u}_i, \mathbf{v}_i), \mathrm{ID}(\mathbf{I}'_{O,r}(\mathbf{u}_i, \mathbf{v}_i))) \\ w''_{r,i} &= 1/\mathrm{SSD}(\mathbf{I}^B_O(\mathbf{u}_i, \mathbf{v}_i), \mathrm{ID}(\mathbf{I}''_{O,r}(\mathbf{u}_i, \mathbf{v}_i))) \end{aligned}, \tag{5.8}$$

$$\begin{aligned} \mathbf{I}'_{O,r}(\mathbf{u}_i, \mathbf{v}_i) &= \mathbf{I}^B_O(\mathbf{u}_i, \mathbf{v}_i) + \mathbf{I}^A_{O,r}{}'(\mathbf{u}_i, \mathbf{v}_i) \\ \mathbf{I}''_{O,r}(\mathbf{u}_i, \mathbf{v}_i) &= \mathbf{I}^B_O(\mathbf{u}_i, \mathbf{v}_i) + \mathbf{I}^A_{O,r}{}''(\mathbf{u}_i, \mathbf{v}_i) \end{aligned}. \tag{5.9}$$

Quanto mais semelhante a $\mathbf{I}^B_O(\mathbf{u}_i, \mathbf{v}_i)$ for o bloco $\mathrm{ID}(\mathbf{I}'_{O,r}(\mathbf{u}_i, \mathbf{v}_i))$, menor será a SSD entre estes dois últimos blocos, e maior será o peso $w'_{r,i}$ dado a $\mathbf{I}^A_{O,r}{}'(\mathbf{u}_i, \mathbf{v}_i)$. O mesmo vale para o peso $w''_{r,i}$ e o quadro $\mathbf{I}^A_{O,r}{}''$.

5.4 Resultados experimentais

A fim de avaliar o algoritmo proposto, foram testadas as mesmas sequências apresentadas na Seção 4.4. Novamente, os testes foram feitos com e sem codificação H.264/AVC. Como a arquitetura em consideração neste Capítulo é diferente daquela considerada no Capítulo anterior (Seções 5.2 e 4.2, respectivamente), mudam os quadros de referência disponíveis em algumas das sequências. A Tabela 5.1 apresenta estas considerações com mais detalhes, aonde as vistas e os quadros foram escolhidos de acordo com a disponibilidade de mapas de profundidade correspondentes, e a numeração das vistas e a quantidade de quadros disponíveis foram previamente determinadas pelas fontes das sequências. As Figs. 5.5 e 5.6 apresentam quadros de exemplo dos mapas de profundidade das sequências reais e sintéticas testadas, respectivamente.

5.4.1 Testes sem codificação H.264/AVC

O algoritmo proposto neste Capítulo foi comparado com a interpolação de quadros em baixa resolução, dentro das seguintes condições:

- O método foi empregado sobre os componentes de luminância dos quadros originais em resolução mista, tal como na Fig. 5.1, mas sem codificação, e foram medidos os ganhos de qualidade dos quadros super-resolvidos em relação à interpolação dos mesmos, com base nas

Tabela 5.1: Sequências utilizadas

Nome	**Resolução original**	**Resolução testada**	$\mathbf{I}_O$ **{vista}, {quadros}**	$\mathbf{I}_R$ **{vistas}, {quadros}**
Sequências reais				
Ballet	1024×768	512×384	$\{1\}, \{0-99\}$	$\{0,2\}, \{0-99\}$
Breakdancers	1024×768	512×384	$\{1\}, \{0-99\}$	$\{0,2\}, \{0-99\}$
Cafe	1920×1080	960×528	$\{3\}, \{0-99\}$	$\{2,4\}, \{0-99\}$
Pantomime	1280×960	640×480	$\{39\}, \{0-99\}$	$\{37,41\}, \{0-99\}$
Lovebird	1024×768	512×384	$\{6\}, \{0-99\}$	$\{4,8\}, \{0-99\}$
Newspaper	1024×768	512×384	$\{4\}, \{0-99\}$	$\{2,6\}, \{0-99\}$
Poznan Street	1920×1088	960×544	$\{4\}, \{0-99\}$	$\{3,5\}, \{0-99\}$
Sequências sintéticas				
Barn1	432×381	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Barn2	430×381	416×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Bull	433×381	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Map	284×216	272×208	$\{1\}, \{0\}$	$\{0\}, \{0\}$
Poster	435×383	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Sawtooth	434×380	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Venus	434×383	432×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Cones	450×375	448×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Teddy	450×375	448×368	$\{6\}, \{0\}$	$\{2\}, \{0\}$
Room3D	480×360	480×360	$\{2\}, \{0-99\}$	$\{1\}, \{0-99\}$

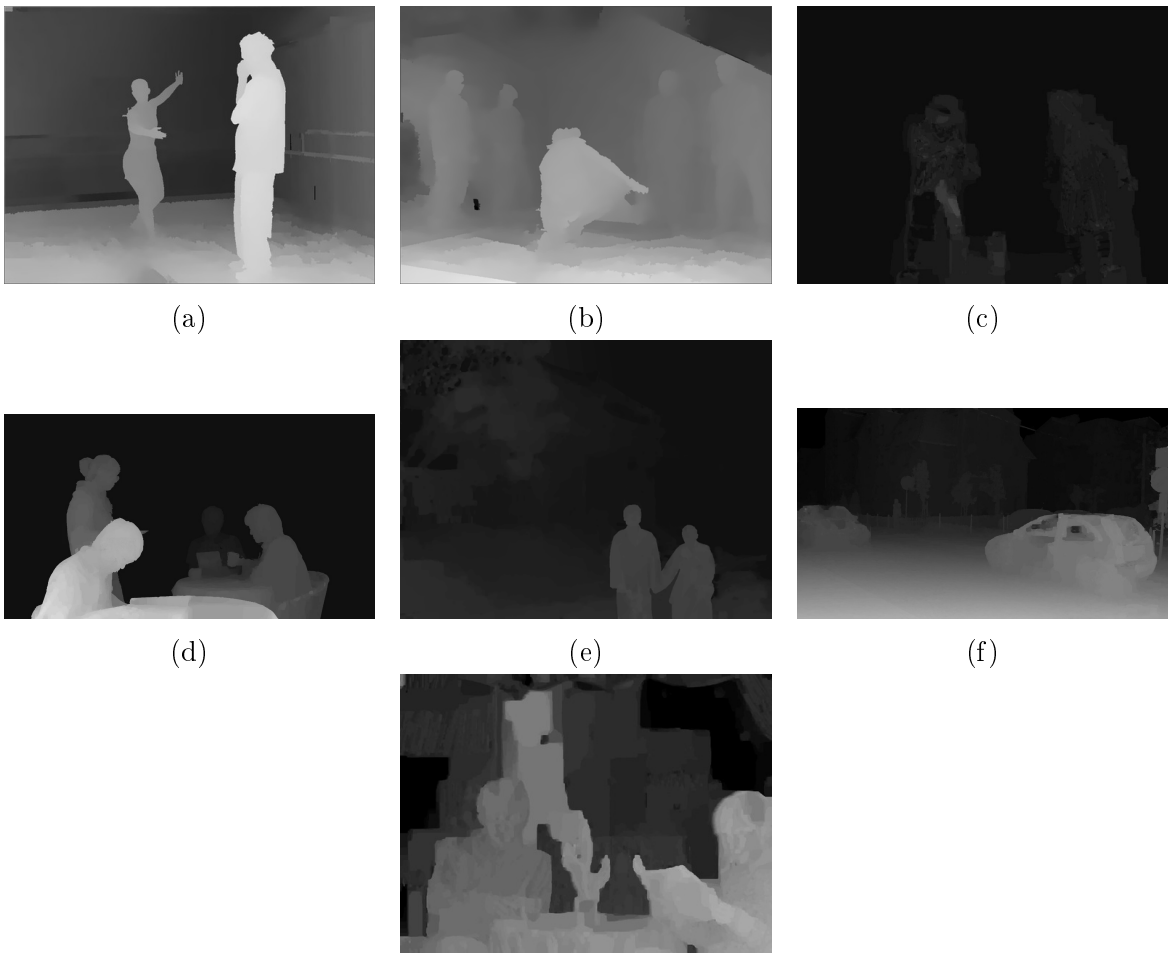


Figura 5.5: Quadros de exemplo dos mapas de profundidade das sequências reais testadas: (a) *Ballet*; (b) *Breakdancers*; (c) *Pantomime*; (d) *Cafe*; (e) *Lovebird1*; (f) *Poznan Street*; (g) *Newspaper*.

médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros interpolados e super-resolvidos.

- Seguindo a Fig. 5.1, foi considerada a vista que possui quadros somente em resolução baixa.
- Utilizou-se um bloco de tamanho 16×16 para a combinar as estimativas de alta frequência, assim como nos testes do Capítulo 4.
- Os ganhos obtidos pela super-resolução proposta em relação à interpolação são apresentados no Anexo I, Tabelas I.9 a I.12, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$). Também se consideram os resultados empregando as referências disponíveis individualmente e combinadas, a fim de avaliar o efeito de se reduzir a quantidade de referências disponíveis (resultados 1 a 7 para as sequências reais, e 1 a 3 para as sinteticas).

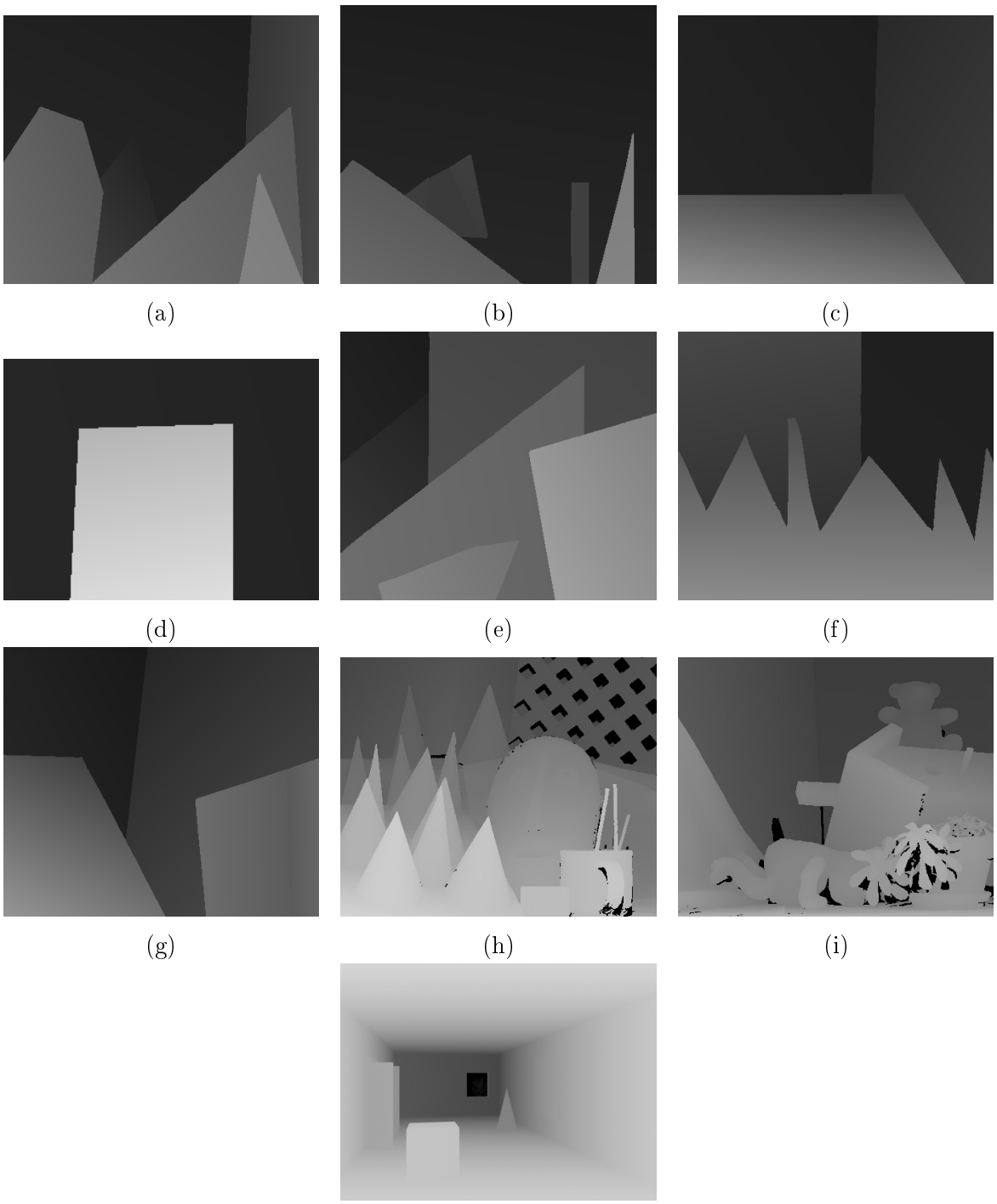


Figura 5.6: Quadros de exemplo dos mapas de profundidade das sequências sintéticas testadas: (a) *Barn1*; (b) *Barn2*; (c) *Bull*; (d) *Map*; (e) *Poster*; (f) *Sawtooth*; (g) *Venus*; (h) *Cones*; (i) *Teddy*; (j) *Room3D*.

Para as sequências reais, a combinação de todas as referências propicia os maiores ganhos em PSNR, sendo todos os ganhos positivos. Além disso, percebe-se que ao se utilizar somente uma referência, para $M = 2$, obtem-se melhores resultados através da combinação de $\mathbf{I}_{Oi}^{A}{}'$ e $\mathbf{I}_{Oi}^{A}{}''$

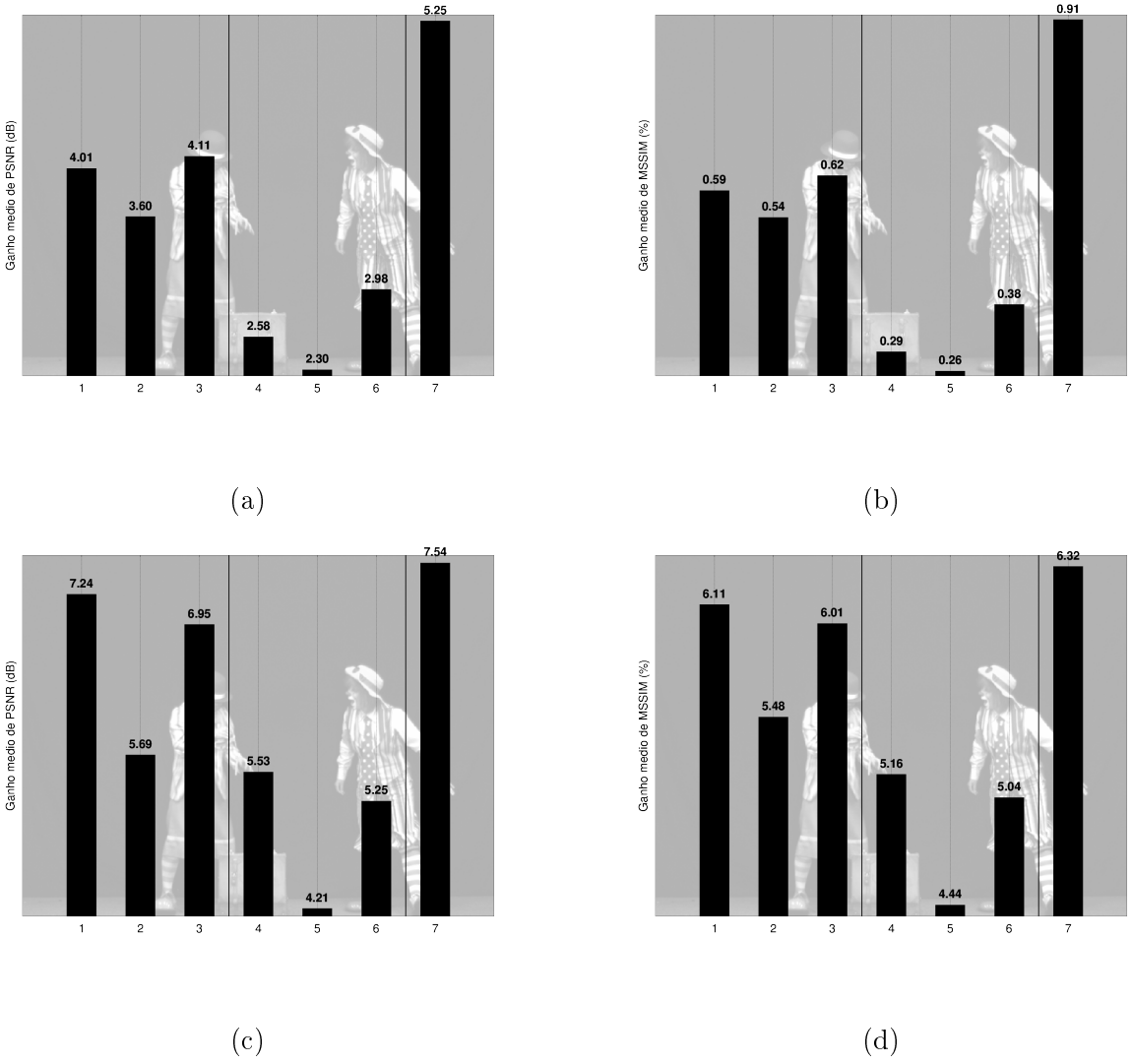


Figura 5.7: Resultados sem codificação para a componente de luminância da sequência *Pantomime*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 13$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_{O}^{B}$; (2) $\mathbf{I}_{O1}^{A}{}'' + \mathbf{I}_{O}^{B}$; (3) $\hat{\mathbf{I}}_{O1}$; (4) $\mathbf{I}_{O2}^{A}{}' + \mathbf{I}_{O}^{B}$; (5) $\mathbf{I}_{O2}^{A}{}'' + \mathbf{I}_{O}^{B}$; (6) $\hat{\mathbf{I}}_{O2}$; (7) $\hat{\mathbf{I}}_{O}$.

(resultados 3 e 6). Isto indica a existência de erros nos mapas de profudidade, cujos valores foram descartados corretamente através da erosão $\mathbf{I}''_{VAL}$ do mapa de *pixels* válidos $\mathbf{I}'_{VAL}$. Já para $M = 4$, verifica-se casos de perda de qualidade ao utilizar-se a combinação de $\mathbf{I}_{Oi}^{A}{}'$ e $\mathbf{I}_{Oi}^{A}{}''$ (resultados 3 e 6), em relação a utilizar somente $\mathbf{I}_{Oi}^{A}{}' + \mathbf{I}_{O}^{B}$ (resultados 1 e 4), exceto para as sequências *Breakdancers* ($O1$ e $O2$), *Lovebird1* ($O1$) e *Newspaper* ($O1$). Isto indica que para $M = 4$, a degradação de $\mathbf{I}_{O}^{B}$ atinge níveis em que é melhor acrescentar altas frequências advindas de valores dúbios de mapas de profundidade do que não acrescentar nada. De qualquer maneira, a combinação de $\mathbf{I}_{Oi}^{A}{}'$ e $\mathbf{I}_{Oi}^{A}{}''$ não representa uma grande perda quando utilizado para $M = 4$.

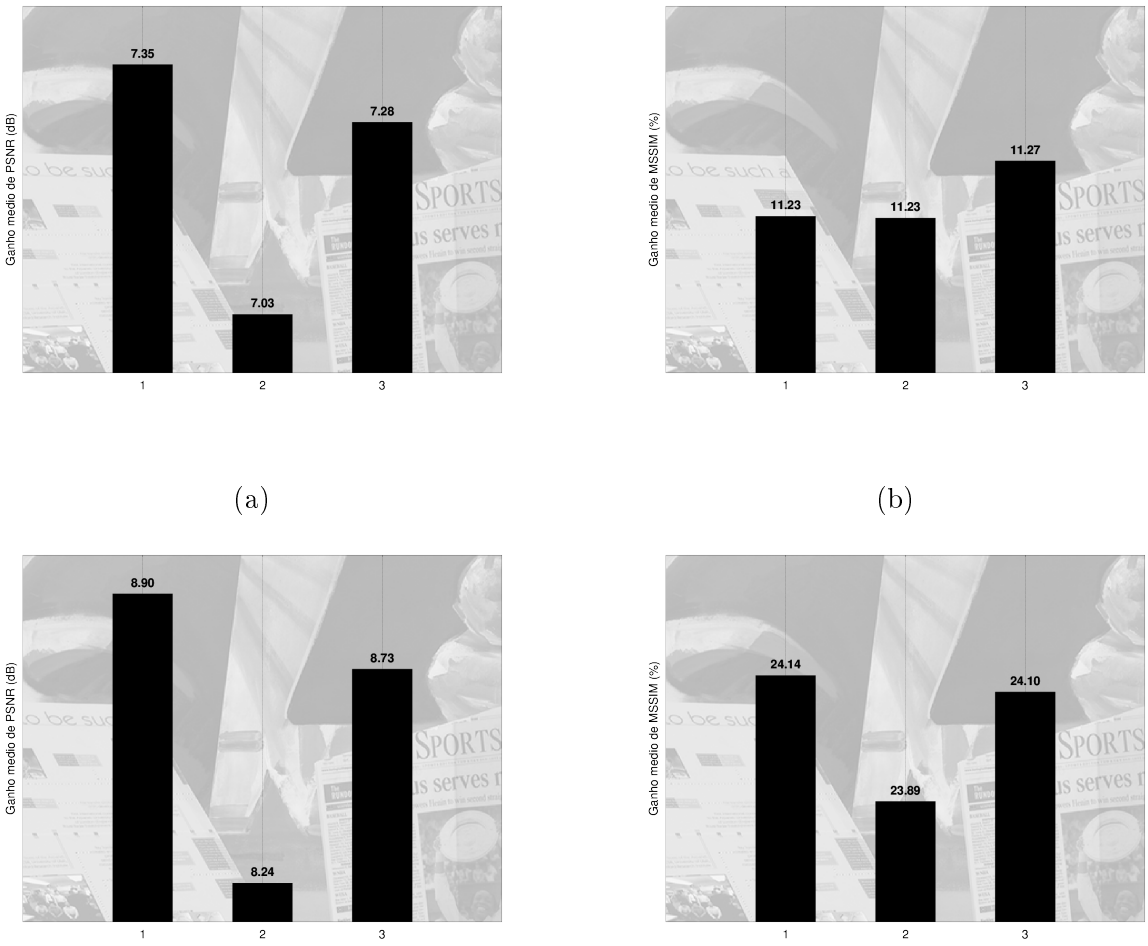


Figura 5.8: Resultados sem codificação para a componente de luminância da sequência *Venus*. São apresentados os ganhos da super-resolução proposta neste Capítulo em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 3$ indicam os ganhos utilizando as diferentes referências: (1) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$; (2) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$; (3) $\hat{\mathbf{I}}_O$.

O algoritmo proposto neste Capítulo apresenta ganhos significativos para as sequências sintéticas. Em nenhum caso verificou-se a perda de qualidade, independente da referência utilizada. Porém, como as sequências sintéticas apresentam mapas de profundidade muito precisos, o uso do quadro $\mathbf{I}_O^{A\prime\prime}$ não se mostrou tão eficaz. Para $M = 2$, a combinação de $\mathbf{I}_O^{A\prime}$ e $\mathbf{I}_O^{A\prime\prime}$ apresentou maior ganho em metade das sequências, e para $M = 4$, quatro sequências foram beneficiadas com esta combinação de altas frequências. É interessante notar que para a sequência *Cones*, verificou-se uma perda de qualidade ao utilizar o método de super-resolução proposto no Capítulo anterior.

As Figs. 5.7 e 5.8 apresentam os ganhos médios obtidos pela super-resolução para as sequências *Pantomime* e *Venus*, respectivamente, representando o comportamento típico obtido para as demais sequências.

A Fig. 5.9 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Ballet*, $M = 2$. Além da clara diferença entre as imagens $\mathbf{I}_O^B$ e $\mathbf{I}_O$ (Figs. 5.9(a) e (f)), nota-se que $\mathbf{I}_{O1}^{A}{}' + \mathbf{I}_O^B$ apresenta um contorno da bailarina projetado na parede, advindo de más projeções dos *pixels* da vista de referência (Fig. 5.9(b)). Através da erosão do mapa de *pixels* válidos, estes artefatos desaparecem, como se pode ver na Fig. 5.9(c), sem perder detalhes de alta frequência no rosto da bailarina, por exemplo.

A Fig. 5.10 apresenta detalhes das vistas decimada e interpolada, super-resolvidas e original para a sequência *Venus*, $M = 2$. Apesar de a sequência ser sintética, ainda se encontram pequenos erros nos mapas de profundidade, devido ao arredondamento dos valores de profundidade na representação por 8 *bits*, o que se reflete no surgimento de uma linha diagonal indevida à esquerda da cena, como se pode ver na Fig. 5.10(b). Os quadros $\mathbf{I}_O^{A''} + \mathbf{I}_O^B$ e $\hat{\mathbf{I}}_O$ (Figs. 5.10(c) e (d), respectivamente) eliminam parte desta linha diagonal, mas não completamente.

A Fig. 5.11 apresenta uma sobreposição dos quadros $\mathbf{I}'_{VAL}$ e $\mathbf{I}'_{O|R}$, e dos quadros $\mathbf{I}''_{VAL}$ e $\mathbf{I}''_{O|R}$. Nas Figs. 5.11(a) e (c), o mapa de *pixels* válidos ainda não foi erodido, e nas Figs. 5.11(b) e (d), ele foi erodido, retirando projeções indevidas, como aquelas vistas nas Figs. 5.9(c) e 5.10(c).

As Figs. 5.12 e 5.13 apresentam detalhes das vistas original, decimada e interpolada e super-resolvida para as sequências *Pantomime* e *Poster*, respectivamente, para $M = 4$, demonstrando a evidente melhora em qualidade obtida através do método de super-resolução proposto.

5.4.2 Testes com codificação H.264/AVC

Assim como na Seção 4.4.2, o algoritmo deste Capítulo foi aplicado às mesmas sequências após estas serem codificadas em resolução mista (Fig. 5.1), a fim de avaliar seu desempenho em uma situação prática. Aplicou-se as seguintes condições:

- Todas as sequências foram codificadas separadamente utilizando o padrão H.264/AVC em modo Intra, incluindo os mapas de profundidade. Os quadros $\mathbf{I}_{Ri}$, $\mathbf{D}_{Ri}$ e $\mathbf{D}_O$ foram codificados em resolução normal, de acordo com a Tabela 5.1, e o quadro $\mathbf{I}_O$ foi codificado em resolução inferior ($\mathbf{I}_O^D$), utilizando a decimação por $M = 2$ e $M = 4$.

- Aplicou-se os mesmo valores $QP = \{22, 27, 32, 37\}$ aos parâmetros de escalonamento, inclusive para os mapas de profundidade. Testes similares aos desta tese indicam que a utilização dos mesmos parâmetros de escalonamento para a luminância e para a profundidade das sequências oferece resultados adequados em termos de taxa e distorção [17].

- Para todas as sequências testadas, mediu-se a média quadro-a-quadro de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) de $\mathbf{I}_O^B$, que corresponde à versão interpolada de $\mathbf{I}_O^D$ após a codificação, e das versões super-resolvidas de $\mathbf{I}_O^B$.

Figura 5.9: Detalhes da sequência *Ballet*, vista 1, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 33,65 dB / MSSIM×100 = 94,40%); (b) $\mathbf{I}_{O1}^A{}' + \mathbf{I}_O^B$ (36,35 dB / 95,88%); (c) $\mathbf{I}_{O1}^A{}'' + \mathbf{I}_O^B$ (36,23 dB / 95,96%); (d) $\hat{\mathbf{I}}_{O1}$ (36,42 dB / 96,14%); (e) $\hat{\mathbf{I}}_O$ (37,57 dB / 96,92%); (f) $\mathbf{I}_O$.

- A taxa considerada foi igual à soma de toda arquitetura. Assim, considera-se novamente um cenário de *free-viewpoint television*, em que o usuário seleciona assistir à vista em consideração, Tabela 5.1. A mesma taxa foi considerada tanto para os quadros interpolados

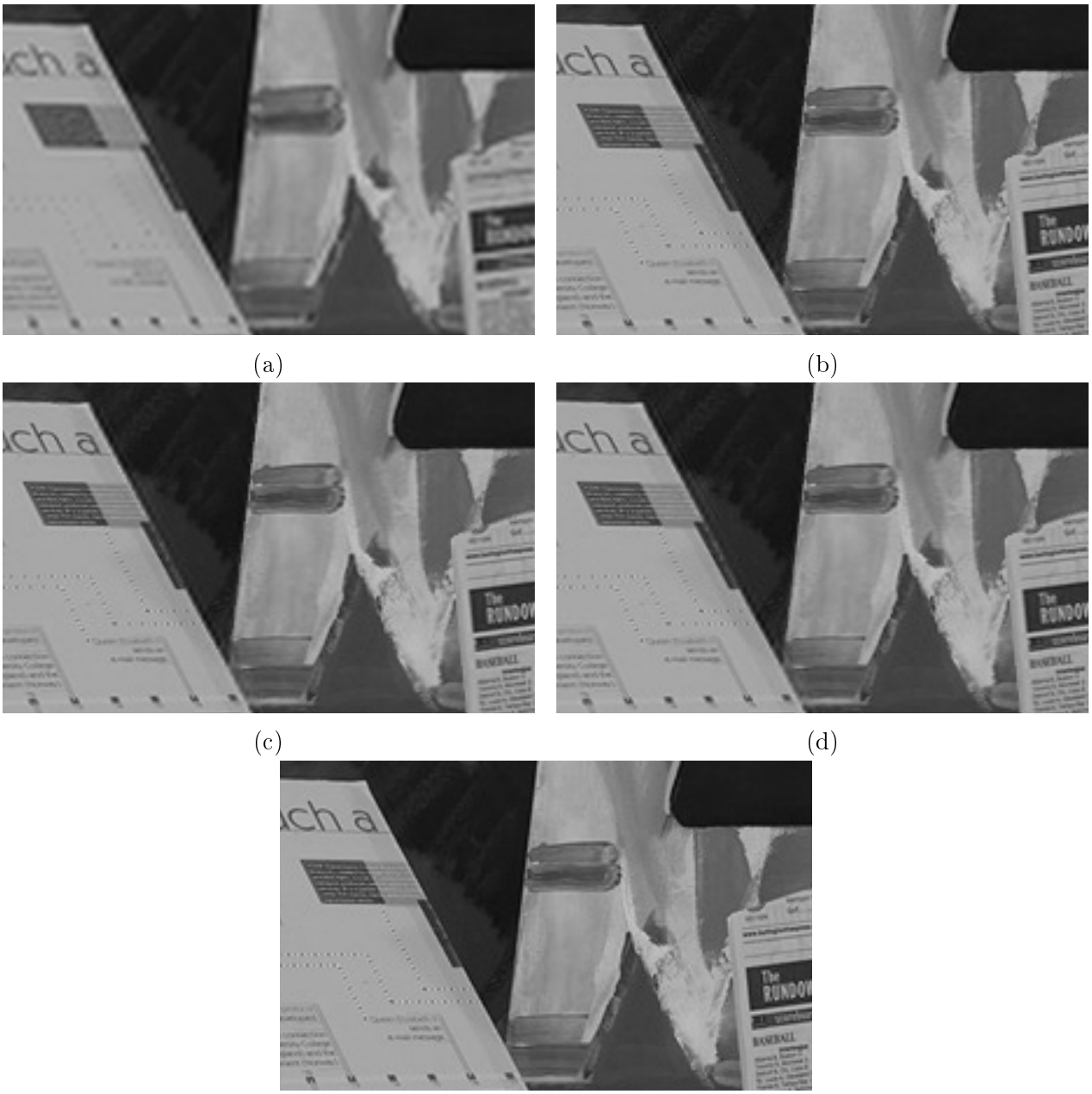


Figura 5.10: Detalhes da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 28,47 dB / MSSIM×100 = 86,24%); (b) $\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$ (35,81 dB / 97,40%); (c) $\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$ (35,50 dB / 97,40%); (d) $\hat{\mathbf{I}}_O$ (35,74 dB / 97,44%); (e) $\mathbf{I}_O$.

quanto para os super-resolvidos, visto que a super-resolução não precisa de nenhuma informação extra além dos quadros em resolução completa e dos mapas de profundidade.

- A partir destes valores de PSNR, MSSIM e taxa, foram medidos os ganhos médios [62] em termos de PSNR e MSSIM para o algoritmo proposto em relação a interpolar os quadros em baixa resolução, utilizando tanto cada referência individualmente quanto todas combinadas.

(a) (b)

(c) (d)

Figura 5.11: Detalhes dos *pixels* advindos de valores de profundidade válidos: (a) $\mathbf{I}'_{VAL}$ e $\mathbf{I}'_{O|R}$ sobrepostos para a sequência *Venus*, vista 6, quadro 0; (b) $\mathbf{I}''_{VAL}$ e $\mathbf{I}''_{O|R}$ sobrepostos para a mesma sequência; (c) $\mathbf{I}'_{VAL}$ e $\mathbf{I}'_{O|R}$ sobrepostos para a sequência *Ballet*, vista 1, quadro 1; (d) $\mathbf{I}''_{VAL}$ e $\mathbf{I}''_{O|R}$ sobrepostos para a mesma sequência.

- O *software* de referência JM 17.2 para o padrão H.264/AVC foi utilizado [61].
- Utilizou-se um bloco de tamanho 16×16 para a combinar as estimativas de alta frequência.
- Os ganhos médios do algoritmo proposto sobre à interpolação são apresentados no Anexo I, Tabelas I.13 a I.16, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$).

Verifica-se que a combinação de todas as referências propicia os maiores ganhos em PSNR e MSSIM para todas as sequências reais, mas não para todas as sequências sintéticas. Todas as

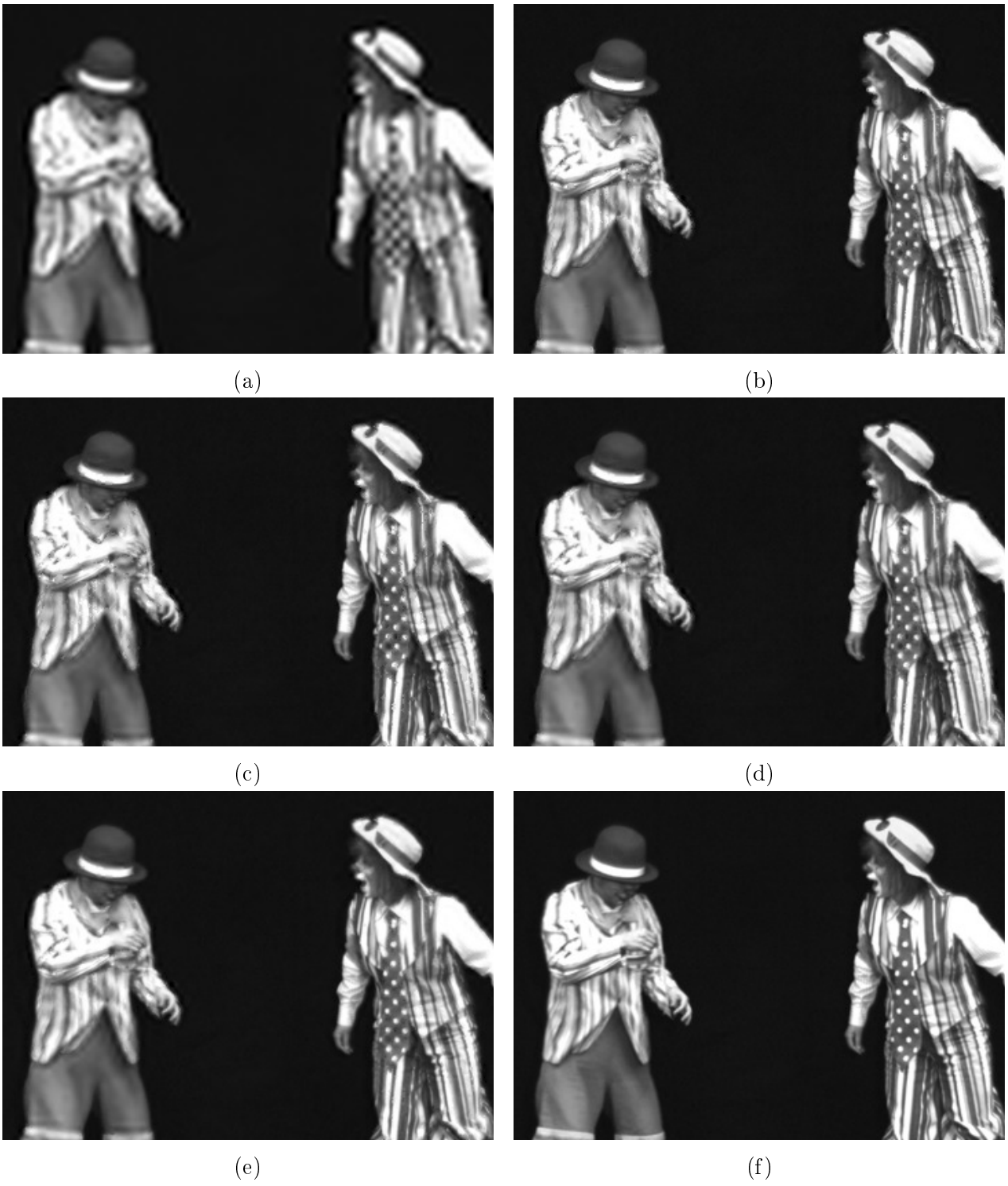


Figura 5.12: Detalhes da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR = 28,53 dB / MSSIM×100 = 93,01%); (b) $\mathbf{I}_{O1}^A{}' + \mathbf{I}_O^B$ (35,77 dB / 96,87%); (c) $\mathbf{I}_{O1}^A{}'' + \mathbf{I}_O^B$ (33,95 dB / 96,37%); (d) $\hat{\mathbf{I}}_{O1}$ (35,36 dB / 96,76%); (e) $\hat{\mathbf{I}}_O$ (36,05 dB / 97,09%); (f) $\mathbf{I}_O$.

sequências apresentaram ganho de qualidade, exceto para a MSSIM dos quadros $\mathbf{I}_{O1}^A{}'$ e $\mathbf{I}_{O1}^A{}''$ de *Breakdancers* e de *Newspaper*, quando $M = 2$, resultados 1 e 2. Em relação ao uso de somente um

quadro como referência, verifica-se que $\hat{\mathbf{I}}'_{Oi}$ (resultados 1 e 4) nem sempre possui melhor qualidade do que $\hat{\mathbf{I}}''_{Oi}$ (resultados 2 e 5), e que para $M = 2$ a combinação de ambos (resultados 3 e 6) oferece melhora de qualidade em relação a estes quadros individualmente.

As Figs. 5.14 e 5.15 apresentam o desempenho da super-resolução proposta em termos de taxa e distorção para as sequências *Pantomime* e *Poster*, respectivamente. Estas curvas são representativas do comportamento típico do desempenho para as sequências testadas. Assim como na Seção 4.4.2, verificou-se uma diferença no desempenho do algoritmo de acordo com o nível de escalonamento aplicado aos quadros decimado e de referência. Além disso, as curvas de taxa-distorção baseadas em PSNR e em MSSIM não apresentaram comportamentos idênticos, e algumas sequências apresentaram maiores melhoras em termos de PSNR, mas não em MSSIM, e vice-versa.

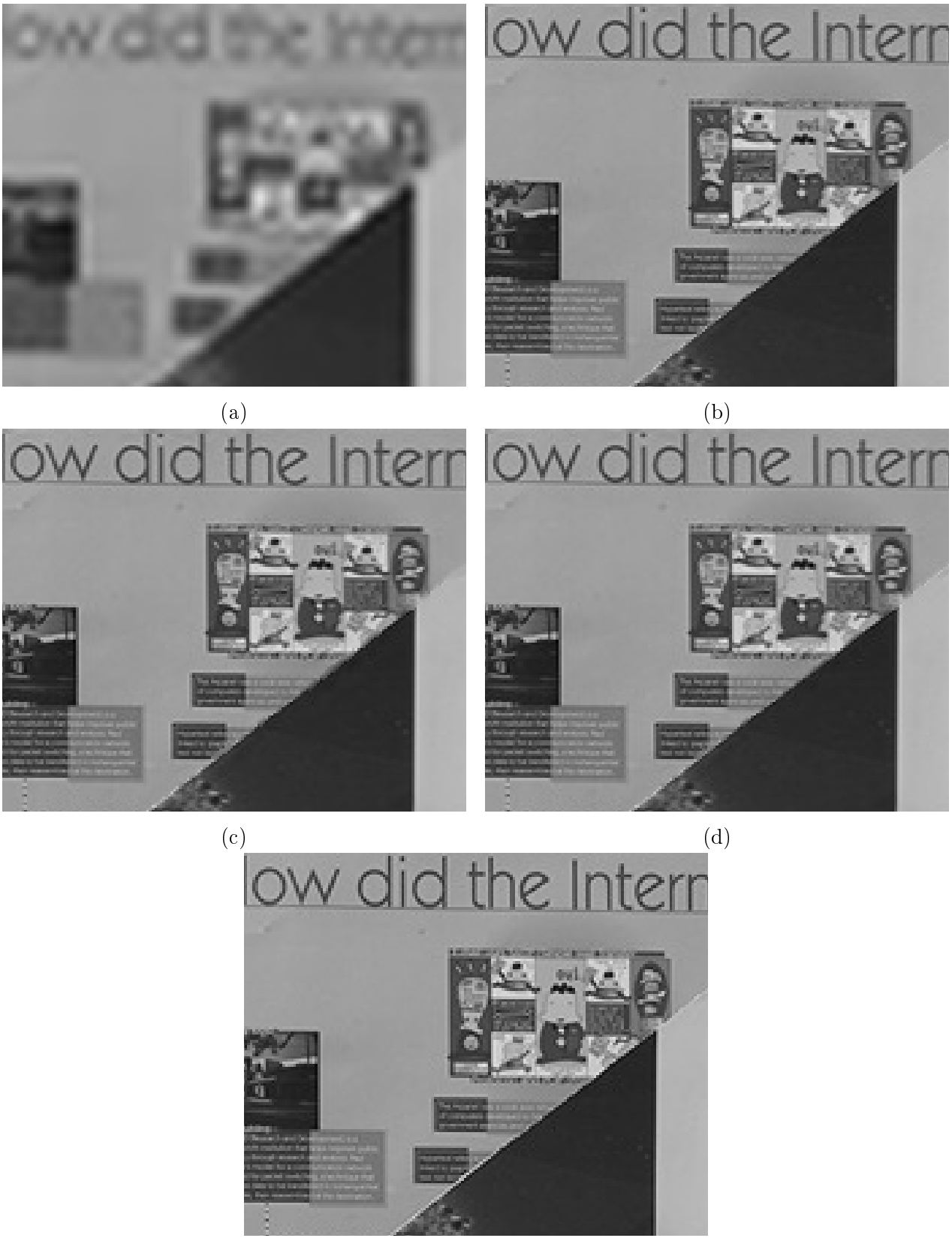


Figura 5.13: Detalhes da sequência *Poster*, vista 6, quadro 0, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR = 22,65 dB / MSSIM×100 = 57,59%); (b) $\hat{\mathbf{I}}'_O$ (31,93 dB / 96,34%); (c) $\hat{\mathbf{I}}''_O$ (31,53 dB / 96,10%); (d) $\hat{\mathbf{I}}_O$ (31,82 dB / 96,28%); (e) $\mathbf{I}_O$.

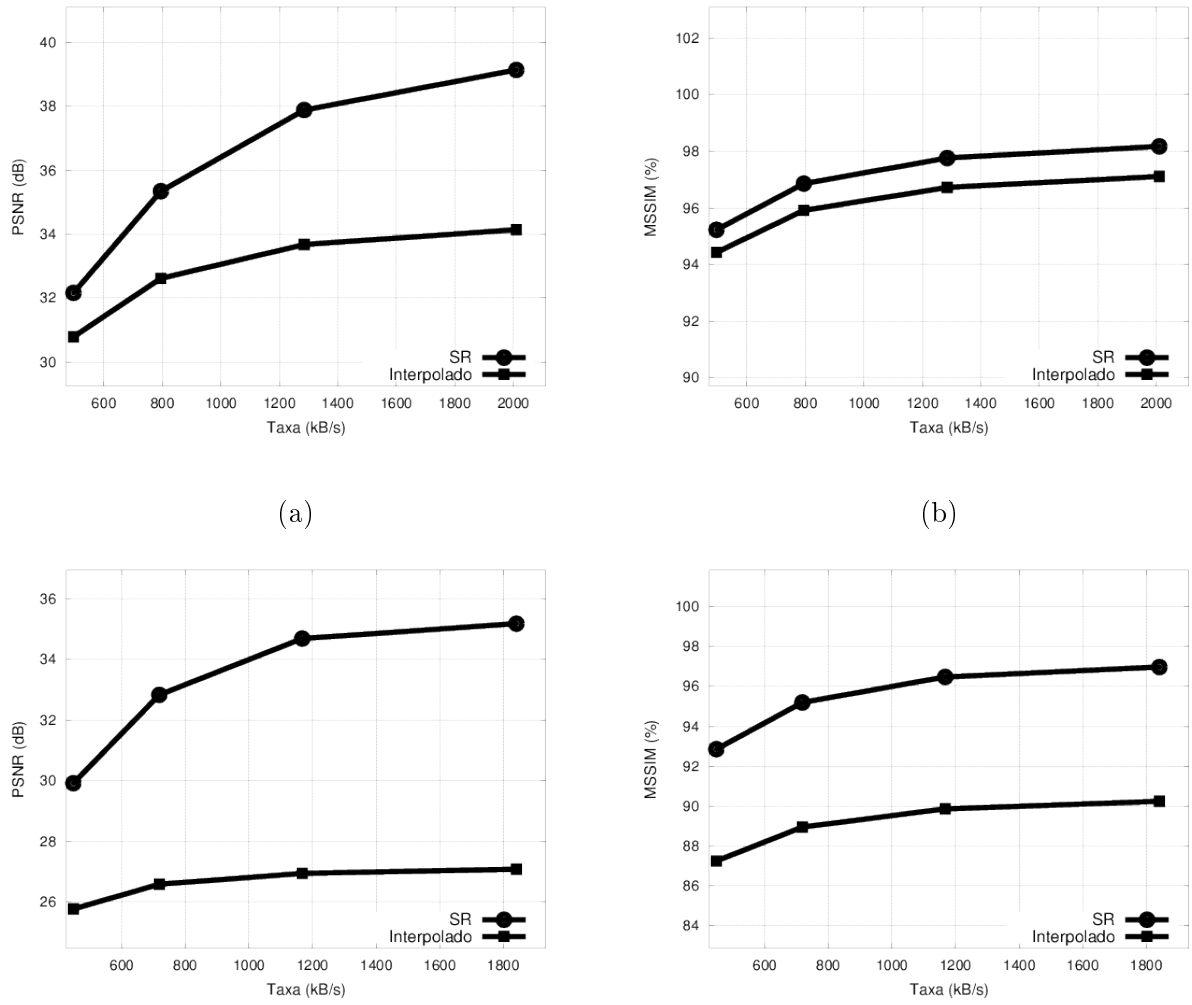
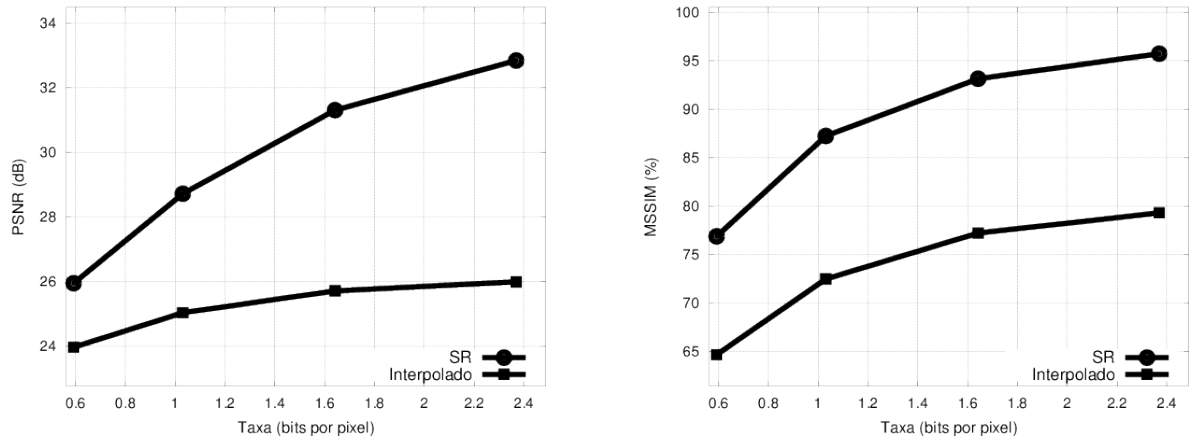
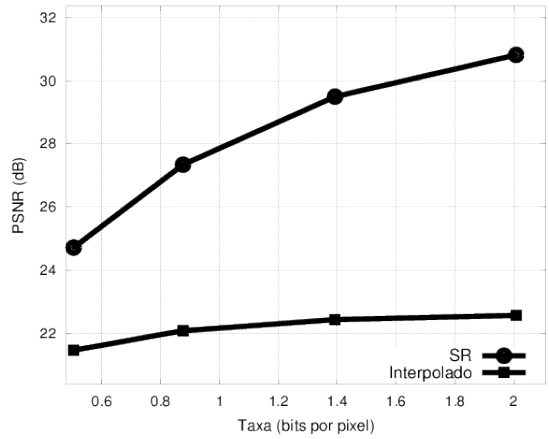


Figura 5.14: Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Pantomime*, vista 39, quadro 1: (a) taxa e PSNR, $M = 2$ (ganho médio de $3,37$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $0,98\%$); (c) taxa e PSNR, $M = 4$ (ganho médio de $6,75$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $6,35\%$).

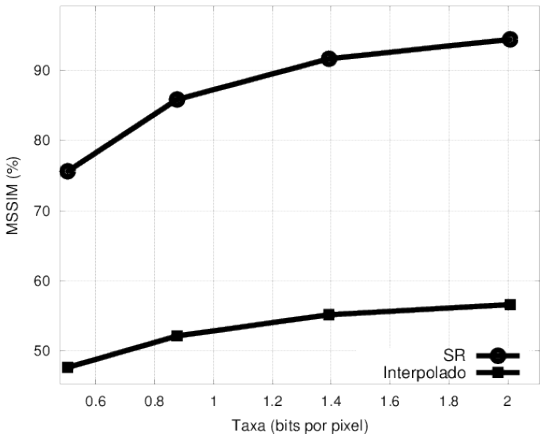


(a)

(b)



(c)



(d)

Figura 5.15: Desempenho em termos de taxa e distorção para o algoritmo proposto, utilizando todas as referências disponíveis, e para a interpolação da sequência *Poster*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$ (ganho médio de $4,30$ dB); (b) taxa e MSSIM, $M = 2$ (ganho médio de $14,89\%$); (c) taxa e PSNR, $M = 4$ (ganho médio de $5,79$ dB); (d) taxa e MSSIM, $M = 4$ (ganho médio de $34,12\%$).

Capítulo 6

Super-resolução de mapas de profundidade em baixa resolução

6.1 Introdução

O Capítulo anterior apresentou um algoritmo de super-resolução de múltiplas vistas em resolução mista baseado em mapas de profundidade. O algoritmo tem por entrada uma imagem em resolução reduzida $\mathbf{I}_O^D$, V imagens de referência $\mathbf{I}_{Rr}$ na resolução original, $r \in \{1, ..., V\}$, e mapas de profundidade correspondentes $\mathbf{D}_O$ e $\mathbf{D}_{Rr}$, também na resolução original. A consistência entre os mapas de profundidade é conferida a fim de detectar possíveis erros, devidos a oclusões, imprecisões de representação e escalonamento. Neste Capítulo, apresenta-se a terceira arquitetura de codificação de sequências de múltiplas vistas em resolução mista, aonde os mapas de profundidade são codificados em baixa resolução. Uma terceira técnica de super-resolução é proposta para estes mapas na Seção seguinte, que são posteriormente utilizados pelo processo de super-resolução proposto no Capítulo anterior. Novamente, resultados experimentais são apresentados para sequências de múltiplas vistas reais e sintéticas, sem e com codificação pelo padrão H.264/AVC.

6.2 Arquitetura em consideração

Como foi visto na Seção 3.3.2, o fato de possuirem grandes áreas planas faz com que mapas de profundidade possam ser codificados em baixa resolução, dado que boa parte dos *pixels* serão posteriormente recuperados no decodificador por um simples processo de interpolação. A codificação em baixa resolução permite uma grande redução na taxa de transmissão, dado que quadros em resolução menor possuem menor quantidade de *pixels* a serem codificados. A mesma ideia foi apresentada na Seção 3.3.4, porém aplicada a sequências de video em múltipla vista ao invés de mapas de profundidade, e com a ressalva de que não há perda de qualidade subjetiva, mas sim objetiva. Os mapas de profundidade também apresentarão perda de qualidade objetiva, mas dependendo do conteúdo do mapa, esta perda poderá não ser significativa.

A Fig. 6.1 ilustra a arquitetura proposta, em que os mapas de profundidade são codificados em baixa resolução. Assim como nas Seções 4.2 e 5.2, as vistas codificadas em baixa resolução deverão ser interpoladas de volta às suas resoluções originais.

A arquitetura presente destina-se a reduzir a taxa de transmissão e a complexidade de codificação de sistemas que devem atender a diversos tipos de receptores, incluindo telas autoestereoscópicas e *free-viewpoint television*. A técnica de super-resolução apresentada neste Capítulo seria empregada a fim de melhorar os mapas que serão posteriormente empregados na super-resolução da vista em baixa resolução, adequando a sequência transmitida a aplicações em *free-viewpoint television*.

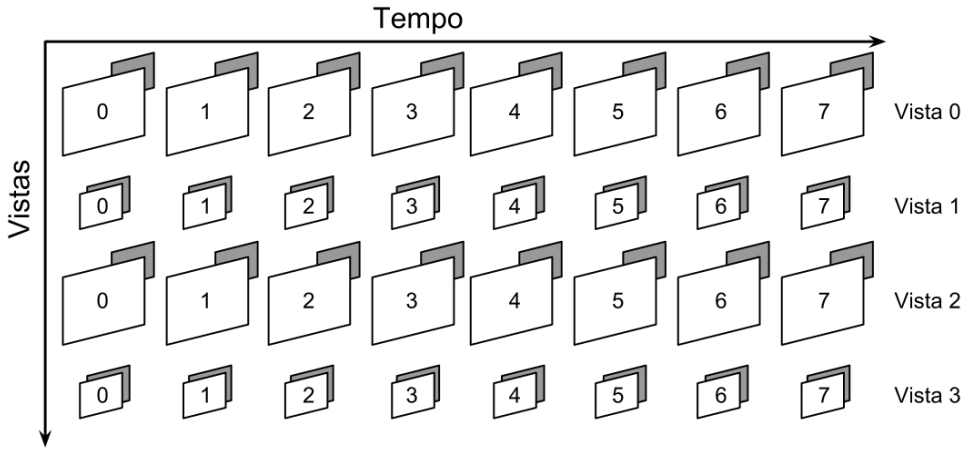


Figura 6.1: Arquitetura de codificação de múltiplas vistas com mapas de profundidade em baixa resolução.

6.3 Solução proposta

Baseado na arquitetura apresentada na Seção anterior, não é possível aplicar os processos de super-resolução das Seções 4.3 ou 5.3 diretamente aos mapas de profundidade codificados em baixa resolução, visto que não há mapas em resolução normal para servirem de referência. Ao invés disso, introduz-se uma terceira forma de super-resolução, baseada na combinação de múltiplas imagens, como apresentado na Seção 2.4.1.

A super-resolução proposta consiste em três etapas: projeção de mapas, preenchimento de buracos e filtragem, como mostrado na Fig. 6.2. A super-resolução baseada na combinação de múltiplas imagens pressupõe o registro desta imagens, mas essa etapa não se faz necessária, dado que os mapas de profundidade representam indiretamente o registro *pixel*-a-*pixel* entre eles. Os mapas em baixa resolução são mutuamente super-resolvidos com a ajuda dos mapas das outras vistas para o mesmo instante em consideração. Nas Seções a seguir, define-se as três etapas da solução proposta.

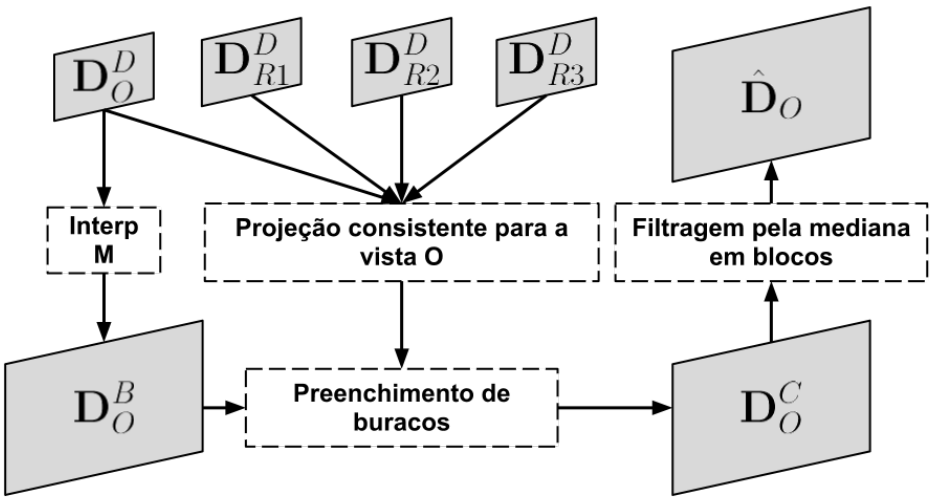


Figura 6.2: Super-resolução de um mapa de profundidade $\mathbf{D}_O^D$ em baixa resolução utilizando três mapas $\mathbf{D}_{R1}^D$, $\mathbf{D}_{R2}^D$ e $\mathbf{D}_{R3}^D$, também em baixa resolução.

6.3.1 Projeção dos mapas

Seguindo a nomenclatura de Capítulos anteriores, tem-se um mapa de profundidade decimado na j-ésima vista e k-ésimo instante, $\mathbf{D}_{j,k}^D = \mathbf{D}_O^D$, que será processado, gerando a versão super-resolvida $\hat{\mathbf{D}}_O$. A primeira etapa do método proposto consiste em projetar consistentemente para a vista O os mapas decimados das vistas adjacentes $\mathbf{D}_{Rr}^D$, $r \in \{1, ..., V\}$, também em baixa resolução, acrescentando novas informações ao mapa $\mathbf{D}_O^D$.

Seguindo as Eqs. 3.4 a 3.7, os *pixels* do mapa $\mathbf{D}_{Rr}^D$ são projetados para a vista O, utilizando o teste de consistência apresentado na Seção 5.3.1, Fig. 5.3. Todavia, a projeção proposta é direta, diferente da projeção apresentada na Seção 5.3.1, que é reversa. Isto acontece porque naquele caso os mapas de profundidade já se encontram em alta resolução, e o que se deseja conhecer é a posição (u', v') na vista de referência R correspondente à posição (u, v) na vista O. No método proposto nesta Seção, deseja-se o oposto, isto é, projetar a posição (u, v) na vista Rr à posição (u', v') na vista O, a fim de acrescentar informação de fração de *pixel* a $\mathbf{D}_O^D$.

Cada *pixel* $D_{Rr}^D(\mathbf{m}_{Ri}) = D_{Rr}^D(u, v)$ que passa no teste de consistência com o mapa $\mathbf{D}_O^D$ é projetado para uma posição $\mathbf{m}'_{Ri} = (u', v', 1)^T$ na versão decimada $\mathbf{D}_O^D$ da vista O, e para a posição $\mathbf{m}''_{Ri} = (Mu', Mv', 1)^T$ na versão interpolada $\mathbf{D}_O^I$ da mesma vista (onde M é o fator de decimação e interpolação). Sendo assim, quaisquer posições inteiras $\{\mathbf{m}_{O1}, \mathbf{m}_{O2}, \mathbf{m}_{O3}, \mathbf{m}_{O4}\}$ em $\mathbf{D}_O^I$ terão uma nuvem de C candidatos próximos, denominados $\{\mathbf{m}''_{R1}, ..., \mathbf{m}''_{RC}\}$, onde $\|\mathbf{m}_{Oi} - \mathbf{m}''_{Rr}\| < 1, i \in \{1, 2, 3, 4\}, r \in \{1, ..., C\}$. A Fig. 6.3 ilustra o caso para três candidatos ($C = 3$).

A interpolação da nuvem de candidatos segue a ponderação pelo inverso da distância quadrática, que oferece resultados empíricos satisfatórios e simplifica os cálculos em relação a outros expoentes para a distância [63]. Para as quatro posições inteiras $\{\mathbf{m}_{O1}, \mathbf{m}_{O2}, \mathbf{m}_{O3}, \mathbf{m}_{O4}\}$ em $\mathbf{D}_O^I$, o valor da interpolação é dada por:

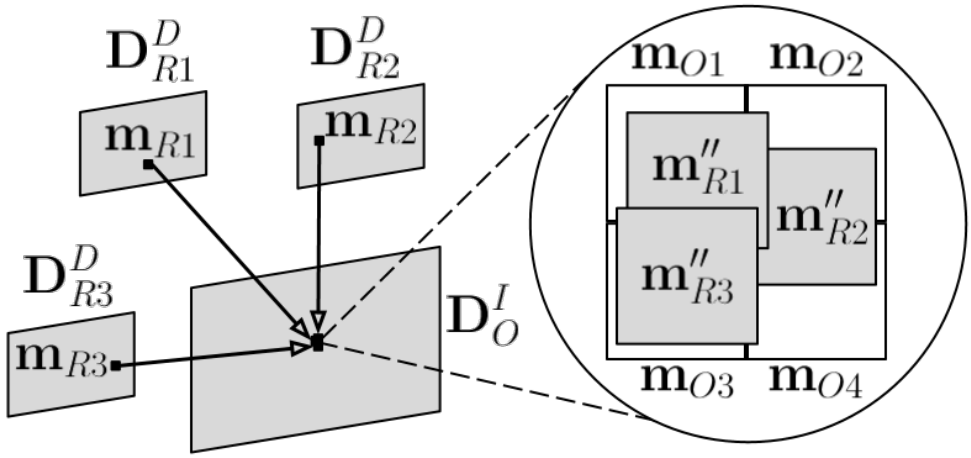


Figura 6.3: Nuvem de três pontos $\{\mathbf{m}''_{R1}, \mathbf{m}''_{R2}, \mathbf{m}''_{R3}\}$ projetados de posições $\{\mathbf{m}_{R1}, \mathbf{m}_{R2}, \mathbf{m}_{R3}\}$ em vistas adjacentes. $\{\mathbf{m}_{O1}, \mathbf{m}_{O2}, \mathbf{m}_{O3}, \mathbf{m}_{O4}\}$ representam posições inteiras próximas na vista O.

$$D^I_O(\mathbf{m}_{Oi}) = \frac{\sum_{r=1}^{C} d_{r,i} D^D_{Rr}(\mathbf{m}_{Rr})}{\sum_{r=1}^{C} d_{r,i}} \qquad d_{r,i} = \|\mathbf{m}_{Oi} - \mathbf{m}''_{Rr}\|^{-2} \tag{6.1}$$

Note que os *pixels* de $D^D_O(u,v)$ correspondem diretamente aos *pixels* $D^I_O(Mu, Mv)$, como indica a Eq. 2.8. Nesses casos, a Eq. 6.1 não é aplicada, pois $d_{r,i} \to \infty$, e $D^I_O(Mu, Mv) = D^D_O(u,v)$.

6.3.2 Preenchimento de buracos

O processo apresentado na Seção anterior não garante que todos os *pixels* de $\mathbf{D}^I_O$ serão preenchidos, pois nem todos os *pixels* de $\mathbf{D}^D_{Rr}$ passam pelos teste de consistência com $\mathbf{D}^D_O$. Os *pixels* $D^I_O(u,v)$ não preenchidos na etapa anterior são então substituídos por $D^B_O(u,v)$, que é a versão interpolada de $\mathbf{D}^D_O$. Desta maneira, gera-se o mapa $\mathbf{D}^C_O$, que representa a combinação de $\mathbf{D}^D_O$, $\mathbf{D}^B_O$ e $\mathbf{D}^D_{Rr}$.

6.3.3 Filtragem de mapas de profundidade

A Seção 3.2.2 apresentou algumas características básicas de mapas de profundidade, tais como a presença de grandes áreas suaves delimitadas por transições bruscas nas bordas de objetos. Buscando tais características, procura-se filtrar o mapa advindo da etapa anterior.

Um filtro simples que mantém as características citadas é o filtro de mediana. Assim como o filtro passa-baixas, ele suaviza áreas planas ruidosas, mas ao contrário do filtro passa-baixas, ele não borra as bordas de objetos. Uma imagem $\mathbf{I}$ filtrada pela mediana $\mathbf{I}^{MED}$ é dada por:

$$\begin{gathered} I^{MED}(u,v) = \text{mediana}(\mathbf{I}(\mathbf{u}_j, \mathbf{v}_j)) \\ \mathbf{u}_j = \{u-w, u-w+1, ..., u+w-1, u+w\} \\ \mathbf{v}_j = \{v-w, v-w+1, ..., v+w-1, v+w\} \\ w = (bl-1)/2 \end{gathered} , \tag{6.2}$$

onde *bl* é o tamanho do bloco com o qual a filtragem de mediana é aplicada. Assim, cada *pixel* em $\mathbf{I}^{MED}$ é dado pela mediana dos *pixels* da imagem original em uma janela em volta da posição atual. Se *bl* for ímpar, $\mathbf{I}^{MED}$ não possuirá valores inexistentes em $\mathbf{I}$, já que a mediana simplesmente ordena os *pixels* e escolhe aquele localizado na posição central do vetor de ordenação. Sendo assim, se o filtro de mediana for aplicado a um mapa de profundidade, as regiões planas que forem ruidosas terão menos ruído adicionado, e as bordas dos objetos não serão borradas.

A última etapa do método proposto consiste em passar o mapa $\mathbf{D}_O^C$ por um filtro de mediana, gerando a versão super-resolvida $\hat{\mathbf{D}}_O$.

6.4 Resultados experimentais

A fim de avaliar o método de super-resolução proposto neste Capítulo, mediu-se a qualidade dos quadros $\mathbf{I}_O^D$ super-resolvidos com diferentes mapas de profundidade pré-processados. A qualidade dos mapas de profundidade em si não é tão importante, dado que eles não são o produto final de uma arquitetura em múltiplas vistas, e sim as imagens $\hat{\mathbf{I}}_O$ e $\mathbf{I}_{Ri}$. Foram utilizadas as mesmas sequências apresentadas nas Seções 4.4 e 5.4, com e sem codificação H.264/AVC. Os quadros de referência disponíveis são os mesmos da Tabela 5.1.

A fim de avaliar a contribuição de cada uma das etapas do método proposto (Seções 6.3.1, 6.3.2 e 6.3.3), foram testados diferentes processamentos para os mapas de profundidade, a saber:

- $\mathbf{D}_r$: mapas de profundidade em resolução completa (o equivalente à arquitetura do Capítulo 5);
- $\mathbf{D}_r^B$: mapas $\mathbf{D}_r$ decimados e interpolados;
- $\mathbf{D}_r^{B,MED}$: $\mathbf{D}_r^B$ filtrado pela mediana;
- $\hat{\mathbf{D}}_{r1}$: $\mathbf{D}_r^B$ mutuamente super-resolvidos com uma referência (método proposto neste Capítulo);
- $\hat{\mathbf{D}}_r$: $\mathbf{D}_r^B$ mutuamente super-resolvidos com todas as referências (método proposto neste Capítulo).

Dada a enorme quantidade de resultados gerados, serão apresentados os valores de qualidade objetiva de $\hat{\mathbf{I}}_O$ para cada sequência utilizando todas as referências disponíveis, Eq. (5.7). Estes resultados representam com fidelidade a tendência seguida pelos quadros de super-resolução utilizando as referências individualmente, $\hat{\mathbf{I}}_{O1}$ e $\hat{\mathbf{I}}_{O2}$.

6.4.1 Testes sem codificação H.264/AVC

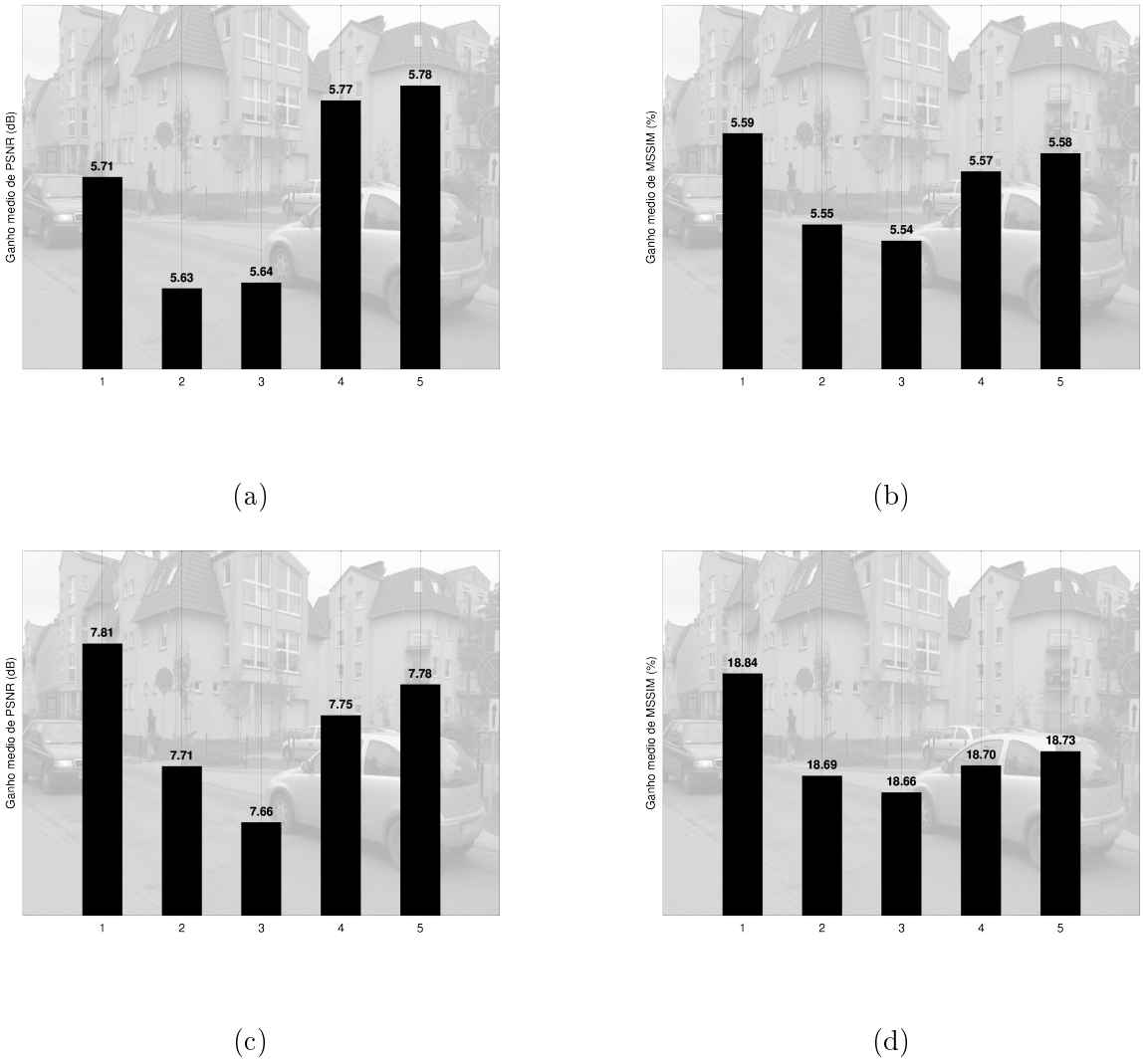


Figura 6.4: Resultados sem codificação para a componente de luminância da sequência *Poznan Street*. São apresentados os ganhos de $\hat{\mathbf{I}}_O$ em relação a $\mathbf{I}_O^B$ utilizando os mapas de profundidade super-resolvidos com o método proposto neste Capítulo. Os ganhos baseiam-se na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros de $\hat{\mathbf{I}}_O$ e $\mathbf{I}_O^B$: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 5$ indicam os ganhos utilizando os seguintes mapas de profundidade: (1) $\mathbf{D}_r$; (2) $\mathbf{D}_r^B$; (3) $\mathbf{D}_r^{B,MED}$; (4) $\hat{\mathbf{D}}_{r1}$; (5) $\hat{\mathbf{D}}_r$.

As diferentes formas de processamento dos mapas de profundidade foram primeiramente comparadas da seguinte maneira:

- Mediram-se os ganhos da super-resolução da componente de luminância de $\mathbf{I}_O^D$ em relação à sua versão interpolada, baseado nos mapas de profundidade processados, sem qualquer codificação, e utilizando todas as referências disponíveis. A qualidade foi medida através das

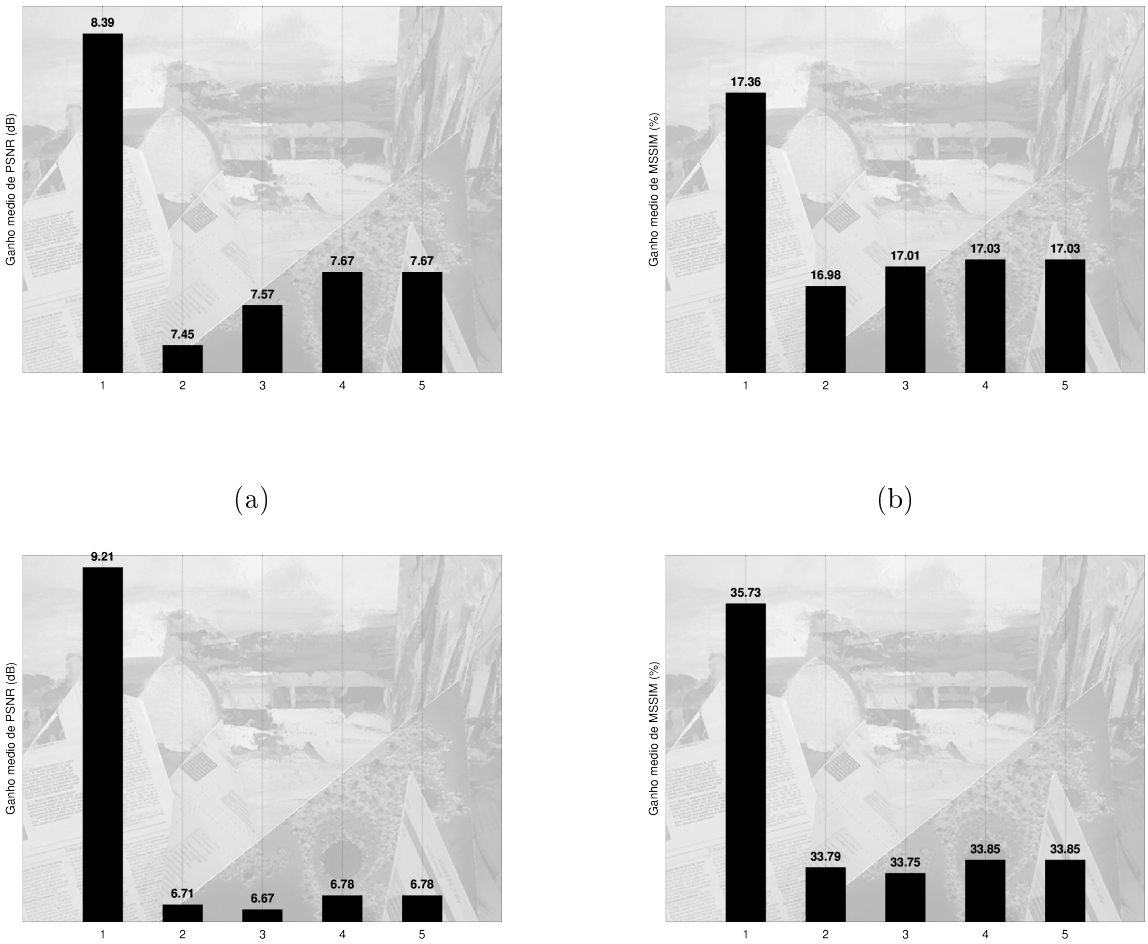


Figura 6.5: Resultados sem codificação para a componente de luminância da sequência *Barn1*. São apresentados os ganhos de $\hat{\mathbf{I}}_O$ em relação a $\mathbf{I}_O^B$ utilizando os mapas de profundidade super-resolvidos com o método proposto neste Capítulo. Os ganhos baseiam-se na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros de $\hat{\mathbf{I}}_O$ e $\mathbf{I}_O^B$: (a) PSNR, $M = 2$; (b) MSSIM, $M = 2$; (c) PSNR, $M = 4$; (d) MSSIM, $M = 4$. Os números $1 - 5$ indicam os ganhos utilizando os seguintes mapas de profundidade: (1) $\mathbf{D}_r$; (2) $\mathbf{D}_r^B$; (3) $\mathbf{D}_r^{B,MED}$; (4) $\hat{\mathbf{D}}_{r1}$; (5) $\hat{\mathbf{D}}_r$.

médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros de $\mathbf{I}_O^D$ interpolados e super-resolvidos.

- Seguindo a Fig. 5.1, foi considerada a vista que possui quadros somente em resolução baixa.

- Aplicou-se blocos de tamanho 3×3 para o filtro de mediana utilizado em $\mathbf{D}_r^{B,MED}$, $\hat{\mathbf{D}}_{r1}$ e $\hat{\mathbf{D}}_r$. Para a super-resolução dos quadros $\mathbf{I}_O^D$, aplicou-se blocos de tamanho 16×16 para a combinação de altas frequências, assim como nos testes do Capítulo 4.

- Os ganhos obtidos pela super-resolução proposta em relação à interpolação são apresentados no Anexo I, Tabelas I.17 a I.18, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$).

Observando os resultados para $M = 2$, percebe-se uma gradual melhora de qualidade de $\mathbf{D}_r^B$, para $\mathbf{D}_r^{B,MED}$, para $\hat{\mathbf{D}}_{r1}$, para $\hat{\mathbf{D}}_r$ (resultados 2 a 5), com exceção das sequências *Breakdancers*, *Lovebird1*, *Bull*, *Venus* e *Teddy*. A queda média de qualidade para estas cinco sequências não ultrapassa $0,13$ dB, constituindo perda muito baixa. Em compensação, os ganhos alcançados nas demais sequências por $\hat{\mathbf{D}}_r$ em relação a $\mathbf{D}_r^B$ variam de $0,05$ dB para *Breakdancers* até $0,47$ dB para *Pantomime*. Resultados semelhantes são obtidos para $M = 4$. Sendo assim, pode-se concluir que a super-resolução proposta oferece ganhos de qualidade para a arquitetura proposta neste Capítulo em relação a super-resolver $\mathbf{I}_O^D$ com os mapas de profundidade decimados e interpolados. Além disso, os resultados de $\mathbf{D}_r^{B,MED}$ e $\hat{\mathbf{D}}_r$ indicam que o filtro de mediana não foi o único responsável pelos ganhos da super-resolução de mapas, cuja etapa de projeção dos mapas adjacentes ofereceu claros ganhos em relação à simples filtragem.

As sequências *Pantomime* ($M = 2$ e $M = 4$), *Newspaper* ($M = 2$ e $M = 4$) e *Poznan Street* ($M = 2$) apresentam uma interessante característica: as super-resoluções baseadas em $\hat{\mathbf{D}}_O$ possuem melhor qualidade do que aquelas baseadas em $\mathbf{D}_O$. Isto indica que os mapas de profundidade destas sequências criam correspondências mais fidedignas entre as vistas após o processamento por super-resolução mútua. É importante notar que os testes apresentados nesta Seção foram feitos sem codificação, o que torna estes resultados contra-intuitivos, já que o processo de decimação e interpolação teoricamente reduziria a qualidade dos mapas de profundidade.

Comparando os resultados obtidos com $\mathbf{D}_r^B$ e $\mathbf{D}_r^{B,MED}$, resultados 2 e 3, percebe-se um aumento de qualidade do primeiro para o segundo processamento, indicando que as imprecisões dos mapas de profundidade decimados e interpolados podem ser parcialmente remediadas com o filtro de mediana.

As Figs. 6.4 e 6.5 apresentam os ganhos médios obtidos pela super-resolução para as sequências *Poznan Street* e *Barn1*, respectivamente. Estas sequências ilustram o comportamento típico obtido para as demais sequências.

As Figs. 6.6 e 6.8 apresentam detalhes das vistas original, decimada e interpolada e super-resolvidas com os diferentes mapas de profundidade processados, para as sequências *Pantomime* com $M = 4$ e *Venus* com $M = 2$, respectivamente. Os mapas de profundidade processados correspondentes também são apresentados nas Figs. 6.7 e 6.9. Apesar da variedade de resultados, não se observa diferenças significativas entre cada uma das versões super-resolvidas, o que se reflete na proximidade dos valores de MSSIM obtidos para cada versão. É importante notar que o mapa processado $\mathbf{D}_O^B$ possui valores de PSNR e MSSIM mais altos do que os mapas $\mathbf{D}_O^{B,MED}$, $\hat{\mathbf{D}}_{O1}$ e $\hat{\mathbf{D}}_O$, mas a componente de luminância super-resolvida com $\mathbf{D}_O^B$ apresenta valores de PSNR e MSSIM mais baixos. Ou seja, a melhora do mapa de profundidade não necessariamente reflete uma melhora de super-resolução da componentes de luminância da vista correspondente.

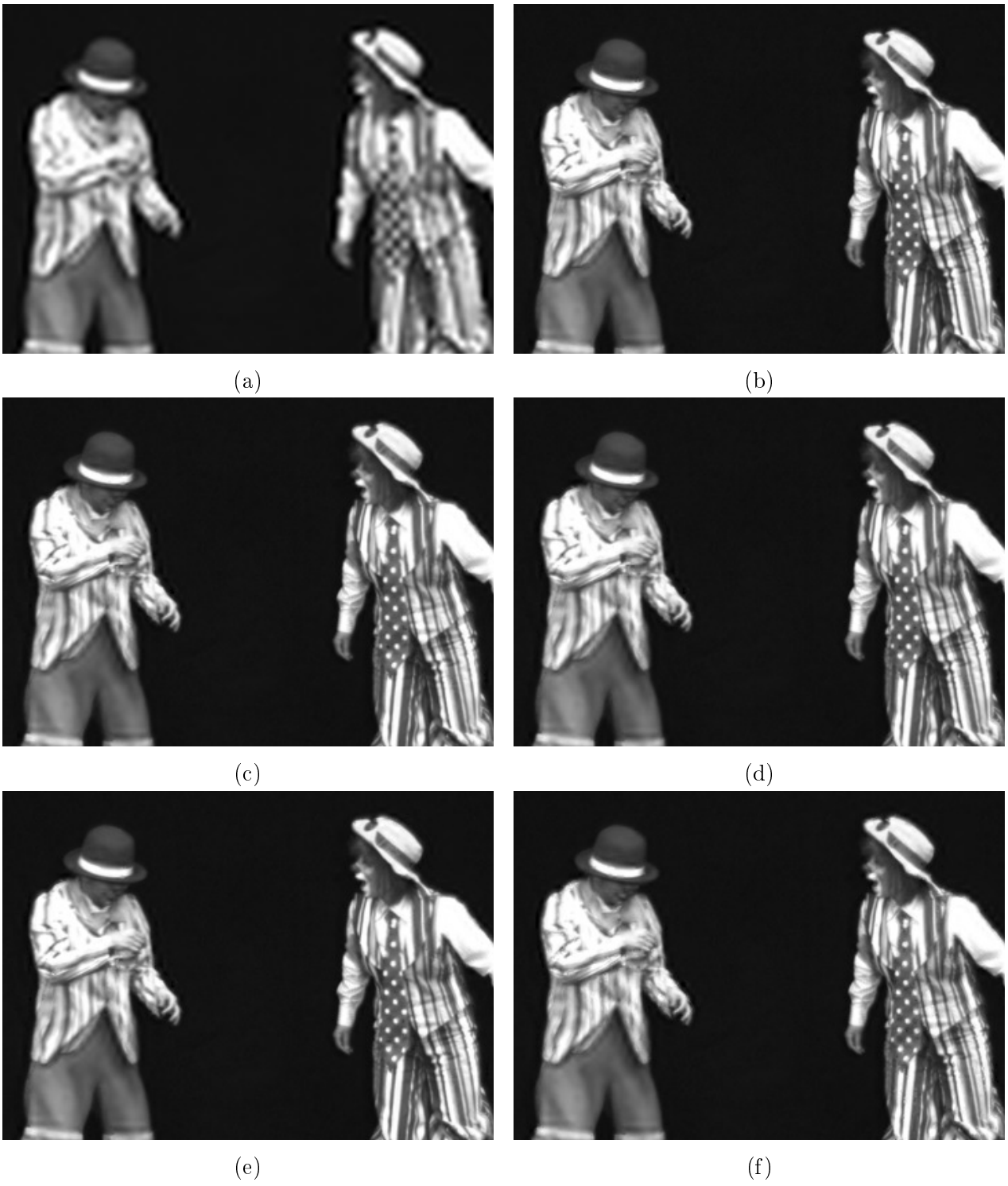


Figura 6.6: Detalhes da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{I}_O^B$ (PSNR = $28,53$ dB / MSSIM×100 = $93,01\%$); (b) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^B$ ($37,64$ dB / $97,33\%$); (c) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^{B,MED}$ ($37,56$ dB / $97,31\%$); (d) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_{O1}$ ($37,74$ dB / $97,35\%$); (e) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_O$ ($37,85$ dB / $97,38\%$); (f) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O$ ($36,05$ dB / $97,09\%$).

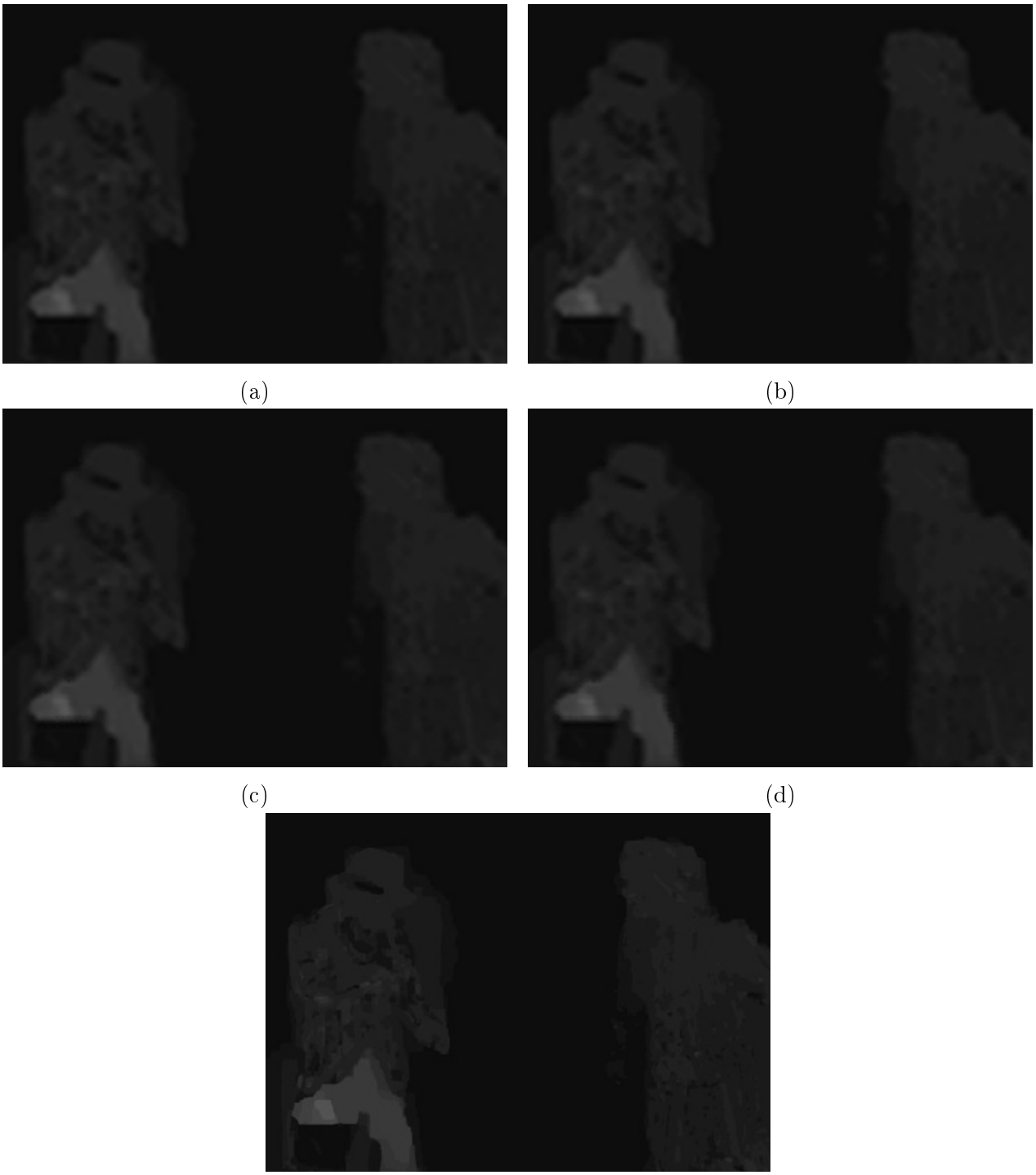


Figura 6.7: Detalhes dos mapas de profundidade da sequência *Pantomime*, vista 39, quadro 0, $M = 4$: (a) $\mathbf{D}_O^B$ (PSNR = 46,85 dB / MSSIM×100 = 98,65%); (b) $\mathbf{D}_O^{B,MED}$ (45,87 dB / 98,59%); (c) $\hat{\mathbf{D}}_{O1}$ (45,62 dB / 98,52%); (d) $\hat{\mathbf{D}}_O$ (45,45 dB / 98,47%); (e) $\mathbf{D}_O$.

6.4.2 Testes com codificação H.264/AVC

Assim como nas Seções 4.4.2 e 5.4.2, os processamentos apresentados neste Capítulo foram aplicados às mesmas sequências após estas serem codificadas em resolução mista, tal como na Fig. 5.1, a fim de avaliar seu desempenho em uma situação prática. Aplicou-se as seguintes condições:

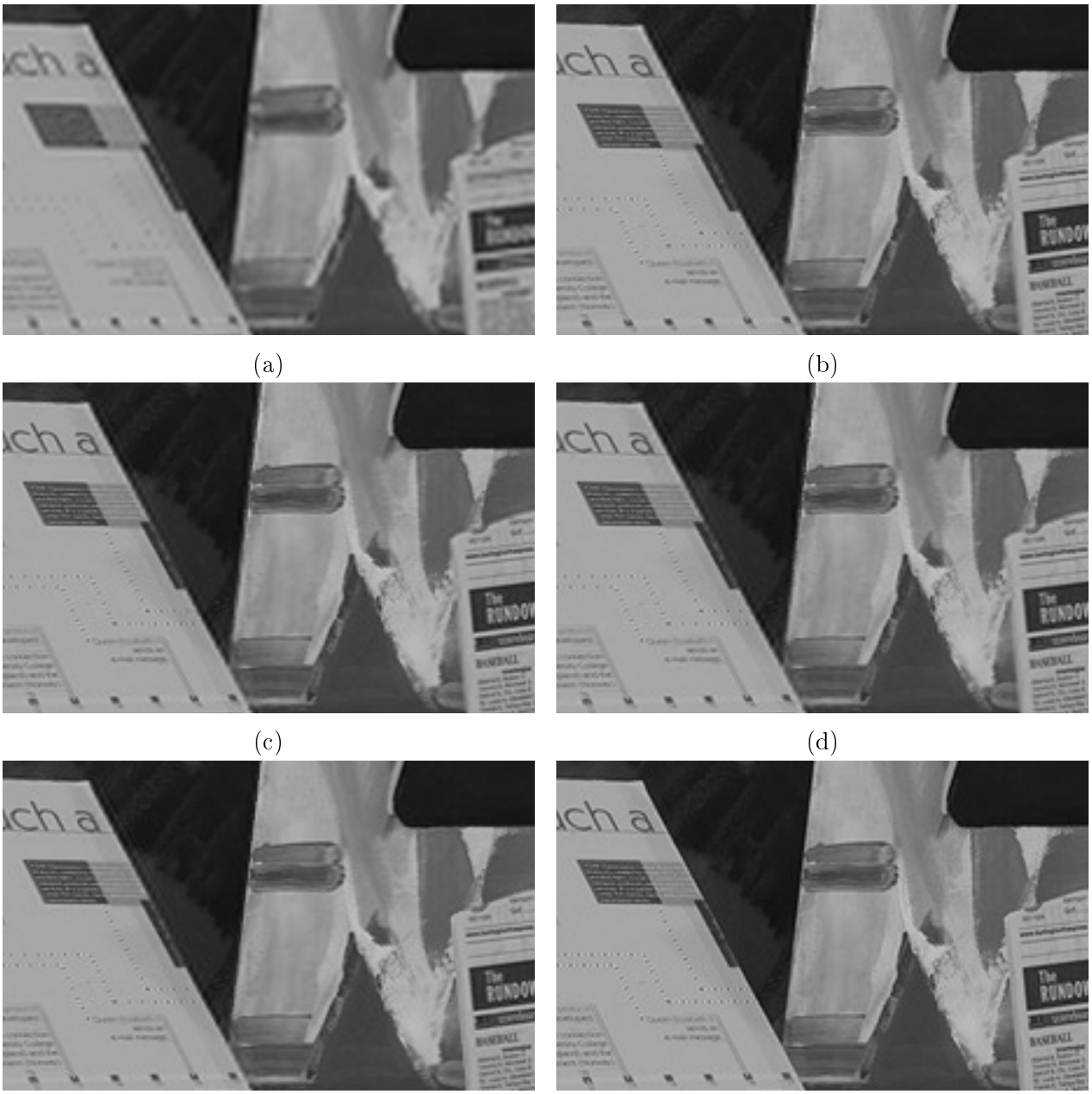
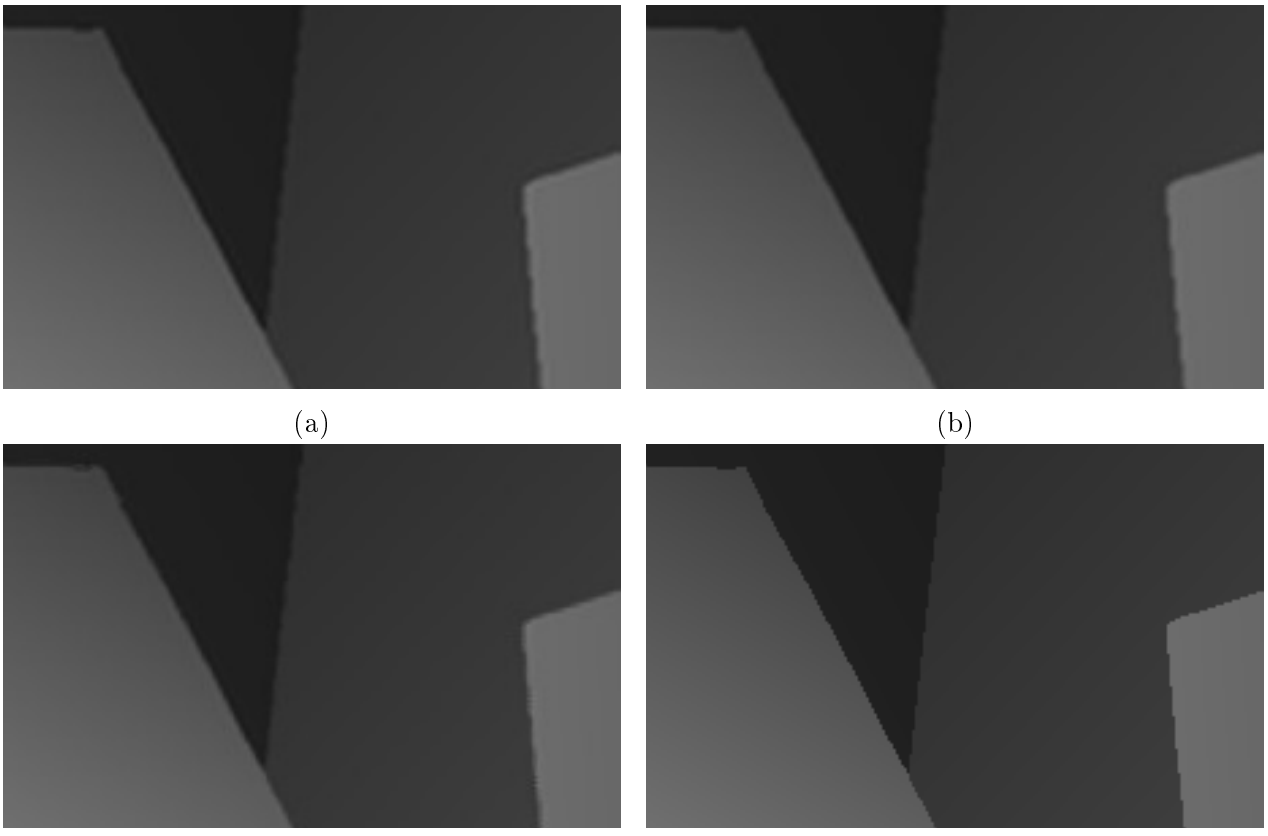


Figura 6.8: Detalhes da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{I}_O^B$ (PSNR = 28,47 dB / MSSIM×100 = 86,24%); (b) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^B$ (35,13 dB / 97,21%); (c) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O^{B,MED}$ (35,17 dB / 97,23%); (d) $\hat{\mathbf{I}}_O$ baseado em $\hat{\mathbf{D}}_O$ com todos os mapas de referência (35,10 dB / 97,22%); (e) $\hat{\mathbf{I}}_O$ baseado em $\mathbf{D}_O$ (35,74 dB / 97,44%); (f) $\mathbf{I}_O$.

- Todas as sequências foram codificadas separadamente utilizando o padrão H.264/AVC em modo Intra, inclusive os mapas de profundidade. Os quadros de $\mathbf{I}_{Ri}$ foram codificados em resolução normal, de acordo com a Tabela 5.1, e os quadros de $\mathbf{I}_O$ foram codificados em resolução inferior ($\mathbf{I}_O^D$), utilizando a decimação por $M = 2$ e $M = 4$. Para gerar os mapas



(a) (b)

(c) (d)

Figura 6.9: Detalhes dos mapas de profundidade da sequência *Venus*, vista 6, quadro 0, $M = 2$: (a) $\mathbf{D}_O^B$ (PSNR = 45,98 dB / MSSIM×100 = 99,38%); (b) $\mathbf{D}_O^{B,MED}$ (35,35 dB / 99,39%); (c) $\hat{\mathbf{D}}_O$ com todos os mapas de referência (35,30 dB / 99,34%); (d) $\mathbf{D}_O$.

$\mathbf{D}_r^B$, $\mathbf{D}_r^{B,MED}$, $\hat{\mathbf{D}}_{r1}$ e $\hat{\mathbf{D}}_r$, codificou-se os mapas $\mathbf{D}_O$ e $\mathbf{D}_r$ em resoluções reduzidas $\mathbf{D}_O^D$ e $\mathbf{D}_r^D$, e os processamentos foram aplicados diretamente a estes mapas após a decodificação.

- Aplicou-se os mesmo valores $QP = \{22, 27, 32, 37\}$ aos parâmetros de escalonamento, inclusive para os mapas de profundidade em resolução completa e reduzida, assim como no Capítulo 5.

- Para todas as sequências testadas, mediu-se a média quadro-a-quadro de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) de $\mathbf{I}_O^B$, que corresponde à versão interpolada de $\mathbf{I}_O^D$ após a codificação, e das versões super-resolvidas de $\mathbf{I}_O^B$ com os diferentes mapas de profundidade processados.

- A taxa considerada para a super-resolução utilizando $\mathbf{D}_r^B$, $\mathbf{D}_r^{B,MED}$, $\hat{\mathbf{D}}_{r1}$ e $\hat{\mathbf{D}}_r$ foi diferente daquela considerada para a super-resolução utilizando $\mathbf{D}_r$. Nos primeiros quatro casos, a taxa considerada foi igual à soma das taxas das versões codificadas de $\mathbf{I}_O^D$, $\mathbf{I}_{Ri}$, $\mathbf{D}_O^D$ e $\mathbf{D}_{Ri}^D$. No quinto caso ($\mathbf{D}_r$), considerou-se a soma das taxas das versões codificadas de $\mathbf{I}_O^D$, $\mathbf{I}_{Ri}$, $\mathbf{D}_O$ e $\mathbf{D}_{Ri}$.

- A partir destes valores de PSNR, MSSIM e taxa, foram medidos os ganhos médios [62] em termos de PSNR e MSSIM para o algoritmo proposto em relação a interpolar os quadros em baixa resolução, utilizando os diferentes mapas de profundidade processados

- O *software* de referência JM 17.2 para o padrão H.264/AVC foi utilizado [61].

- Aplicou-se blocos de tamanho 3×3 para o filtro de mediana utilizado em $\mathbf{D}_r^{B,MED}$, $\hat{\mathbf{D}}_{r1}$ e $\hat{\mathbf{D}}_r$. Para a super-resolução dos quadros $\mathbf{I}_O^D$, aplicou-se blocos de tamanho 16×16 para a combinação de altas frequências.

- Os ganhos médios da super-resolução de $\mathbf{I}_O^D$ com os mapas processados sobre à interpolação de $\mathbf{I}_O^D$ são apresentados no Anexo I, Tabelas I.19 e I.20, que se diferenciam pelo tipo de sequência testada (real ou sintética), pelo tipo de medida de qualidade empregada (médias de PSNR ou MSSIM) e pelo fator de decimação e interpolação utilizado ($M = 2$ ou $M = 4$).

Assim como na Seção anterior, os resultados para $M = 2$ apresentam um gradual aumento no ganho médio de qualidade de $\mathbf{D}_r^B$, para $\mathbf{D}_r^{B,MED}$, para $\hat{\mathbf{D}}_{r1}$, para $\hat{\mathbf{D}}_r$ (resultados 2 a 5). Além disso, $\hat{\mathbf{D}}_r$ oferece aumento no ganho médio de qualidade em relação a $\mathbf{D}_r$ para a maioria dos casos, com a exceção de quatro sequências sintéticas (*Barn1*, *Poster*, *Sawtooth* e *Venus*). Desta forma, o método proposto de super-resolução de mapas de profundidade oferece ganhos médios de qualidade tanto em relação a codificar estes mapas em resolução completa quanto em relação a codificá-los em baixa resolução e os interpolar.

Para $M = 4$, o método proposto oferece os melhores resultados para as sequências reais, e que $\mathbf{D}_r$ oferece os melhores resultados para as sequências sintéticas. O fator de decimação e interpolação elevado para os mapas de profundidade impede a obtenção de ganhos mais significativos para as sequências sintéticas.

As Tabelas 6.1 e 6.2 apresentam os tempos médios de codificação e de super-resolução para os testes desta Seção. Observa-se que a codificação dos mapas de profundidade em baixa resolução oferece redução no tempo de codificação, aliado a ganhos médios de qualidade. Em compensação, aumentam gradualmente os tempos de processamento no lado do decodificador, devido ao acréscimo do processo de super-resolução dos mapas de profundidade, culminando no maior tempo de processamento para $\hat{\mathbf{D}}_r$. E importante notar que o código para todos os métodos e o modo de operação do sistema operacional não foram otimizados para desempenho, e que os tempos reportados são tempos médios, considerando os quatro valores aplicados ao parâmetro de escalonamento.

As Figs. 6.10 a 6.13 apresentam o desempenho da super-resolução baseada nos métodos propostos de pré-processamento de mapas para as sequências *Pantomime*, *Cafe*, *Venus* e *Bull*. É possível ver para as sequências *Pantomime* e *Cafe* (Figs. 6.10 e 6.11) que o uso dos mapas em resolução normal oferece um pior desempenho em termos de taxa e distorção, tanto para $M = 2$ quanto para $M = 4$. Já os desempenhos com os mapas processados $\mathbf{D}_r^B$, $\mathbf{D}_r^{B,MED}$, $\hat{\mathbf{D}}_{r1}$ e $\hat{\mathbf{D}}_r$ são bastante semelhantes.

Tabela 6.1: Tempo médio de codificação e de super-resolução para os testes de super-resolução com os mapas de profundidade processados, em segundos, para $M = 2$.

Nome	**Codificação** (s)		**Super-resolução** (s)				
	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
Sequências reais							
Ballet	**54,30**	45,69	121,14	128,81	139,15	**198,60**	198,60
Breakdancers	**54,09**	45,61	122,00	129,72	139,74	**199,11**	199,08
Cafe	**129,20**	115,21	310,67	328,57	353,63	492,71	**493,20**
Pantomime	**77,84**	68,58	184,49	195,82	211,07	294,46	**294,60**
Lovebird1	**52,32**	45,63	121,19	128,92	138,20	**192,58**	192,55
Newspaper	**53,90**	46,24	116,83	124,55	134,94	192,48	**192,51**
Poznan Street	**144,18**	122,23	319,50	338,55	363,72	510,46	**511,00**
Sequências sintéticas							
Barn1	**0,40**	0,38	0,53	0,59	0,64	**0,72**	**0,72**
Barn2	**0,40**	0,39	0,52	0,57	0,63	**0,71**	**0,71**
Bull	**0,40**	0,38	0,53	0,60	0,65	**0,71**	**0,71**
Map	**0,22**	0,22	0,21	0,25	0,27	**0,28**	**0,28**
Poster	**0,40**	0,39	0,55	0,60	0,66	**0,73**	**0,73**
Sawtooth	**0,41**	0,39	0,55	0,61	0,66	**0,73**	**0,73**
Venus	**0,41**	0,39	0,55	0,61	0,66	**0,73**	**0,73**
Cones	**0,44**	0,41	0,57	0,63	0,69	**0,76**	**0,76**
Teddy	**0,44**	0,41	0,56	0,63	0,68	**0,75**	**0,75**
Room3D	**31,06**	26,05	51,96	56,76	61,86	**66,66**	**66,66**

Tabela 6.2: Tempo médio de codificação e de super-resolução para os testes de super-resolução com os mapas de profundidade processados, em segundos, para $M = 4$.

Nome	**Codificação** (s)		**Super-resolução** (s)				
	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
Sequências reais							
Ballet	**53,41**	42,01	121,14	128,81	139,15	**198,60**	198,60
Breakdancers	**53,57**	42,09	122,00	129,72	139,74	**199,11**	199,08
Cafe	**126,12**	105,57	310,67	328,57	353,63	492,71	**493,20**
Pantomime	**76,29**	64,28	184,49	195,82	211,07	294,46	**294,60**
Lovebird1	**51,24**	42,41	121,19	128,92	138,20	**192,58**	192,55
Newspaper	**52,78**	42,70	116,83	124,55	134,94	192,48	**192,51**
Poznan Street	**141,10**	111,10	319,50	338,55	363,72	510,46	**511,00**
Sequências sintéticas							
Barn1	**0,41**	0,36	0,53	0,59	0,64	**0,72**	**0,72**
Barn2	**0,39**	0,37	0,52	0,57	0,63	**0,71**	**0,71**
Bull	**0,40**	0,37	0,53	0,60	0,65	**0,71**	**0,71**
Map	**0,21**	0,21	0,21	0,25	0,27	**0,28**	**0,28**
Poster	**0,40**	0,37	0,55	0,60	0,66	**0,73**	**0,73**
Sawtooth	**0,40**	0,36	0,55	0,61	0,66	**0,73**	**0,73**
Venus	**0,41**	0,36	0,55	0,61	0,66	**0,73**	**0,73**
Cones	**0,43**	0,38	0,57	0,63	0,69	**0,76**	**0,76**
Teddy	**0,43**	0,37	0,56	0,63	0,68	**0,75**	**0,75**
Room3D	**30,20**	23,95	51,96	56,76	61,86	**66,66**	**66,66**

Para a sequência *Venus* (Fig. 6.12), os mapas em resolução normal oferecem desempenho superior para as taxas mais altas, especialmente para $M = 4$, aonde o ganho médio ao se utilizar o mapa $\mathbf{D}_r$ é $0,65$ dB a mais do que utilizar o mapa $\mathbf{D}_r^B$ ($5,20$ dB e $4,55$ dB, respectivamente). Finalmente, para a sequência *Bull* (Fig. 6.13), percebe-se que os mapas $\hat{\mathbf{D}}_O$, $\hat{\mathbf{D}}_O^{MED}$, $\mathbf{D}_O$ ou $\mathbf{D}_O^{MED}$ também oferecem desempenhos muito próximos.

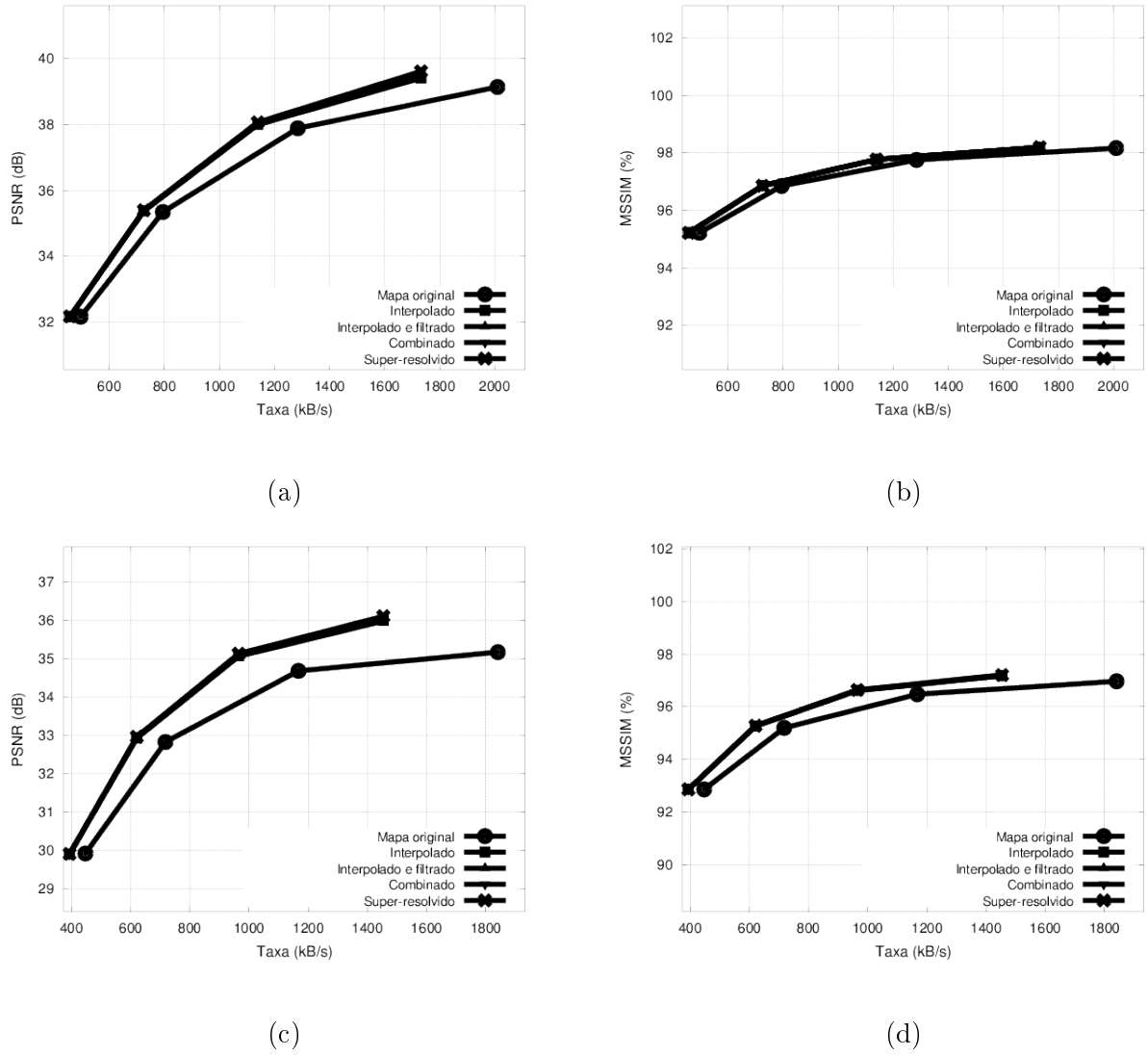


Figura 6.10: Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Pantomime*, vista 39, quadro 1: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$.

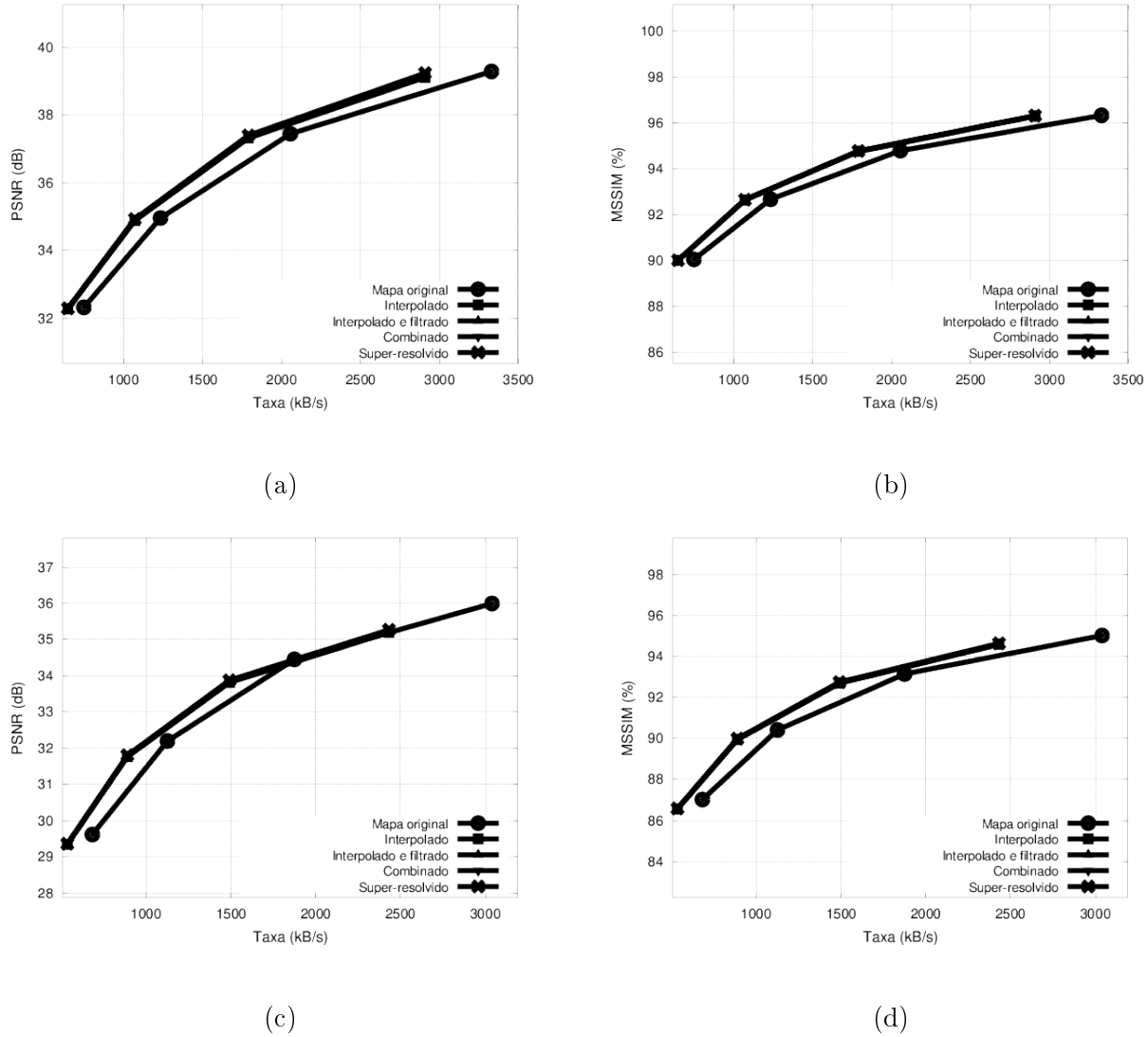


Figura 6.11: Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Cafe*, vista 3: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$.

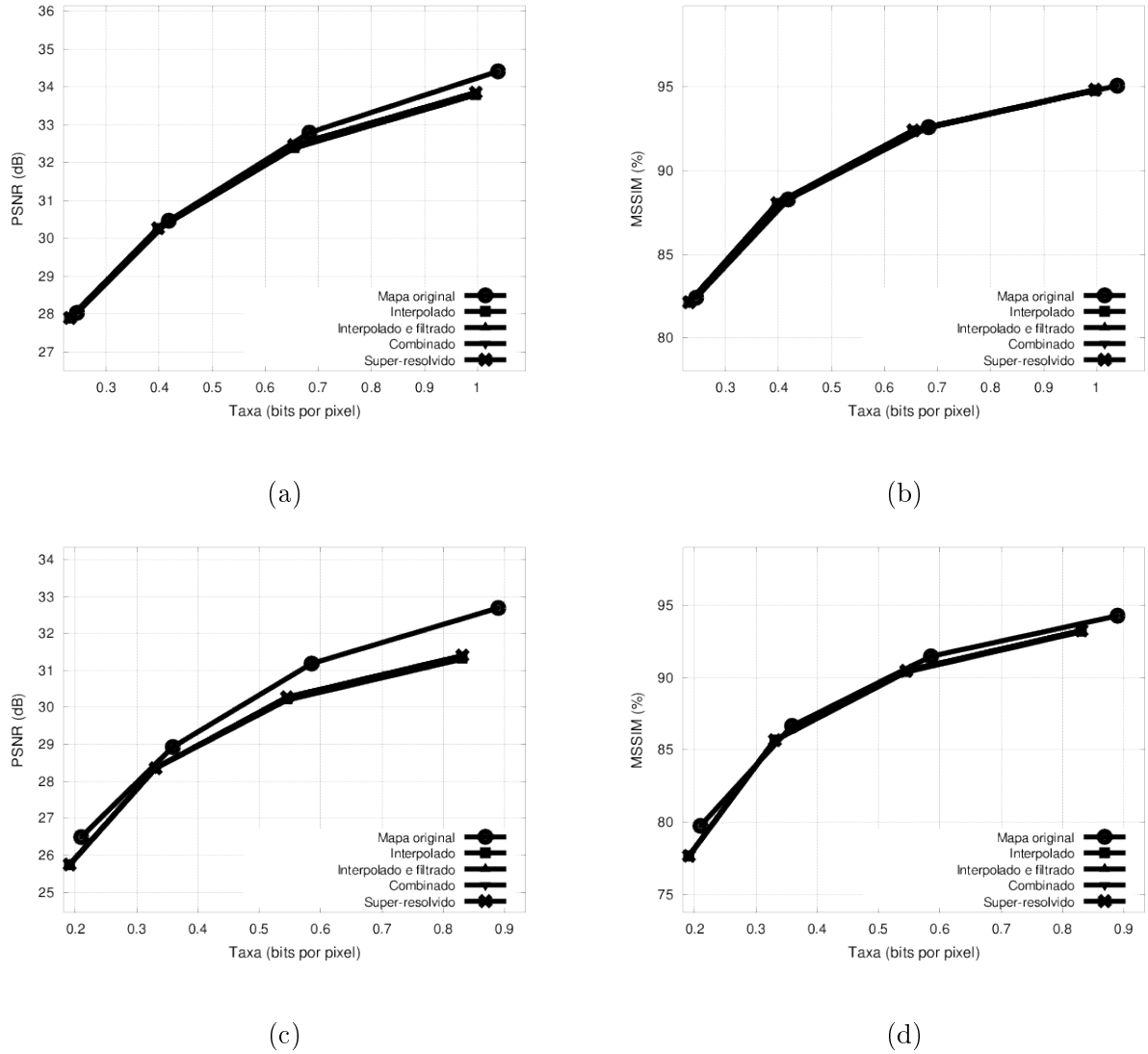
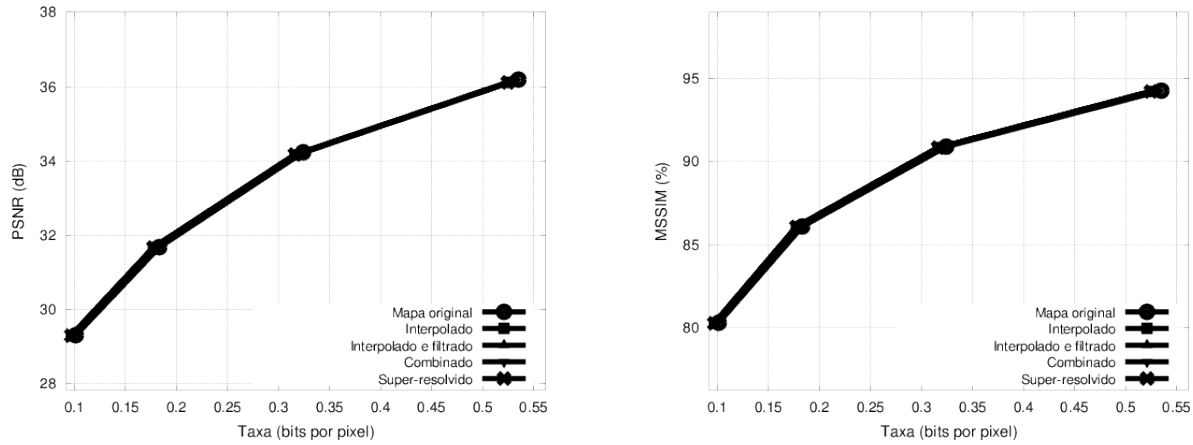
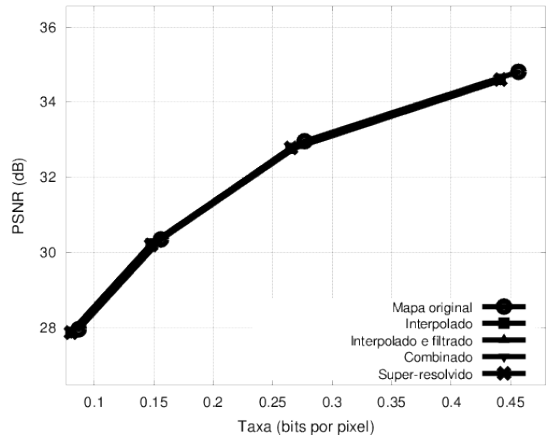


Figura 6.12: Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Venus*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$.

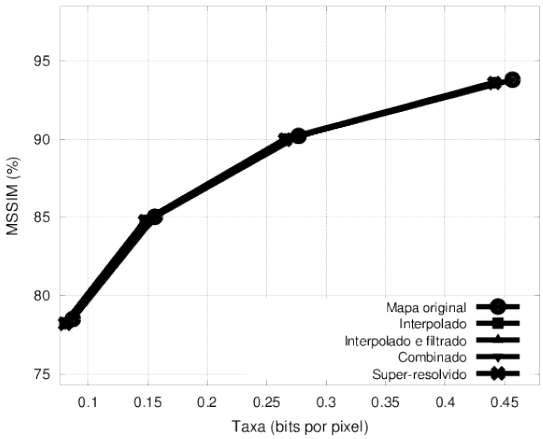


(a)

(b)



(c)



(d)

Figura 6.13: Desempenho em termos de taxa e distorção para o método de super-resolução utilizando os pré-processamentos de mapas de profundidade propostos para a sequência *Bull*, vista 6, quadro 0: (a) taxa e PSNR, $M = 2$; (b) taxa e MSSIM, $M = 2$; (c) taxa e PSNR, $M = 4$; (d) taxa e MSSIM, $M = 4$.

Capítulo 7

Conclusões

Foram desenvolvidos nesta tese três métodos de super-resolução de sequências de múltiplas vistas em resolução mista. Cada método adequa-se a uma arquitetura diferente, dada a grande variedade de sistemas que podem utilizar esse tipo de sequência, não havendo um formato único e universal. A seguir, serão tecidas as conclusões para cada um dos métodos.

7.1 Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade

O primeiro dos métodos oferece aumento de qualidade de quadros em resolução reduzida, tendo por base quadros de vistas ou instantes próximos em resolução completa. A partir de cada um destes quadros, são extraídas informações de alta frequência utilizando métodos de estimação e compensação de movimento e/ou disparidade, de forma a gerar duas versões do quadro de alta frequência para cada referência disponível. Em seguida, todas as versões geradas são combinadas em um único quadro de alta frequência.

Diversos testes com sequências reais e sintéticas demonstram que o método apresentado oferece ganhos substanciais de qualidade aos quadros em resolução reduzida, tanto para quadros de referência originais quanto para quadros corrompidos pelo escalonamento por codificação. Constata-se que os ganhos do algoritmo são maiores quando se possui o maior número de referências permitido pela arquitetura em consideração.

Em relação aos quadros individuais, os resultados indicam que os quadros pertencentes à mesma vista sendo super-resolvida geralmente resultam em melhores estimativas de alta frequência do que os quadros de vistas adjacentes. Dentre as duas versões obtidas a partir de uma dada referência, a primeira é geralmente melhor do que a segunda, e a combinação de ambas oferece os melhores resultados.

Os resultados com quadros codificados pelo padrão H.264/AVC apresentam ganhos de qualidade, ainda que inferiores aos resultados obtidos com os quadros de referência originais. Este era um resultado esperado, dado que a codificação corrompe os quadros de referência através do

processo de escalonamento. À medida que se aumenta o nível de escalonamento dos quadros, pior se torna o desempenho do algoritmo proposto, dada a degradação das imagens de referência.

7.2 Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade

O primeiro método assume que os mapas de profundidade de cada vista não estão disponíveis, sendo necessário estimar de alguma maneira os melhores candidatos de alta frequência para cada bloco do quadro a ser super-resolvido. No segundo método, assume-se que estes mapas estão disponíveis, de forma que é possível obter correspondências entre o quadro em baixa resolução e os quadros de referência. Assim, obtem-se para cada *pixel* do quadro em baixa resolução os candidatos de alta frequência, ao invés de obter esta informação em blocos.

Mapas de profundidade são suscetíveis a erros, devidos a oclusões, imprecisões de representação e, no caso de codificação com perdas, escalonamento, de forma que é possível obter informações inválidas a partir destes mapas. Sendo assim, o segundo método confere a consistência entre os mapas disponíveis antes de realizar a projeção das informações de alta frequência, para evitar o acréscimo de informações errôneas à imagem sendo super-resolvida. Duas versões de alta frequência são geradas a partir destes valores válidos, e de forma similar ao primeiro método combina-se as diversas informações de alta frequência em um quadro final super-resolvido.

Assim como no primeiro método, verificou-se ganho substancial de qualidade para o método proposto, testando sequências reais e sintéticas originais e codificadas com o padrão H.264/AVC. A combinação de todas as referências disponíveis resultou nos maiores ganhos, e o algoritmo obteve desempenho inferior à medida que se aumentou o ruído de codificação por escalonamento.

Com relação ao uso de uma única vista de referência, verificou-se para as sequências reais que a combinação das duas versões de alta frequência geradas oferece ganho em relação a cada versão individualmente. Isto indica que os mapas de profundidade utilizados contêm erros, que são suprimidos quando se confere a consistência entre os mapas. Já para as sequências sintéticas, não se verificou tal ganho, dado que os mapas de profundidade correspondentes são muito mais precisos.

7.3 Super-resolução de mapas de profundidade em baixa resolução

Baseado em características intrínsecas aos mapas de profundidade, o terceiro método oferece a super-resolução de mapas de profundidade codificados em baixa resolução, que são depois utilizados pelo segundo método de super-resolução para obter informação de alta frequência dos quadros codificados em baixa resolução. A codificação dos mapas em baixa resolução baseia-se no fato de que eles costumam ser compostos por grandes áreas suaves demarcadas por transições abruptas entre os objetos, de forma que decimação dos mapas antes da codificação reduz a taxa

de transmissão e a complexidade de codificação, e a super-resolução mútua destes mapas recupera informação de alta frequência no lado do decodificador.

A super-resolução dos mapas de profundidade baseia-se em métodos de super-resolução por combinação de múltiplas imagens, e consiste em três etapas: projeção dos mapas, preenchimento de buracos e filtragem. A primeira etapa acrescenta ao mapa a ser super-resolvido informação de mapas adjacentes, de acordo com a consistência entre eles. A segunda etapa corrige os espaços não-preenchidos pela etapa anterior, e a terceira etapa processa os mapas através da filtragem de mediana, para eliminar eventuais ruídos presentes nas áreas suaves sem borrar as bordas dos objetos.

Testes realizados com as mesmas sequências reais e sintéticas, com e sem codificação H.264/AVC, demonstram a eficiência do método proposto, em que a arquitetura proposta oferece reduções de taxa e de complexidade no codificador. Além disso, o método proposto de super-resolução para mapas de profundidade oferece ganho médio de qualidade nas vistas super-resolvidas em relação a transmitir os mapas em resolução completa e também em relação a transmitir os mapas em baixa resolução e os interpolar.

7.4 Comparações entre arquiteturas

Como indicado anteriormente, cada um dos métodos apresentados presta-se a uma arquitetura diferente, atendendo a requisitos de taxa e complexidade. Nesta Seção, comparam-se as arquiteturas consideradas.

A primeira arquitetura de resolução mista é voltada para a redução de complexidade em codificadores que não possuem poder computacional para calcular mapas de profundidade, ao contrário das outras duas arquiteturas. Sendo assim, a primeira arquitetura oferece maior rapidez de codificação e menor taxa de transmissão, visto que não é necessário calcular e nem codificar os mapas de profundidade.

O processamento do lado do decodificador é provavelmente mais complexo para a primeira arquitetura, devido à ausência de mapas de profundidade. O primeiro método de super-resolução torna necessária a estimação de movimento e/ou disparidade do lado do decodificador, enquanto os outros dois métodos assumem que os mapas de profundidade podem oferecer correspondências fidedignas entre as sequências em alta e em baixa resolução. Sendo assim, estas correspondências devem ser calculadas no primeiro método de super-resolução, enquanto os outros dois métodos possuem as correspondências definidas, ao custo de maior taxa de transmissão e de maior complexidade no codificador, devido aos mapas de profundidade.

Como os mapas de profundidade são codificados em baixa resolução na terceira arquitetura, e são mutuamente super-resolvidos no terceiro método proposto, ocorre uma redução de taxa e de complexidade no codificador e um aumento de complexidade no decodificador da segunda para a terceira arquitetura.

7.5 Considerações finais

A presente tese apresentou e desenvolveu três métodos de super-resolução de sequências de múltiplas vistas em resolução mista. Cada método se adequa a uma arquitetura diferente, de acordo com as necessidades do sistema de transmissão. Para futuros trabalhos, sugere-se o estudo de novos métodos de estimativa de mapas de profundidade, a criação de métodos que levem em conta o nível de escalonamento empregado sobre as imagens, e o uso do valor de SSIM ao invés da SSD para a estimação de movimento/disparidade utilizada pelo primeiro método de super-resolução.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] VETRO, A.; WIEGAND, T.; SULLIVAN, G. J. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proceedings of the IEEE*, v. 99, n. 4, p. 626–642, 2011.

[2] SCHARSTEIN, D.; SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, v. 47, p. 7–42, 2002.

[3] VETRO, A.; YEA, S.; SMOLIC, A. Towards a 3D video format for auto-stereoscopic displays. In: *SPIE Conference on Applications of Digital Image Processing XXXI*. [S.l.: s.n.], 2008.

[4] HUANG, Y.-W. et al. Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC. *IEEE Transactions on Circuits and Systems Video Technology*, v. 16, n. 4, p. 507–522, 2006.

[5] FEHN, C. et al. Asymmetric coding of stereoscopic video for transmission over T-DMB. In: *3DTV Conference*. [S.l.: s.n.], 2007. p. 1 –4.

[6] BRUST, H.; MUELLER, G. T. K.; WIEGAND, T. Mixed resolution coding with inter view prediction for mobile 3DTV. In: *3DTV Conference*. [S.l.: s.n.], 2010. p. 1 –4.

[7] EKMEKCIOGLU, E.; WORRALL, S.; KONDOZ, A. Utilisation of downsampling for arbitrary views in multi-view video coding. *IEEE Transactions on Circuits and Systems Video Technology*, v. 44, n. 5, p. 339–340, 2008.

[8] AKSAY, A. et al. Temporal and spatial scaling for stereoscopic video compression. In: *3DTV Conference*. [S.l.: s.n.], 2006.

[9] CHEN, Y. et al. Regionally adaptive filtering for asymmetric stereoscopic video coding. In: *IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2009. p. 2585–2588.

[10] SAWHNEY, H. et al. Hybrid stereo camera: an IBR approach for synthesis of very high resolution stereoscopic image sequences. In: *Conference on Computer Graphics and Interactive Techniques*. [S.l.: s.n.], 2001. p. 451–460.

[11] JULESZ, B. *Foundations of Cyclopean Perception*. [S.l.]: University of Chicago Press, 1971.

[12] TAM, W. Image and depth quality of asymmetrically coded stereoscopic video for 3D-TV. In: *JVT-W094*. [S.l.: s.n.], 2007.

[13] AFLAKI, P. et al. Subjective study on compressed asymmetric stereoscopic video. In: *IEEE International Conference on Image Processing*. [S.l.: s.n.], 2010. p. 4021–4024.

[14] SAYGILI, G.; GURLER, C.; TEKALP, A. Quality assessment of assymmetric stereo video coding. In: *IEEE International Conference on Image Processing*. [S.l.: s.n.], 2010. p. 4009–4012.

[15] TANIMOTO, M. Free viewpoint systems. In: SCHREER, O.; KAUFF, P.; SIKORA, T. (Ed.). *3D Video Communication Algorithms - Concepts and Real Time Systems in Human Centred Communication*. [S.l.]: John Wiley Sons, Ltd., 2005. p. 55–74.

[16] GARCIA, D. C. et al. Mixed resolution framework for distributed multiview coding. In: *IST/SPIE Symposium on Electronic Imaging, Multimedia on Mobile Devices*. [S.l.: s.n.], 2010.

[17] GARCIA, D. C.; DOREA, C.; QUEIROZ, R. L. de. Super resolution for multiview images using depth information. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 22, n. 9, p. 1249–1256, Setembro 2012.

[18] GARCIA, D. C.; DOREA, C. C.; QUEIROZ, R. L. de. Super-resolution for multiview images using depth information. In: *IEEE International Conference on Image Processing*. [S.l.: s.n.], 2010. p. 1793–1796.

[19] ZITNICK, C. et al. High-quality video view interpolation using a layered representation. In: *Conference on Computer Graphics and Interactive Techniques*. [S.l.: s.n.], 2004. p. 600–608.

[20] RICHARDSON, I. *The H.264 Advanced Video Compression Standard*. [S.l.]: John Wiley Sons, Ltd., 2010.

[21] ITU-R. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. *ITU-R Recommendation BT.601-6*, 2007.

[22] DINIZ, P. S. R.; SILVA, E. A. B. da; NETTO, S. L. *Digital Signal Processing: System Analysis and Design*. [S.l.]: Cambridge, 2002.

[23] BAKER, S.; KANADE, T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 9, p. 1167–1183, 2002.

[24] PARK, S. C.; PARK, M. K.; KANG, M. G. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, v. 20, n. 3, p. 21–36, 2003.

[25] FARSIU, S. et al. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, v. 14, p. 47–57, 2004.

[26] FREEMAN, W.; JONES, T.; PASZTOR, E. Example-based super-resolution. *IEEE Computer Graphics and Applications*, v. 22, n. 2, p. 56–65, 2002.

[27] FIELD, D. What is the goal of sensory coding? *Neural Computation*, v. 6, n. 4.

[28] SCHWARTZ, O.; SIMONCELLI, E. P. Natural signal statistics and sensory gain control. *Nature Neuroscience*, v. 4, p. 819–825, 2001.

[29] ITU-T and ISO/IEC JTC 1. Advanced video coding for generic audiovisual services. *ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)*, 2010.

[30] J.LUBINAND; FIBUSH, D. Sarnoff JND vision model. *T1 A1.5 Working group Document, T1 Standards Committee*, 1997.

[31] TAN, T.; SULLIVAN, G.; WEDI, T. Recommended simulation common conditions for coding efficiency experiments. In: *VCEG-AA10*. [S.l.: s.n.], 2005.

[32] WANG, Z. et al. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, 2004.

[33] HARTLEY, R.; ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. [S.l.]: Cambridge University Press, 2003.

[34] MERKLE, P. et al. Multi-view video plus depth representation and coding. In: *IEEE International Conference on Image Processing*. [S.l.: s.n.], 2007. p. I–201–I–204.

[35] MÜLLER, K.; MERKLE, P.; WIEGAND, T. 3-D video representation using depth maps. *Proceedings of the IEEE*, v. 99, n. 4, p. 643–656, 2011.

[36] VETRO, A. et al. 3D-TV content storage and transmission. *IEEE Transactions on Broadcasting*, v. 57, n. 2, p. 384–394, 2011.

[37] R.; BAKER, B. H.; HANNAH, M. The JISCT stereo evaluation.

[38] HANNAH, M. *Computer Matching of Areas in Stereo Images*. Tese (Doutorado) — Stanford University, 1974.

[39] BARNARD, S. T. Stochastic stereo matching over scale. *International Journal of Computer Vision*, v. 3, n. 1, p. 17–32, 1989.

[40] MARROQUIN, J.; MITTER, S.; POGGIO, T. *Probabilistic Solution of Ill-Posed Problems in Computational Vision*. [S.l.], 1987.

[41] SCHARSTEIN, D.; SZELISKI, R. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, v. 28, n. 2, p. 155–174, 1998.

[42] BOYKOV, Y.; VEKSLER, O.; ZABIH, R. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[43] LEE, Y. et al. CE11: Illumination compensation. In: *JVT-U052*. [S.l.: s.n.], 2006.

[44] HUR, J.; CHO, S.; LEE, Y. Adaptive local illumination change compensation method for H.264/AVC-based multiview video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2007.

[45] LAI, P. et al. Adaptive reference filtering for MVC. In: *JVT-W065*. [S.l.: s.n.], 2007.

[46] LAI, P. et al. Focus mismatches in multiview systems and efficient adaptive reference filtering for multiview video coding. In: *Visual Communications and Image Processing*. [S.l.: s.n.], 2007.

[47] MARTINIAN, E. et al. View synthesis for multiview video compression. In: *Picture Coding Symposium*. [S.l.: s.n.], 2006.

[48] YEA, S.; VETRO, A. View synthesis prediction for multiview video coding. *Image Commun.*, v. 24, n. 1-2, p. 89–100, 2009.

[49] KITAHARA, M. et al. Multi-view video coding using view interpolation and reference picture selection. In: *IEEE International Conference on Multimedia and Expo*. [S.l.: s.n.], 2006. p. 97–100.

[50] JEON, H. K. Y.; JEON, B. MVC motion skip mode. In: *JVT-W081*. [S.l.: s.n.], 2007.

[51] KOO, H.; JEON, Y.; JEON, B. Motion information inferring scheme for multi-view video coding. *IEICE Transactions on Communications*, p. 1247–1250, 2008.

[52] DOMAŃSKI, M. et al. Poznań multiview video test sequences and camera parameters. In: *ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050*. [S.l.: s.n.], 2009.

[53] KIM, W. et al. Depth map coding with distortion estimation of rendered view. In: *Proceedings of the SPIE Visual Information Processing and Communication*. [S.l.: s.n.], 2010.

[54] MÜLLER, K. et al. View synthesis for advanced 3D video systems. *EURASIP Journal on Image and Video Processing*, v. 2008, p. 1–11, 2008.

[55] DARIBO, I.; TILLIER, C.; PESQUET-POPESCU, B. Adaptive wavelet coding of the depth map for stereoscopic view synthesis. In: *IEEE 10th Workshop on Multimedia Signal Processing*. [S.l.: s.n.], 2008.

[56] MAITRE, M.; DO, M. Shape-adaptivewavelet encoding of depth maps. In: *Picture Coding Symposium*. [S.l.: s.n.], 2009.

[57] MERKLE, P. et al. The effects of multiview depth video compression on multiview rendering. *Sig. Proc.: Image Comm.*, v. 24, n. 1-2, p. 73–88, 2009.

[58] OH, K. et al. Depth reconstruction filter and down/up sampling for depth coding in 3D video. *IEEE Signal Processing Letters*, 2009.

[59] ISO/IEC JTC1/SC29/WG11. Description of exploration experiments in 3D video coding. In: *Doc. N9783*. [S.l.: s.n.], 2008.

[60] NAGOYA University FTV test sequences. nov 2011. Disponível em: <http://www.tanimoto.nuee.nagoya-u.ac.jp/>.

[61] JM H.264/AVC reference software. nov 2011. Disponível em: <http://iphome.hhi.de/suehring/tml/>.

[62] BJONTEGAARD, G. Calculation of average PSNR differences between RD-curves. In: *VCEG-M33*. [S.l.: s.n.], 2001.

[63] SHEPARD, D. A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 23rd ACM national conference*. [S.l.: s.n.], 1968. p. 517–524.

ANEXOS

I. RESULTADOS EXPERIMENTAIS

Neste Anexo, apresenta-se de forma tabulada os resultados obtidos nos Capítulos 4, 5 e 6.

I.1 Super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade

I.1.1 Resultados sem codificação

As Tabelas I.1 a I.4 apresentam os ganhos sem codificação da super-resolução proposta no Capítulo 4 em relação a interpolar os quadros em baixa resolução, baseado nas médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. Os melhores resultados para cada sequência são indicados por números em negrito.

Nas Tabelas I.1 e I.2, os ganhos obtidos quando se utiliza somente uma referência são apresentados da terceira à sexta coluna. Para cada entrada nestas colunas, tem-se três linhas: na primeira, utiliza-se $\mathbf{I}_{Oi}^{A\,\prime} + \mathbf{I}_{O}^{B}$; na segunda, utiliza-se $\mathbf{I}_{Oi}^{A\,\prime\prime} + \mathbf{I}_{O}^{B}$; e na terceira linha, combina-se $\mathbf{I}_{Oi}^{A\,\prime}$ e $\mathbf{I}_{Oi}^{A\,\prime\prime}$ em um único quadro. A última coluna apresenta os resultados obtidos aplicando a Eq. (4.6) com todas as referências disponíveis.

A terceira coluna das Tabelas I.1 e I.2, ($Oi = O1$) indica os ganhos obtidos pela super-resolução em relação à interpolação quando se utiliza somente um quadro da vista anterior no mesmo instante de tempo como referência ($\{\text{vista,tempo}\} = \{j-1, k\}$), de acordo com os quadros indicados na Tabela 4.1. A quarta coluna ($O2$) corresponde ao uso de somente um quadro da vista posterior no mesmo instante ($\{j+1, k\}$), a quinta coluna ($O3$) corresponde ao instante anterior na mesma vista ($\{j, k-1\}$), e a sexta coluna ($O4$), ao instante posterior na mesma vista ($\{j, k+1\}$).

Tabela I.1: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**												
		$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O2}$	$\mathbf{I}_{O3}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O3}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O3}$	$\mathbf{I}_{O4}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O4}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O4}$	$\hat{\mathbf{I}}_O$
PSNR (dB)														
Ballet	33,49	0,26	-1,70	0,69	1,15	0,81	1,67	9,54	1,37	9,86	9,48	1,38	9,81	**12,25**
Breakdancers	39,16	0,01	-0,31	0,51	0,40	-0,26	0,78	0,64	0,46	1,14	0,62	0,43	1,11	**3,85**
Cafe	36,82	2,01	1,45	2,53	2,78	2,92	3,36	6,75	4,94	7,18	6,77	4,91	7,19	**9,42**
Pantomime	34,52	3,94	1,08	4,39	2,50	1,16	3,31	2,35	0,42	2,79	2,31	0,40	2,77	**7,40**
Lovebird1	34,27	1,63	0,42	2,12	2,07	0,79	2,59	12,41	9,17	12,65	12,37	9,10	12,62	**14,63**
Newspaper	34,00	-0,55	0,27	0,22	0,42	0,36	1,10	10,18	6,81	10,56	10,18	6,81	10,57	**13,15**
Poznan Street	33,16	2,21	0,61	2,70	2,60	0,93	2,99	7,18	4,65	7,42	7,18	4,63	7,42	**10,01**
MSSIM (%)														
Ballet	94,46	-0,04	-1,99	0,38	-0,21	-0,53	0,40	3,28	0,70	3,41	3,27	0,72	3,40	**4,13**
Breakdancers	96,49	-1,44	-1,57	-0,91	-1,20	-1,50	-0,77	-0,74	-1,08	-0,34	-0,74	-1,08	-0,34	**1,05**
Cafe	97,38	-0,16	-0,53	0,04	-0,10	-0,31	0,11	0,71	-0,04	0,79	0,71	-0,05	0,79	**1,41**
Pantomime	97,79	0,21	-0,40	0,51	-0,02	-0,41	0,35	0,09	-0,49	0,38	0,08	-0,49	0,37	**1,15**
Lovebird1	94,91	1,95	0,05	2,26	2,16	0,65	2,42	4,49	3,77	4,51	4,49	3,76	4,51	**4,70**
Newspaper	95,09	-2,11	-0,70	-1,08	-0,17	-0,82	0,33	3,85	2,08	3,91	3,85	2,06	3,91	**4,26**
Poznan Street	91,33	2,13	-0,97	2,69	2,48	-0,59	2,92	5,63	3,05	5,74	5,63	3,04	5,74	**6,73**

Tabela I.2: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**												
		$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O2}$	$\mathbf{I}_{O3}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O3}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O3}$	$\mathbf{I}_{O4}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O4}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O4}$	$\hat{\mathbf{I}}_O$
PSNR (dB)														
Ballet	29,62	0,10	-2,13	0,37	0,53	-2,27	0,79	9,50	-0,73	9,86	9,47	-0,74	9,83	**13,05**
Breakdancers	33,60	1,07	-0,83	1,47	1,79	-0,81	2,12	1,28	-0,31	1,78	1,35	-0,32	1,82	**5,51**
Cafe	30,23	3,08	0,36	3,51	3,61	1,21	4,04	8,69	2,88	9,12	8,57	2,85	9,04	**11,69**
Pantomime	27,17	6,16	-0,24	6,39	4,56	-0,03	4,92	4,19	-0,57	4,49	4,27	-0,59	4,56	**9,49**
Lovebird1	27,75	3,27	-0,81	3,65	3,00	0,31	3,52	15,19	3,83	15,59	15,14	3,82	15,53	**18,16**
Newspaper	27,58	-0,18	-0,81	0,48	1,74	-0,46	2,26	12,25	3,07	12,65	12,33	3,06	12,70	**15,81**
Poznan Street	27,97	3,89	-0,18	4,25	4,45	0,09	4,73	9,58	2,19	9,81	9,60	2,17	9,83	**12,82**
MSSIM (%)														
Ballet	86,10	2,32	-5,69	2,76	1,64	-6,78	2,31	10,02	-1,53	10,22	10,00	-1,61	10,20	**11,64**
Breakdancers	90,48	-0,65	-4,42	0,13	0,46	-4,24	1,04	1,02	-3,35	1,65	1,04	-3,34	1,66	**4,99**
Cafe	90,39	3,16	-1,41	3,50	3,13	-0,70	3,55	6,43	1,01	6,54	6,42	0,99	6,53	**7,58**
Pantomime	90,60	5,27	-0,54	5,69	4,46	-0,52	4,95	4,62	-1,03	4,96	4,60	-1,02	4,94	**7,19**
Lovebird1	78,43	13,62	-1,34	14,14	13,78	1,39	14,31	20,48	9,23	20,53	20,47	9,24	20,53	**20,91**
Newspaper	82,20	-1,67	-3,44	0,44	5,13	-2,62	6,03	15,47	6,36	15,60	15,50	6,31	15,62	**16,45**
Poznan Street	76,19	11,85	-2,45	12,50	13,34	-1,49	13,74	18,97	5,32	19,09	18,98	5,33	19,11	**20,75**

I.1.2 Resultados com codificação

As Tabelas I.5 a I.8 apresentam os ganhos com codificação da super-resolução proposta no Capítulo 4 em relação a interpolar os quadros em baixa resolução, baseado nas médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. Os melhores resultados para cada sequência são indicados por números em negrito. É mantida a nomenclatura utilizada nas colunas das Tabelas I.1 a I.4.

Tabela I.3: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**		
		$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)				
Barn1	27,50	4,78	3,55	**5,45**
Barn2	30,88	4,56	3,23	**4,92**
Bull	32,35	3,69	2,31	**4,11**
Map	28,25	0,64	0,39	**1,06**
Poster	26,15	3,67	1,54	**4,21**
Sawtooth	28,08	3,53	2,32	**4,00**
Venus	28,47	4,02	2,24	**4,46**
Cones	28,41	-0,18	0,26	**0,41**
Teddy	29,69	0,64	0,50	**1,31**
Room3D	24,92	7,31	1,43	**7,70**
MSSIM (%)				
Barn1	80,02	13,92	10,83	**14,53**
Barn2	87,50	7,80	4,95	**8,11**
Bull	91,02	4,86	2,46	**5,19**
Map	91,30	0,85	0,59	**1,71**
Poster	80,68	12,68	8,01	**13,29**
Sawtooth	86,10	7,47	4,54	**8,01**
Venus	86,24	8,53	4,66	**8,92**
Cones	84,80	-2,02	-0,50	**-0,13**
Teddy	89,19	1,56	-0,41	**2,55**
Room3D	87,24	9,57	3,50	**9,81**

Tabela I.4: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), sem codificação, para a componente de luminância das sequências sintética, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**		
		$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)				
Barn1	24,84	3,64	-0,09	**4,18**
Barn2	27,39	5,55	1,66	**5,91**
Bull	28,56	5,11	0,10	**5,48**
Map	21,20	2,78	-0,27	**3,29**
Poster	22,65	3,99	-0,54	**4,43**
Sawtooth	24,53	4,37	0,35	**4,80**
Venus	25,16	4,61	0,37	**4,98**
Cones	24,59	-0,44	-1,15	**0,03**
Teddy	25,91	0,77	-0,58	**1,39**
Room3D	19,23	8,84	-0,04	**9,09**
MSSIM (%)				
Barn1	60,47	27,36	5,18	**27,99**
Barn2	73,83	18,19	6,05	**18,64**
Bull	79,55	13,15	-0,49	**13,65**
Map	49,57	28,83	11,60	**30,34**
Poster	57,62	29,70	6,98	**30,37**
Sawtooth	69,30	19,91	3,70	**20,56**
Venus	72,59	18,06	3,05	**18,54**
Cones	62,77	1,76	-6,52	**3,66**
Teddy	73,82	7,09	-3,43	**8,43**
Room3D	55,83	36,10	4,85	**36,19**

Tabela I.5: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**												
	$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O2}$	$\mathbf{I}_{O3}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O3}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O3}$	$\mathbf{I}_{O4}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O4}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O4}$	$\hat{\mathbf{I}}_{O}$
PSNR (dB)													
Ballet	0,18	-0,62	0,32	0,47	0,27	0,62	2,37	0,15	2,40	2,34	0,11	2,37	**2,55**
Breakdancers	0,11	-0,02	0,17	0,18	-0,01	0,22	0,19	0,07	0,26	0,18	0,04	0,24	**0,55**
Cafe	0,50	0,26	0,60	0,66	0,58	0,75	1,22	0,85	1,26	1,20	0,83	1,25	**1,45**
Pantomime	1,13	0,18	1,23	0,77	0,23	0,95	0,73	0,04	0,85	0,70	0,01	0,82	**1,89**
Lovebird1	0,32	-0,08	0,42	0,37	-0,01	0,46	1,20	0,62	1,22	1,19	0,61	1,20	**1,37**
Newspaper	-0,11	-0,02	0,06	0,12	-0,04	0,27	1,34	0,73	1,36	1,32	0,71	1,34	**1,53**
Poznan Street	0,60	0,06	0,69	0,68	0,12	0,75	1,34	0,59	1,37	1,32	0,57	1,35	**1,61**
MSSIM (%)													
Ballet	0,41	-0,58	0,51	0,34	0,08	0,49	1,60	0,24	1,62	1,58	0,22	1,61	**1,75**
Breakdancers	-0,05	-0,17	0,01	0,01	-0,16	0,05	0,03	-0,11	0,09	0,02	-0,13	0,07	**0,32**
Cafe	0,09	-0,06	0,14	0,11	0,02	0,16	0,33	0,10	0,35	0,32	0,09	0,34	**0,47**
Pantomime	0,33	-0,03	0,36	0,25	-0,01	0,30	0,23	-0,08	0,28	0,23	-0,09	0,27	**0,54**
Lovebird1	0,65	-0,34	0,75	0,67	-0,17	0,74	1,23	0,28	1,25	1,21	0,26	1,23	**1,46**
Newspaper	-0,51	-0,26	-0,25	-0,02	-0,30	0,12	0,94	0,28	0,96	0,92	0,26	0,95	**1,13**
Poznan Street	0,98	-0,22	1,10	1,07	-0,12	1,16	1,80	0,49	1,83	1,78	0,46	1,81	**2,11**

Tabela I.6: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**												
	$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O2}$	$\mathbf{I}_{O3}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O3}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O3}$	$\mathbf{I}_{O4}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O4}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O4}$	$\hat{\mathbf{I}}_{O}$
PSNR (dB)													
Ballet	-0,08	-0,89	0,04	-0,01	-0,90	0,10	1,32	-0,58	1,44	1,30	-0,64	1,42	**1,85**
Breakdancers	0,34	-0,25	0,42	0,50	-0,27	0,56	0,30	-0,17	0,41	0,28	-0,21	0,38	**1,05**
Cafe	0,91	-0,13	1,05	1,03	0,03	1,14	2,02	0,30	2,09	1,96	0,24	2,04	**2,39**
Pantomime	2,32	-0,26	2,39	1,83	-0,16	1,93	1,70	-0,31	1,81	1,68	-0,37	1,79	**3,29**
Lovebird1	0,70	-0,44	0,83	0,71	-0,16	0,87	2,11	0,35	2,17	2,08	0,30	2,14	**2,39**
Newspaper	-0,06	-0,42	0,16	0,60	-0,32	0,76	2,55	0,25	2,59	2,52	0,20	2,57	**2,86**
Poznan Street	0,94	-0,32	1,05	1,06	-0,26	1,15	1,75	-0,01	1,82	1,71	-0,06	1,79	**2,16**
MSSIM (%)													
Ballet	1,20	-2,19	1,29	0,77	-2,63	0,93	2,92	-1,44	3,04	2,89	-1,51	3,01	**3,49**
Breakdancers	0,34	-1,07	0,47	0,58	-1,10	0,68	0,37	-0,98	0,52	0,34	-1,04	0,49	**1,36**
Cafe	1,19	-0,57	1,30	1,22	-0,45	1,33	1,97	-0,14	2,03	1,94	-0,19	2,01	**2,33**
Pantomime	2,66	-0,30	2,73	2,35	-0,24	2,43	2,28	-0,47	2,36	2,28	-0,49	2,36	**3,23**
Lovebird1	3,37	-1,51	3,64	3,66	-0,82	3,89	5,32	0,25	5,44	5,26	0,15	5,39	**5,98**
Newspaper	-0,14	-1,70	0,45	2,17	-1,48	2,45	5,10	0,37	5,18	5,08	0,30	5,16	**5,57**
Poznan Street	3,42	-1,67	3,67	3,66	-1,43	3,83	4,83	-0,60	4,97	4,77	-0,69	4,92	**5,66**

I.2 Super-resolução de múltiplas vistas em resolução mista com mapas de profundidade

I.2.1 Resultados sem codificação

As Tabelas I.9 a I.12 apresentam os ganhos sem codificação da super-resolução proposta no Capítulo 5 em relação a interpolar os quadros em baixa resolução, baseado nas médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

De forma semelhante à Seção I.1, os ganhos obtidos quando se utiliza somente uma referência são apresentados na terceira e quarta colunas. Para cada entrada nestas colunas, tem-se três linhas: na primeira, utiliza-se $\mathbf{I}_{Oi}^{A\,\prime} + \mathbf{I}_{O}^{B}$; na segunda, utiliza-se $\mathbf{I}_{Oi}^{A\,\prime\prime} + \mathbf{I}_{O}^{B}$; e na terceira linha, combina-se $\mathbf{I}_{Oi}^{A\,\prime}$ e $\mathbf{I}_{Oi}^{A\,\prime\prime}$ em um único quadro. A última coluna apresenta os resultados obtidos aplicando a Eq. (5.7) com todas as referências disponíveis. Novamente, os melhores resultados para cada sequência são indicados por números em negrito.

I.2.2 Resultados com codificação

As Tabelas I.13 a I.16 apresentam os ganhos com codificação da super-resolução proposta no Capítulo 5 em relação a interpolar os quadros em baixa resolução, baseado nas médias de PSNR (Eq. 2.16) e MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados. Os melhores resultados para cada sequência são indicados por números em negrito. É mantida a nomenclatura utilizada nas colunas das Tabelas I.9 a I.12.

I.3 Super-resolução de mapas de profundidade em baixa resolução

I.3.1 Resultados sem codificação

A fim de avaliar o método de super-resolução proposto neste Capítulo, mediu-se a qualidade das quadros $\mathbf{I}_{O}^{D}$ super-resolvidos com diferentes mapas de profundidade pré-processados. A qualidade dos mapas de profundidade em si não é tão importante, dado que eles não são o produto final de uma arquitetura em múltiplas vistas, e sim as imagens $\hat{\mathbf{I}}_{O}$ e $\mathbf{I}_{Ri}$. Foram utilizadas as mesmas sequências apresentadas nas Seções 4.4 e 5.4, com e sem codificação H.264/AVC. Os quadros de referência disponíveis são os mesmos da Tabela 5.1.

As Tabelas I.17 e I.18 apresentam os ganhos de qualidade dos quadros $\mathbf{I}_{O}^{D}$ super-resolvidos com diferentes mapas de profundidade pré-processados do Capítulo 6 em relação a interpolar os quadros em baixa resolução, para testes sem codificação. Os mapas pré-processados são:

- $\mathbf{D}_r$: mapas de profundidade em resolução completa (o equivalente à arquitetura do Capítulo 5);
- $\mathbf{D}_r^B$: mapas $\mathbf{D}_r$ decimados e interpolados;

Tabela I.7: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**		
	$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)			
Barn1	2,36	0,73	**2,65**
Barn2	1,86	0,62	**1,98**
Bull	1,55	0,25	**1,67**
Map	0,47	-0,12	**0,75**
Poster	2,23	0,24	**2,52**
Sawtooth	2,16	0,72	**2,37**
Venus	2,31	0,57	**2,57**
Cones	-0,01	-0,09	**0,32**
Teddy	0,38	0,12	**0,74**
Room3D	2,55	0,41	**2,65**
MSSIM (%)			
Barn1	9,81	2,70	**10,17**
Barn2	4,37	0,72	**4,51**
Bull	2,63	-0,02	**2,76**
Map	1,52	-1,02	**2,22**
Poster	10,37	2,04	**10,65**
Sawtooth	5,57	1,14	**5,85**
Venus	6,04	1,22	**6,32**
Cones	-0,44	-0,96	**0,64**
Teddy	1,03	-0,46	**1,63**
Room3D	4,03	0,46	**4,13**

Tabela I.8: Resultados do método de super-resolução de múltiplas vistas em resolução mista sem mapas de profundidade (Cap. 4), com codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**		
	$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)			
Barn1	0,25	-1,02	**0,49**
Barn2	1,67	-0,46	**1,91**
Bull	1,69	-0,68	**1,85**
Map	1,61	-1,02	**2,01**
Poster	1,01	-1,11	**1,36**
Sawtooth	2,11	-0,70	**2,35**
Venus	1,61	-0,74	**1,94**
Cones	-0,53	-1,15	**-0,19**
Teddy	0,25	-0,86	**0,60**
Room3D	3,15	-0,26	**3,27**
MSSIM (%)			
Barn1	9,06	-5,08	**9,50**
Barn2	7,23	-2,48	**7,79**
Bull	5,94	-3,34	**6,24**
Map	21,19	-0,32	**22,46**
Poster	14,26	-3,97	**14,63**
Sawtooth	11,57	-3,02	**12,11**
Venus	8,62	-3,44	**9,17**
Cones	-0,18	-7,55	**1,15**
Teddy	3,47	-4,67	**4,36**
Room3D	13,54	-1,08	**13,68**

Tabela I.9: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**						
		$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_O^B$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O2}$	$\hat{\mathbf{I}}_O$
PSNR (dB)								
Ballet	33,49	3,09	3,01	3,18	2,90	2,75	2,97	**3,93**
Breakdancers	39,17	1,09	1,23	1,56	1,20	1,55	1,70	**2,54**
Cafe	36,83	2,88	3,00	3,30	1,45	1,49	1,72	**4,76**
Pantomime	34,50	4,01	3,60	4,11	2,58	2,30	2,98	**5,25**
Lovebird1	34,27	-0,05	0,11	0,46	1,62	1,69	1,81	**3,07**
Newspaper	34,00	0,38	0,67	1,12	1,51	1,38	1,81	**2,47**
Poznan Street	33,16	3,87	3,62	3,97	4,04	3,85	4,17	**5,71**
MSSIM (%)								
Ballet	94,46	1,77	1,85	1,98	1,52	1,57	1,78	**2,50**
Breakdancers	96,50	-0,18	-0,06	0,07	0,02	0,18	0,25	**0,74**
Cafe	97,38	0,40	0,51	0,55	0,11	0,19	0,25	**0,99**
Pantomime	97,79	0,59	0,54	0,62	0,29	0,26	0,38	**0,91**
Lovebird1	94,91	1,89	1,97	2,07	2,13	2,15	2,21	**3,15**
Newspaper	95,09	-0,04	0,29	0,71	1,01	0,98	1,20	**1,56**
Poznan Street	91,33	4,51	4,41	4,56	4,44	4,39	4,50	**5,59**

Tabela I.10: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**						
		$\mathbf{I}_{O1}^{A\,\prime} + \mathbf{I}_O^B$	$\mathbf{I}_{O1}^{A\,\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime} + \mathbf{I}_O^B$	$\mathbf{I}_{O2}^{A\,\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_{O2}$	$\hat{\mathbf{I}}_O$
PSNR (dB)								
Ballet	29,62	3,73	3,42	3,72	3,82	3,47	3,79	**4,88**
Breakdancers	33,62	3,44	3,09	3,54	3,62	3,47	3,79	**4,63**
Cafe	30,24	5,13	4,46	5,12	3,35	2,87	3,34	**7,15**
Pantomime	27,16	7,24	5,69	6,95	5,53	4,21	5,25	**7,54**
Lovebird1	27,75	1,11	1,26	1,36	3,57	3,29	3,56	**5,12**
Newspaper	27,58	1,86	1,33	1,97	2,74	1,94	2,65	**3,22**
Poznan Street	27,98	6,09	5,09	5,91	6,46	5,60	6,35	**7,81**
MSSIM (%)								
Ballet	86,10	6,79	6,39	6,85	6,75	6,37	6,88	**8,32**
Breakdancers	90,50	3,35	3,17	3,51	3,72	3,66	3,88	**4,72**
Cafe	90,39	5,38	5,13	5,42	4,48	4,18	4,52	**6,35**
Pantomime	90,60	6,11	5,48	6,01	5,16	4,44	5,04	**6,32**
Lovebird1	78,44	13,66	13,42	13,74	14,32	14,02	14,31	**17,02**
Newspaper	82,20	4,99	3,93	5,36	6,98	5,44	6,75	**7,78**
Poznan Street	76,19	17,42	16,62	17,28	17,35	16,69	17,24	**18,84**

Tabela I.11: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**		
		$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)				
Barn1	27,50	8,35	8,06	**8,39**
Barn2	30,88	**7,32**	6,90	7,28
Bull	32,35	**5,95**	5,78	5,92
Map	28,25	2,93	3,00	**3,05**
Poster	26,15	**7,89**	7,63	7,82
Sawtooth	28,08	**5,31**	5,00	5,25
Venus	28,47	**7,35**	7,03	7,28
Cones	28,41	3,80	4,35	**4,46**
Teddy	29,69	3,21	3,56	**3,84**
Room3D	24,92	2,06	2,07	**2,11**
MSSIM (%)				
Barn1	80,02	17,25	17,28	**17,36**
Barn2	87,50	9,95	9,96	**10,03**
Bull	91,02	6,58	6,60	**6,61**
Map	91,30	4,30	4,46	**4,47**
Poster	80,68	**16,78**	16,70	16,77
Sawtooth	86,10	9,61	9,73	**9,76**
Venus	86,24	11,23	11,23	**11,27**
Cones	84,80	9,91	10,49	**10,56**
Teddy	89,19	6,29	6,82	**6,94**
Room3D	87,23	4,41	4,42	**4,44**

Tabela I.12: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), sem codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{I}_O^B$	**Ganhos**		
		$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)				
Barn1	24,84	**9,31**	8,68	9,21
Barn2	27,39	**9,12**	8,40	8,95
Bull	28,56	**8,10**	7,83	8,05
Map	21,20	6,88	6,70	**6,90**
Poster	22,65	**9,28**	8,88	9,17
Sawtooth	24,53	**6,93**	6,09	6,61
Venus	25,16	**8,90**	8,24	8,73
Cones	24,59	4,82	4,78	**5,07**
Teddy	25,91	4,34	4,06	**4,55**
Room3D	19,22	3,70	3,68	**3,71**
MSSIM (%)				
Barn1	60,47	**35,78**	35,42	35,73
Barn2	73,83	**22,83**	22,56	22,82
Bull	79,55	**17,41**	17,34	17,41
Map	49,57	41,18	41,06	**41,24**
Poster	57,62	**38,76**	38,52	38,70
Sawtooth	69,30	**24,84**	24,39	24,74
Venus	72,59	**24,14**	23,89	24,10
Cones	62,77	27,53	27,20	**27,78**
Teddy	73,82	18,66	18,34	**18,92**
Room3D	55,82	26,09	26,04	**26,10**

Tabela I.13: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências reais, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**						
	$\mathbf{I}_{O1}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O1}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime}+\mathbf{I}_{O}^{B}$	$\mathbf{I}_{O2}^{A\,\prime\prime}+\mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O2}$	$\hat{\mathbf{I}}_{O}$
PSNR (dB)							
Ballet	2,03	1,98	2,11	1,72	1,66	1,79	**2,34**
Breakdancers	0,33	0,38	0,52	0,34	0,46	0,53	**0,74**
Cafe	1,28	1,33	1,45	0,62	0,64	0,75	**1,86**
Pantomime	3,03	2,93	3,06	2,27	2,19	2,42	**3,37**
Lovebird1	-0,00	0,12	0,25	0,57	0,59	0,65	**1,07**
Newspaper	0,18	0,29	0,52	0,79	0,75	0,93	**1,09**
Poznan Street	1,86	1,81	1,91	1,87	1,82	1,91	**2,46**
MSSIM (%)							
Ballet	1,71	1,79	1,96	1,12	1,23	1,50	**2,08**
Breakdancers	-0,11	-0,01	0,16	0,02	0,15	0,24	**0,44**
Cafe	0,34	0,45	0,49	0,11	0,19	0,26	**0,68**
Pantomime	0,88	0,86	0,90	0,67	0,66	0,73	**0,98**
Lovebird1	1,33	1,36	1,46	1,37	1,37	1,43	**2,07**
Newspaper	-0,39	-0,17	0,22	0,46	0,44	0,63	**0,78**
Poznan Street	3,40	3,37	3,45	3,20	3,17	3,24	**3,98**

Tabela I.14: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências reais, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**						
	$\mathbf{I}_{O1}^{A\,\prime} + \mathbf{I}_{O}^{B}$	$\mathbf{I}_{O1}^{A\,\prime\prime} + \mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O1}$	$\mathbf{I}_{O2}^{A\,\prime} + \mathbf{I}_{O}^{B}$	$\mathbf{I}_{O2}^{A\,\prime\prime} + \mathbf{I}_{O}^{B}$	$\hat{\mathbf{I}}_{O2}$	$\hat{\mathbf{I}}_{O}$
PSNR (dB)							
Ballet	2,66	2,38	2,64	2,57	2,32	2,55	**3,16**
Breakdancers	1,70	1,51	1,75	1,74	1,63	1,81	**2,11**
Cafe	3,33	2,94	3,29	2,26	1,96	2,23	**4,04**
Pantomime	6,68	6,24	6,59	5,24	4,80	5,17	**6,75**
Lovebird1	0,93	1,04	1,10	2,29	2,13	2,28	**2,94**
Newspaper	1,44	1,16	1,51	2,09	1,65	2,04	**2,27**
Poznan Street	3,94	3,62	3,89	4,07	3,80	4,04	**4,73**
MSSIM (%)							
Ballet	5,86	5,38	5,93	5,38	4,93	5,57	**6,67**
Breakdancers	2,48	2,23	2,65	2,72	2,57	2,86	**3,34**
Cafe	4,38	4,19	4,45	3,64	3,40	3,69	**4,89**
Pantomime	6,29	6,10	6,27	5,50	5,25	5,49	**6,35**
Lovebird1	11,33	11,12	11,39	11,66	11,38	11,64	**13,35**
Newspaper	4,05	3,44	4,41	5,73	4,74	5,62	**6,12**
Poznan Street	14,14	13,74	14,06	13,84	13,51	13,79	**15,19**

Tabela I.15: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências sintéticas, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**		
	$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)			
Barn1	3,69	3,57	**3,72**
Barn2	2,76	2,71	**2,81**
Bull	2,31	2,27	**2,31**
Map	1,66	1,70	**1,74**
Poster	4,30	4,16	**4,30**
Sawtooth	**2,88**	2,73	2,88
Venus	**3,86**	3,73	3,86
Cones	1,76	1,88	**2,00**
Teddy	1,59	1,64	**1,81**
Room3D	1,81	1,81	**1,84**
MSSIM (%)			
Barn1	13,89	13,79	**13,96**
Barn2	6,63	6,65	**6,73**
Bull	4,00	4,00	**4,02**
Map	4,45	4,55	**4,58**
Poster	**14,90**	14,75	14,89
Sawtooth	7,15	7,14	**7,25**
Venus	9,07	9,01	**9,10**
Cones	6,52	6,67	**6,94**
Teddy	3,85	4,09	**4,28**
Room3D	4,10	4,10	**4,13**

Tabela I.16: Resultados do método de super-resolução de múltiplas vistas em resolução mista com mapas de profundidade (Cap. 5), com codificação, para a componente de luminância das sequências sintéticas, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	**Ganhos**		
	$\mathbf{I}_O^{A\prime} + \mathbf{I}_O^B$	$\mathbf{I}_O^{A\prime\prime} + \mathbf{I}_O^B$	$\hat{\mathbf{I}}_O$
PSNR (dB)			
Barn1	**4,69**	4,33	4,64
Barn2	**4,31**	4,00	4,25
Bull	**4,00**	3,91	3,99
Map	**5,21**	5,04	5,21
Poster	**5,88**	5,50	5,79
Sawtooth	**4,35**	3,86	4,21
Venus	**5,29**	4,93	5,20
Cones	2,71	2,51	**2,78**
Teddy	2,44	2,18	**2,50**
Room3D	3,51	3,49	**3,52**
MSSIM (%)			
Barn1	**28,71**	28,01	28,53
Barn2	**16,32**	15,95	16,27
Bull	**12,26**	12,17	12,26
Map	**40,57**	40,19	40,53
Poster	**34,30**	33,68	34,12
Sawtooth	**19,72**	19,03	19,56
Venus	**19,49**	19,07	19,39
Cones	**19,99**	18,90	19,91
Teddy	12,85	12,16	**12,92**
Room3D	23,97	23,92	**23,97**

- $\mathbf{D}_r^{B,MED}$: $\mathbf{D}_r^B$ filtrado pela mediana;
- $\hat{\mathbf{D}}_{r1}$: $\mathbf{D}_r^B$ mutuamente super-resolvidos com uma referência (método proposto neste Capítulo);
- $\hat{\mathbf{D}}_r$: $\mathbf{D}_r^B$ mutuamente super-resolvidos com todas as referências (método proposto neste Capítulo).

I.3.2 Resultados com codificação

Assim como nas Seções I.1.2 e I.2.2, os processamentos apresentados no Capítulo 6 foram aplicados às mesmas sequências após estas serem codificadas em resolução mista, tal como na Fig. 6.1, a fim de avaliar seu desempenho em uma situação prática.

Tabela I.17: Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), sem codificação, para a componente de luminância de todas as sequências, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
PSNR (dB)					
Ballet	3,93	3,52	3,63	3,78	**3,94**
Breakdancers	**2,54**	2,25	2,31	2,30	2,30
Cafe	**4,76**	4,48	4,56	4,61	4,66
Pantomime	5,25	5,83	5,97	6,20	**6,30**
Lovebird1	**3,07**	3,00	2,98	2,98	2,87
Newspaper	2,47	2,48	2,51	2,79	**2,93**
Poznan Street	5,71	5,63	5,64	5,77	**5,78**
Barn1	**8,39**	7,45	7,57	7,67	7,67
Barn2	**7,28**	6,39	6,51	6,60	6,60
Bull	5,92	**5,92**	5,90	5,88	5,88
Map	**3,05**	2,81	2,90	2,99	2,99
Poster	**7,82**	7,33	7,41	7,50	7,50
Sawtooth	**5,25**	4,65	4,79	4,83	4,83
Venus	**7,28**	6,67	6,71	6,63	6,63
Cones	**4,46**	4,05	4,20	4,31	4,31
Teddy	**3,84**	3,57	3,68	3,65	3,65
Room3D	2,11	**2,12**	2,12	2,12	2,12
MSSIM (%)					
Ballet	**2,50**	2,30	2,36	2,41	2,48
Breakdancers	**0,74**	0,70	0,71	0,69	0,69
Cafe	**0,99**	0,96	0,97	0,97	0,98
Pantomime	0,91	0,96	0,97	0,98	**0,98**
Lovebird1	**3,15**	3,13	3,12	3,13	3,12
Newspaper	1,56	1,52	1,53	1,67	**1,77**
Poznan Street	**5,59**	5,55	5,54	5,57	5,58
Barn1	**17,36**	16,98	17,01	17,03	17,03
Barn2	**10,03**	9,75	9,80	9,84	9,84
Bull	**6,61**	6,60	6,59	6,58	6,58
Map	**4,47**	4,29	4,37	4,46	4,46
Poster	**16,77**	16,46	16,51	16,54	16,54
Sawtooth	**9,76**	9,51	9,60	9,61	9,61
Venus	**11,27**	11,04	11,06	11,05	11,05
Cones	**10,56**	9,87	10,08	10,21	10,21
Teddy	**6,94**	6,69	6,79	6,76	6,76
Room3D	4,44	4,48	4,48	**4,48**	**4,48**

Tabela I.18: Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), sem codificação, para a componente de luminância de todas as sequências, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
PSNR (dB)					
Ballet	**4,88**	3,78	3,83	3,88	3,94
Breakdancers	**4,63**	3,78	3,77	3,76	3,77
Cafe	**7,15**	6,16	6,18	6,16	6,24
Pantomime	7,54	9,17	9,12	9,24	**9,32**
Lovebird1	**5,12**	4,86	4,80	4,89	4,91
Newspaper	3,22	3,19	3,17	3,29	**3,34**
Poznan Street	**7,81**	7,71	7,66	7,75	7,78
Barn1	**9,21**	6,71	6,67	6,78	6,78
Barn2	**8,95**	6,78	6,75	6,77	6,77
Bull	**8,05**	7,88	7,83	7,86	7,86
Map	**6,90**	4,86	4,88	4,96	4,96
Poster	**9,17**	7,60	7,55	7,63	7,63
Sawtooth	**6,61**	5,20	5,21	5,14	5,14
Venus	**8,73**	7,14	7,07	7,23	7,23
Cones	**5,07**	3,57	3,55	3,60	3,60
Teddy	**4,55**	3,36	3,34	3,38	3,38
Room3D	**3,71**	3,70	3,71	3,70	3,70
MSSIM (%)					
Ballet	**8,32**	6,93	6,96	7,00	7,07
Breakdancers	**4,72**	4,33	4,32	4,28	4,28
Cafe	**6,35**	5,93	5,94	5,94	5,97
Pantomime	6,32	6,70	6,68	6,72	**6,74**
Lovebird1	**17,02**	16,78	16,76	16,81	16,82
Newspaper	7,78	7,65	7,64	7,80	**7,89**
Poznan Street	**18,84**	18,69	18,66	18,70	18,73
Barn1	**35,73**	33,79	33,75	33,85	33,85
Barn2	**22,82**	21,38	21,36	21,40	21,40
Bull	**17,41**	17,30	17,27	17,24	17,24
Map	**41,24**	37,11	37,27	37,43	37,43
Poster	**38,70**	37,14	37,07	37,17	37,17
Sawtooth	**24,74**	23,17	23,19	23,14	23,14
Venus	**24,10**	23,09	23,04	23,15	23,15
Cones	**27,78**	22,91	22,88	23,13	23,13
Teddy	**18,92**	16,37	16,34	16,41	16,41
Room3D	26,10	26,10	**26,11**	26,08	26,08

Tabela I.19: Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), com codificação, para a componente de luminância de todas as sequências, $M = 2$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
PSNR (dB)					
Ballet	2,34	2,79	2,83	2,91	**3,01**
Breakdancers	0,74	1,41	1,42	1,44	**1,45**
Cafe	1,86	2,28	2,31	2,34	**2,35**
Pantomime	3,37	3,85	3,87	3,91	**3,93**
Lovebird1	1,07	1,21	1,21	**1,22**	1,18
Newspaper	1,09	1,51	1,52	1,58	**1,66**
Poznan Street	2,46	2,59	2,59	2,63	**2,64**
Barn1	**3,72**	3,59	3,62	3,65	3,65
Barn2	2,81	2,76	2,79	**2,82**	**2,82**
Bull	2,31	2,35	2,35	**2,36**	**2,36**
Map	1,74	1,69	1,73	**1,74**	**1,74**
Poster	**4,30**	4,10	4,13	4,17	4,17
Sawtooth	**2,88**	2,82	2,87	2,87	2,87
Venus	**3,86**	3,67	3,66	3,70	3,70
Cones	2,00	2,15	2,17	**2,21**	**2,21**
Teddy	1,81	2,15	2,17	**2,20**	**2,20**
Room3D	1,84	1,86	1,85	**1,86**	**1,86**
MSSIM (%)					
Ballet	2,08	3,00	3,04	3,09	**3,16**
Breakdancers	0,44	1,29	1,29	1,29	**1,31**
Cafe	0,68	1,21	1,22	1,23	**1,23**
Pantomime	0,98	1,19	**1,19**	1,18	1,18
Lovebird1	2,07	2,52	2,52	**2,53**	2,52
Newspaper	0,78	1,56	1,57	1,61	**1,68**
Poznan Street	3,98	4,37	4,36	4,39	**4,40**
Barn1	13,96	14,17	14,21	**14,22**	**14,22**
Barn2	6,73	6,97	7,01	**7,05**	**7,05**
Bull	4,02	**4,20**	4,19	4,19	4,19
Map	4,58	4,66	4,73	**4,78**	**4,78**
Poster	14,89	14,98	15,02	**15,05**	**15,05**
Sawtooth	7,25	7,58	**7,65**	7,65	7,65
Venus	9,10	9,18	9,18	**9,21**	**9,21**
Cones	6,94	8,56	8,65	**8,75**	**8,75**
Teddy	4,28	5,96	6,00	**6,02**	**6,02**
Room3D	4,13	4,30	4,29	**4,31**	**4,31**

Tabela I.20: Resultados do método de super-resolução de profundidade em baixa resolução (Cap. 6), com codificação, para a componente de luminância de todas as sequências, $M = 4$. São apresentados os ganhos da super-resolução proposta em relação a interpolar os quadros em baixa resolução, baseado na média de PSNR (Eq. 2.16) e de MSSIM×100 (Eq. 2.20) dos quadros super-resolvidos e interpolados.

Sequência	$\mathbf{D}_r$	$\mathbf{D}_r^B$	$\mathbf{D}_r^{B,MED}$	$\hat{\mathbf{D}}_{r1}$	$\hat{\mathbf{D}}_r$
PSNR (dB)					
Ballet	3,16	3,20	3,22	3,25	**3,31**
Breakdancers	2,11	2,69	2,68	2,69	**2,71**
Cafe	4,04	4,13	4,14	4,14	**4,18**
Pantomime	6,75	7,43	7,42	7,46	**7,48**
Lovebird1	2,94	2,95	2,94	2,94	**2,95**
Newspaper	2,27	2,51	2,50	2,55	**2,59**
Poznan Street	**4,73**	4,65	4,63	4,67	4,69
Barn1	**4,64**	3,85	3,86	3,88	3,88
Barn2	**4,25**	3,61	3,61	3,61	3,61
Bull	**3,99**	3,95	3,93	3,94	3,94
Map	**5,21**	3,92	3,97	4,03	4,03
Poster	**5,79**	4,99	4,96	5,00	5,00
Sawtooth	**4,21**	3,53	3,53	3,51	3,51
Venus	**5,20**	4,55	4,53	4,57	4,57
Cones	**2,78**	2,22	2,21	2,25	2,25
Teddy	**2,50**	2,46	2,45	2,46	2,46
Room3D	**3,52**	3,51	3,51	3,51	3,51
MSSIM (%)					
Ballet	6,67	6,96	6,96	7,00	**7,07**
Breakdancers	3,34	**4,29**	4,26	4,26	4,28
Cafe	4,89	5,45	5,46	5,48	**5,50**
Pantomime	6,35	6,85	6,85	6,86	**6,87**
Lovebird1	13,35	13,78	13,78	13,79	**13,81**
Newspaper	6,12	7,10	7,09	7,19	**7,26**
Poznan Street	15,19	15,26	15,24	15,28	**15,30**
Barn1	**28,53**	27,24	27,23	27,28	27,28
Barn2	**16,27**	15,39	15,38	15,37	15,37
Bull	12,26	**12,34**	12,32	12,29	12,29
Map	**40,53**	36,63	36,83	36,98	36,98
Poster	**34,12**	32,67	32,62	32,70	32,70
Sawtooth	**19,56**	18,53	18,53	18,53	18,53
Venus	**19,39**	18,75	18,71	18,76	18,76
Cones	**19,91**	17,66	17,63	17,84	17,84
Teddy	12,92	13,17	13,16	**13,21**	**13,21**
Room3D	23,97	24,13	**24,13**	24,11	24,11